УДК 004
ББК 32.81
А 72

Рецензенты:

Кафедра АСОИУ Московского государственного технического университета им. Н.Э. Баумана (зав. кафедрой – д-р техн. наук, проф. *В.М. Черненький*); д-р физ.-мат. наук, проф. *В.В. Нечаев* (зав. кафедрой интеллектуальных технологий и систем Московского государственного института радиотехники, электроники и автоматики)

Антонов, А.В.

А 72 Системный анализ. Учеб. для вузов/А.В. Антонов. – М.: Высш. шк., 2004. – 454 с.: ил.

ISBN 5-06-004862-4

В учебнике изложены методологические вопросы системиого анализа. Описаны этапы и процедуры проведения системных исследований, сформулированы цели и задачи системного анализа. Большое место уделено вопросам построения моделей сложных систем. Изложены вопросы проверки адекватности моделей, процедуры их формирования, методы оценки параметров.

Рассмотрены математические методы и модели системного анализа, типовые постановки задач, описаны области их приложения. Изложены численные методы решения типовых задач системного анализа. Приведены методы выбора и принятия решений, процедур, выполняемых на заключительном этапе системного анализа. Дана характеристика задач принятия решений.

Для студентов, обучающихся по направлению 552800 и 654600 «Информатика и вычислительная техника» и образовательной программе (специальности), реализуемой в рамках направления подготовки дипломированных специалистов 220200 – «Автоматизированные системы обработки информации и управления», а также для аспирантов и инженеров.

УДК 004
ББК 32.81

ISBN 5-06-004862-4

ПРЕДИСЛОВИЕ

Системные исследования – интенсивно развивающаяся область научной деятельности, которая является одним из наиболее результативных проявлений интегративных тенденций в науке. Специфика системных исследований состоит в их направленности на изучение сложных, комплексных, крупномасштабных проблем. В ходе проведения данного вида работ исследователи ориентируются не только на познание существа изучаемых проблем и соответствующих объектов, но и на создание средств, позволяющих обеспечить рациональное управление этими объектами, содействовать разрешению имеющихся проблем. Единство исследовательских функций и решение практических задач, направленных на преобразование объекта исследования, разрешение проблемной ситуации, имеющей место в исследуемой системе, обуславливают комплексный, междисциплинарный характер системных исследований.

Системный анализ является синтетической дисциплиной. В нем находит отражение междисциплинарный характер системных исследований, реализуется современная форма синтеза научных знаний. В своей простейшей интерпретации междисциплинарность выражается в том, что системный анализ занимается изучением объектов такой сложности, для описания которых приходится привлекать понятия, изучаемые в рамках различных традиционных научных дисциплин. Реально содержание этого понятия гораздо глубже. Дело в том, что традиционные дисциплины изучают различные аспекты поведения исследуемых систем. В системных исследованиях такая декомпозиция невозможна, так как при этом могут потеряться основные свойства системы. Иными словами необходимо учитывать системный эффект, когда совокупность объектов, объединенная в систему, приводит к появлению новых свойств. Таким образом, для понимания поведения системы необходимы теоретические знания различных дисциплин. Причем для исследования систем применяются не только формализованные методы, но и неформальные процедуры.

Исторически системный анализ явился развитием таких дисциплин как исследование операций и системотехника. Системный анализ и исторически и содержательно имеет вполне определенный смысл, а именно, он представляет собой совокупность методов исследования систем, методик выработки и принятия решений при проектировании, конструировании и управлении сложными объектами различной природы.

Системный анализ – это, прежде всего, определенный тип научно-технической деятельности, необходимый для исследования, разработки, управления сложными объектами. Результаты системных исследований, для того чтобы быть успешными, должны удовлетворять заранее установленным критериям эффективности, опираться на определенный теоретический фундамент и в процессе своего применения порождать образцы для последующего использования.

ВВЕДЕНИЕ

Современное состояние общества характеризуется внедрением достижений научно-технического прогресса во все сферы деятельности. Переживаемый в настоящее время этап развития является этапом информатизации. Информатизация – это процесс создания, развития и всеобщего применения информационных средств и технологий, обеспечивающих кардинальное улучшение качества труда и условий жизни в обществе. Информатизация тесно связана с внедрением информационно-вычислительных систем, с повышением уровня автоматизации организационно-экономической, технологической, административно-хозяйственной, проектно-конструкторской, научно-исследовательской и других видов деятельности. Создание сложных технических систем, проектирование и управление сложными комплексами, анализ экологической ситуации, особенно в условиях агрессивного техногенного воздействия, исследование социальных проблем коллективов, планирование развития регионов и многие другие направления деятельности требуют организации исследований, которые имеют нетрадиционный характер. По ряду специфических признаков все перечисленные объекты прикладной деятельности обладают свойствами больших систем. Таким образом, в различных сферах деятельности приходится сталкиваться с понятиями больших или сложных систем.

В разных сферах практической деятельности развивались соответствующие методы анализа и синтеза сложных систем: в инженерной деятельности – системотехника, методы проектирования, методы инженерного творчества; в сфере управления – системный подход, политология; в военной сфере – методы исследования операций, теория оптимального управления; в научных исследованиях – имитационное моделирование, теория эксперимента. В 80-е гг. XX в. все эти теоретические и прикладные дисциплины приобретают общую направленность, они образуют «системное движение». Системность стала не только теоретической категорией, но и аспектом практической деятельности. Ввиду того, что сложные системы стали предметом изучения, проектирования и управления, потребовалось обобщение методов исследования систем. Появилась объективная необходимость в возникновении прикладной науки, устанавливающей связь между абстрактными теориями системности и системной практикой. В последнее время это движение оформилось в науку, которая получила название «системный анализ».

Особенности современного системного анализа вытекают из самой природы сложных систем. Имея в качестве цели ликвидацию проблемы или, как минимум, выяснение ее причин, системный анализ привлекает для этого широкий спектр средств, использует возможности различных наук и практических сфер деятельности. Являясь по существу прикладной диалектикой, системный анализ придает большое значение методологическим аспектам любого системного исследования. С другой стороны, прикладная направленность системного анализа приводит к необходимости использования всех современных средств научных исследований – математики, вычислительной техники, моделирования, натурных наблюдений и экспериментов.

Системный анализ является меж- и наддисциплинарным курсом, обобщающим методологию исследования сложных технических, природных и социальных систем. Для проведения анализа и синтеза сложных систем используется широкий спектр математических методов. Основу математического аппарата данной дисциплины составляют линейное и нелинейное программирование, теория принятия решений, теория игр, имитационное моделирование, теория массового обслуживания, теория статистических выводов и т.п.

В настоящее время методы системного анализа получили широкое применение при перспективном и текущем планировании научно-исследовательских работ, проектировании различных объектов, управлении производственными и технологическими процессами, прогнозировании развития отдельных отраслей промышленности и сельского хозяйства. Особенно часто к ним обращаются при решении задач распределения трудовых ресурсов и производственных запасов, назначения сроков профилактического ремонта оборудования, выбора средств транспортировки грузов, составления маршрутов и расписаний перевозок, размещения новых производственных комплексов, сбора информации в автоматизированных системах управления и целого ряда других. Следует также обратить внимание на то обстоятельство, что при решении задач системного анализа наряду со строгим математическим аппаратом применяются эвристические методы. Так, например, при решении задач проектирования принимают участие группы людей, которые оказывают большое влияние как на сам процесс проектирования, так и на принятие решения на отдельных этапах выполнения проекта. Естественно, что при принятии решения проектировщики учитывают не только рекомендации, полученные на основе расчетов, проводимых с помощью вычислительных машин, но и свои соображения, зачастую носящие качественный характер.

Следует отметить еще одну особенность задач системного анализа, а именно, требование оптимальности принимаемых решеиий. То есть, в настоящее время перед системными аналитиками ставится задача не просто разрешения той или иной проблемы, а выработка таких рекомендаций, которые бы гарантировали оптимальность решения.

Решение вопросов проведения и организации системных исследований связано со специфическими особенностями и проблемами, требующими для своего разрешения привлечения результатов широкого спектра научных дисциплин. В ходе исследования реальной системы обычно приходится сталкиваться с самыми разнообразными проблемами; быть профессионалом в каждой из них одному человеку невозможно. Специалист, занимающийся системным анализом, должен иметь образование и опыт, необходимые для анализа и классификации конкретных проблем, для определения перечня специалистов, способных решить конкретные задачи анализа. Это предъявляет особые требования к специалистам-системщикам: они должны обладать широкой эрудицией, раскованностью мышления, умением привлекать людей к работе, организовывать коллективную деятельность.

Глава 1
ОПРЕДЕЛЕНИЯ СИСТЕМНОГО АНАЛИЗА

1.1. Системность – общее свойство материи

Современный этап развития теории и практики характеризуется повышением уровня системности. Ученые, инженеры, представители различных профессий оперируют такими понятиями как системный или комплексный подход. Полезность и важность системного подхода вышла за рамки специальных научных истин и стала привычной, общепринятой. Такая ситуация явилась отражением объективных процессов развития представлений о материальном мире, сформировалась под воздействием объективных факторов.

В своей работе [1] Ф.И. Перегудов и Ф.П. Тарасенко говорят о том, что свойство системности является всеобщим свойством материи. Современные научные данные и современные системные представления позволяют говорить о мире как о бесконечной иерархической системе систем. Причем части системы находятся в развитии, на разных стадиях развития, на разных уровнях системной иерархии и организации. Системность как всеобщее свойство материи проявляется через следующие составляющие: системность практической деятельности, системность познавательной деятельности и системность среды, окружающей человека.

Рассмотрим практическую деятельность человека, т. е. его активное и целенаправленное воздействие на окружающую среду. Покажем, что человеческая практика системна. Отметим очевидные и обязательные признаки системности: *структурированность системы, взаимосвязанность составляющих ее частей, подчиненность организации всей системы определенной цели.* По отношению к человеческой деятельности эти признаки очевидны. Всякое осознанное действие преследует определенную цель. Во всяком действии достаточно просто увидеть его составные части, более мелкие действия. При этом легко убедиться, что эти составные части должны выполняться не в произвольном порядке, а в определенной их последовательности. Это и есть та самая определенная, подчиненная цели взаимосвязанность составных частей, которая и является признаком системности.

Название для такого построения деятельности – *алгоритмичность.* Понятие **алгоритма** возникло сначала в математике и означало задание точно определенной последовательности однозначно понимаемых операций над числами или другими математическими объектами. В настоящее время понятие алгоритма применяется к различным отраслям деятельности. Так говорят не только об алгоритмах принятия управленческих решений, об алгоритмах обучения, алгоритмах написания программ, но и об алгоритмах изобретательства [2]. Алгоритмизуются такие виды деятельности как игра в шахматы, доказательство теорем и т. п. При этом делается отход от математического понимания алгоритма. Важно сознавать, что в алгоритме должна сохраняться логическая последовательность действий. При этом допускается, что в алгоритме определенного вида деятельности могут присутствовать неформализованные виды действия. Важно лишь, чтобы определенные этапы алгоритма успешно, хотя бы и неосознанно, выполнялись человеком.

Р.Х. Зарипов в своей работе [3, с. 12] отмечает: «...подавляющее большинство элементов творческой деятельности, реализуемых человеком «легко и просто», «не думая», «по интуиции», на самом деле являются неосознанной реализацией определенных алгоритмизируемых закономерностей, реализацией неосознаваемых, но объективно существующих и формализуемых критериев красоты и вкуса».

Из данной цитаты можно сделать следующие выводы. Во-первых, всякая деятельность алгоритмична. Во-вторых, не всегда алгоритм реальной деятельности осознается – ряд процессов человек выполняет интуитивно, т. е. его способность решать некоторые задачи доведена до автоматизма. Это есть признак профессионализма, который вовсе не означает, что в действиях профессионала отсутствует алгоритм. В-третьих, в случае неудовлетворенности результатом деятельности возможную причину неудачи следует искать в несовершенстве алгоритма. Это означает пытаться выявить алгоритм, исследовать его, искать «слабые места», устранять их, т. е. совершенствовать алгоритм и, следовательно, повышать системность деятельности. Таким образом, явная алгоритмизация любой практической деятельности является важным средством ее развития.

Системными являются также результаты практической деятельности. Следует отметить, что роль системных представлений в практике постоянно увеличивается, что растет сама системность человеческой деятельности. Данный тезис можно пояснить на примере проектирования технических объектов. Если раньше перед разработчиками новых образцов техники ставилась задача создания работоспособного объекта, то в настоящее время практика ставит задачу создания новых объектов с некоторыми оптимальными свойствами, т. е. к разрабатываемым образцам еще на этапе проектирования предъявляются требования оптимальности. Цели, которые ставятся перед разработчиками, таким образом, являются более глобальными, более сложными.

Далее отметим, что системным является само мышление. Успешное решение поставленной задачи зависит от того, насколько системно подходит специалист к ее анализу. Неудачи в решении тех или иных проблем связаны с отходом от системности, с игнорированием части существенных взаимосвязей компонентов системы. Разрешение возникшей проблемы осуществляется путем перехода на новый, более высокий уровень системности. В связи с этим можно отметить, что системность не столько состояние, сколько процесс.

Свойство системности присуще процессу познания. Системны знания, накопленные человечеством. В качестве особенности процесса познания отметим наличие аналитического и синтетического образов мышления. *Анализ* – это процесс, состоящий в разделении целого на части, в представлении сложного в виде совокупности более простых компонент, но чтобы познать целое, сложное, необходим и обратный процесс – *синтез*. Это относится как к индивидуальному мышлению, так и к общечеловеческому знанию.

Аналитичность человеческого знания находит свое отражение в существовании различных наук, в продолжающейся их дифференциации, во все более глубоком изучении все более узких вопросов. Вместе с тем мы наблюдаем и обратный процесс синтеза знаний. Процесс синтеза проявляется в возникновении междисциплинарных наук, таких как физическая химия, биофизика, биохимия и т. п. Наконец, наиболее высокая форма синтеза знаний реализуется в виде наук о самых общих свойствах природы. К числу таких синтетических наук относится, в первую очередь, философия, которая выявляет и отражает общие свойства всех форм существования материи. К синтетическим можно отнести математику – дисциплину, изучающую всеобщие отношения, взаимосвязи и взаимодействия объектов, а также и системные науки: кибернетику, теорию систем, теорию организации и т. п. В этих дисциплинах органическим образом соединяются технические, естественнонаучные и гуманитарные знания. В качестве методологического подхода к анализу явлений и процессов с точки зрения их системности развился диалектический метод. Именно диалектический метод рассматривает объект как комплекс взаимодействующих и взаимосвязанных компонентов, развивающихся во времени. «Диалектика является методом познания, обеспечивающим согласование системности знаний и системности мира на любом уровне абстракции» [1].

Свойство системности присуще результатам познания. В технических науках это реализуется в построении адекватных моделей, являющихся отражением исследуемых объектов, моделей, описывающих динамическое поведение материальных объектов.

Системна также среда, окружающая человека. Свойство системности является естественным свойством природы. Как уже отмечалось, окружающий нас мир есть бесконечная система систем, иерархическая организация все более сложных объектов. Причем как в живой, так и неживой природе действуют свои законы организации, являющиеся объективными биологическими или физическими законами.

Системно человеческое общество в целом. Системность человеческого общества выражается опять же во взаимосвязи развития отдельных структур (национальных, государственных, религиозных образований) и в их взаимном влиянии друг на друга. Причем следует отметить, что уровень системности человеческого общества постоянно увеличивается. Системность необходимо, таким образом, рассматривать в историческом аспекте. Если в Древнем мире племена жили достаточно отдаленно друг от друга и уровень общения между ними был минимален, то в современном обществе события, происходящие в одних государствах, находят отклик в различных частях мира и имеют на них влияние.

Системны взаимодействия человека со средой. В данном аспекте системность выражается в необходимости комплексного учета всех особенностей и возможных воздействий факторов внешней среды на ее состояние в последующие моменты. В случае недостаточной проработки данных вопросов, игнорирования ряда факторов, наблюдается возникновение проблемы в развитии природы, негативное воздействие на хозяйственную и культурную деятельность человека. Примеров тому можно привести множество. Скажем, строительство гидроэлектростанций в равнинной части континента привело к заболачиванию мест, выводу земель из севооборота, нарушению экологической ситуации в данном регионе, а в некоторых случаях – к изменению климата. Применение различных химикатов ненадлежащего качества и в необоснованном количестве вызвало непоправимые последствия в развитии региона Аральского моря. Примеры такого плана можно продолжать и продолжать. Таким образом, можно сделать вывод, что игнорирование системности взаимодействия человека со средой приводит к возникновению проблемы в развитии среды обитания и соответственно во взаимодействии природы и общества.

1.2. Развитие системных представлений. Становление системного анализа

С позиций современных научных представлений системность всегда была методом любой науки. Возможно, что принципы системности применялись не всегда осознанно, но, тем не менее, любой ученый прошлого, который и не помышлял о системном подходе, так или иначе имел дело с системами и моделями объектов или процессов. Ранее всего системные проблемы были осознаны философами. Следует отметить, что обсуждение системных проблем в таких дисциплинах как философия, логика, математика осуществлялось еще древними учеными. Однако для нас представляет особый интерес развитие системных представлений в применении к системным и техническим дисциплинам.

Первым в явной форме вопрос о научном подходе к управлению сложными системами поставил М.-А. Ампер. Он впервые выделил кибернетику как специальную науку об управлении государством, обозначил ее место в ряду других наук и сформулировал ее системные особенности. Идеи системности применительно к управлению государством развивались в работах польского ученого Б. Трентовского. Он отмечал, что действительно эффективное управление должно учитывать все важнейшие внешние и внутренние факторы, влияющие на объект управления. В своих работах Трентовский пишет, что при выработке управляющего воздействия необходимо учитывать национальные особенности населения с учетом временного аспекта, при одной и той же политической идеологии кибернет (в современной терминологии, лицо, принимающее решение) должен управлять различно в Австрии, России или Пруссии, точно так же и в одной и той же стране он должен управлять завтра иначе, чем сегодня. Трентовский рассматривает общество как систему, которая развивается путем разрешения противоречий. И все-таки общество середины 19-го столетия было не готово к восприятию системных представлений. Прошло еще более полувека, прежде чем системная проблематика прочно заняла свое место в научных публикациях. К числу основоположников теории систем можно заслуженно отнести российского ученого, академика Е.С. Федорова. Основные научные результаты были достигнуты им в области минералогии. Он установил, что существует только 230 типов кристаллической решетки, тем не менее, любое вещество при определенных условиях может кристаллизоваться. Таким образом, было показано, что великое многообразие кристаллов и минералов использует для своего строения ограниченное количество типов структур. Далее им были отмечены аналогичные закономерности в области архитектурных и музыкальных

конструкций, языковых построений, строения вещества и ряда других систем. Развивая системные представления Федоров установил ряд других закономерностей развития систем, в частности, им было замечено такое свойство систем как самоорганизация, способность к приспособлению, к повышению стройности.

Следующим этапом в развитии системных представлений явились работы А.А. Богданова, который в начале XX в. начал создавать теорию организации (тектологию) [4]. Основная идея теории Богданова заключается в том, что все существующие объекты и процессы имеют определенный уровень организованности, который тем выше, чем сильнее свойства целого отличаются от простой суммы свойств комплектующих элементов. Именно анализ свойств целого и его частей был впоследствии заложен в качестве основной характеристики понятия сложной системы. Заслугой Богданова явилось также то, что он изучает не только статическое состояние структур, а занимается исследованием динамического поведения объектов, уделяет внимание вопросам развития организации, подчеркивает значение обратных связей, указывает на необходимость учета собственных целей организации, отмечает роль открытых систем. Он подчеркивает роль моделирования и математических методов как потенциальных методов решения задач теории организации.

Позднее идеи теории организации развивались в трудах выдающихся представителей отечественного естествознания И.И. Шмальгаузена, В.Н. Беклемишева и ряда других специалистов, вклад которых во многих отношениях явился решающим в формировании вышеназванной теории.

Вклад русских и советских исследователей в развитие теории систем и формирование системных представлений явился определяющим, поскольку большинство развиваемых ныне идей связано с работами Богданова и трудами его последователей. Однако нельзя не отметить также и зарубежных ученых, работы которых являются основополагающими в области развития теории систем и системного анализа. В первую очередь следует обратить внимание на труды австрийского ученого Л. фон Берталанфи, который в 50-х гг. XX в. организовал в Канаде центр системных исследований. Он опубликовал большое количество работ (например [5]), в которых исследовал взаимодействие систем с окружающей средой. Подчеркнуто большое значение обмена системы веществом, энергией и энтропией с внешним миром, отмечено, что в системе устанавливается динамическое равновесие, которое может быть направлено в сторону усложнения организации, функционирование системы является не просто откликом на изменение внешних условий, а сохранением старого или установлением нового внутреннего равновесия системы. В своих работах Берталанфи исследует общие закономерности, присущие любым достаточно сложным организациям материи как биологической, так и общественной природы. Берталанфи и организованная им школа последователей в своих трудах пытаются придать общей теории систем формальный характер.

Массовое распространение системных представлений, осознание системности мира, общества и человеческой деятельности связано с именем американского математика Н. Винера. В 1948 г. он опубликовал книгу «Кибернетика» [6] и далее «Кибернетика и общество» [7]. В своих трудах он развивает идеи управления и связи в животном мире и машинах, анализирует с позиций кибернетики процессы, происходящие в обществе. Н.Винером и его последователями было указано, что предметом кибернетики является исследование систем. Причем отмечается, что хотя при изучении системы на каком-то этапе потребуется проводить учет ее конкретных свойств, для кибернетики в принципе несущественно, какова природа системы. То есть для изучения систем различных типов, будь она физической, биологической, экономической, организационной или вовсе представленной в виде модели, кибернетика предлагает единые подходы к ее исследованию. Ф.И. Перегудов и Ф.П. Тарасенко в своей книге отмечают, что с кибернетикой Винера связаны такие продвижения в развитии системных представлений как типизация моделей систем, выявление особого значения обратных связей в системе, подчеркивание принципа оптимальности в управлении и синтезе систем, осознание информации как всеобщего свойства материи и возможности ее количественного описания, развитие методологии моделирования вообще и в особенности идеи математического эксперимента с помощью ЭВМ.

Существенное место в развитии кибернетики занимают советские ученые. Можно отметить многочисленные работы академика А.И. Берга. Фундаментальный вклад в развитие кибернетики внес также академик А.Н. Колмогоров. Так в период, когда в Советском Союзе кибернетику считали лженаукой и в стране шли горячие дискуссии о сути кибернетики, были сформулированы достаточно общие и полные определения кибернетики. Приведем эти определения: «Кибернетика – это наука об оптимальном управлении сложными динамическими системами» (А.И. Берг); «Кибернетика – это наука о системах, воспринимающих, хранящих, перерабатывающих и использующих информацию» (А.Н. Колмогоров).

Наконец, отметим достижения в области исследования систем бельгийской школы во главе с И. Пригожиным. Ученые этой школы исследовали механизм самоорганизации систем. Они отмечают, что в результате взаимодействия с окружающей средой система может перейти в неравновесное состояние. В результате такого взаимодействия изменяется организованность системы. Переломные точки, в которых наблюдается неустойчивость неравновесных состояний, называются точками бифуркации. Таким образом, согласно теории И. Пригожина [8], материя не является пассивной субстанцией, ей присуща спонтанная активность.

1.3. Определения системного анализа

Системный анализ как дисциплина сформировался в результате возникновения необходимости исследовать и проектировать сложные системы, управлять ими в условиях неполноты информации, ограниченности ресурсов и дефицита времени. Системный анализ является дальнейшим развитием целого ряда дисциплин, таких как исследование операций, теория оптимального управления, теория принятия решений, экспертный анализ, теория организации эксплуатации систем и т.д. Для успешного решения поставленных задач системный анализ использует всю совокупность формальных и неформальных процедур. Перечисленные теоретические дисциплины являются базой и методологической основой системного анализа. Таким образом, *системный анализ – междисциплинарный курс, обобщающий методологию исследования сложных технических, природных и социальных систем* [1]. Широкое распространение идей и методов системного анализа, а главное – успешное их применение на практике стало возможным только с внедрением и повсеместным использованием ЭВМ. Именно применение ЭВМ как инструмента решения сложных задач позволило перейти от построения теоретических моделей систем к широкому их практическому применению. В связи с этим Н.Н. Моисеев пишет [9], что *системный анализ – это совокупность методов, основанных на использовании ЭВМ и ориентированных на исследование сложных систем – технических, экономических, экологических и т.д.* Центральной проблемой системного анализа является проблема принятия решения. Применительно к задачам исследования, проектирования и управления сложными системами проблема принятия решения связана с выбором определенной альтернативы в условиях различного рода неопределенности. Неопределенность обусловлена многокритериальностью задач оптимизации, неопределенностью целей развития систем, неоднозначностью сценариев развития системы, недостаточностью априорной информации о системе, воздействием случайных факторов в ходе динамического развития системы и прочими условиями. Учитывая данные обстоятельства, *системный анализ можно определить как дисциплину, занимающуюся проблемами принятия решений в условиях, когда выбор альтернативы требует анализа сложной информации различной физической природы.*

Главным содержанием дисциплины «Системный анализ» являются сложные проблемы принятия решений, при изучении которых неформальные процедуры, представления здравого смысла и способы описания ситуаций играют не меньшую роль, чем формальный математический аппарат.

Системный анализ является дисциплиной синтетической. В нем можно выделить три главных направления. Эти три направления соответствуют трем этапам, которые всегда присутствуют в исследовании сложных систем:

1) построение модели исследуемого объекта;
2) постановка задачи исследования;
3) решение поставленной математической задачи.

Рассмотрим данные этапы. **Построение модели** (формализация изучаемой системы, процесса или явления) есть описание процесса на языке математики. При построении модели осуществляется математическое описание явлений и процессов, происходящих в системе. Поскольку знание всегда относительно, описание на любом языке отражает лишь некоторые стороны происходящих процессов и никогда не является абсолютно полным. С другой стороны, следует отметить, что при построении модели необходимо уделять основное внимание тем сторонам изучаемого процесса, которые интересуют исследователя. Глубоко ошибочным является желание при построении модели системы отразить все стороны существования системы. При проведении системного анализа, как правило, интересуются динамическим поведением системы, причем при описании динамики с точки зрения проводимого исследования есть первостепенные параметры и взаимодействия, а есть несущественные в данном исследовании параметры. Таким образом, качество модели определяется соответствием выполненного описания тем требованиям, которые предъявляются к исследованию, соответствием получаемых с помощью модели результатов ходу наблюдаемого процесса или явления. Построение математической модели есть основа всего системного анализа, центральный этап исследования или проектирования любой системы. От качества модели зависит результат всего системного анализа.

Постановка задачи исследования. На данном этапе формулируется цель анализа. Цель исследования предполагается внешним фактором по отношению к системе. Таким образом, цель становится самостоятельным объектом исследования. Цель должна быть формализована. Задача системного анализа состоит в проведении необходимого анализа неопределенностей, ограничений и формулировании, в конечном счете, некоторой оптимизационной задачи:

$$f(x) \to \max,\ x \subset G. \tag{1.1}$$

Здесь x – элемент некоторого нормированного пространства G, определяемого природой модели, $G \subset E$, где E – множество, которое может иметь сколь угодно сложную природу, определяемую структурой модели и особенностями исследуемой системы. Таким образом, задача системного анализа на этом этапе трактуется как некоторая оптимизационная проблема. Анализируя требования к системе, т. е. цели, которые предполагает достигнуть исследователь, и те неопределенности, которые при этом неизбежно присутствуют, исследователь должен сформулировать цель анализа на языке математики. Язык оптимизации оказывается здесь естественным и удобным, но вовсе не единственно возможным.

Решение поставленной математической задачи. Только этот третий этап анализа можно отнести собственно к этапу, использующему в полной степени математические методы. Хотя без знания математики и возможностей ее аппарата успешное выполнение двух первых этапов невозможно, так как и при построении модели системы, и при формулировании цели и задач анализа широкое применение должны находить методы формализации. Однако отметим, что именно на завершающем этапе системного анализа могут потребоваться тонкие математические методы. Но следует иметь в виду, что задачи системного анализа могут иметь ряд особенностей, которые приводят к необходимости применения наряду с формальными процедурами эвристических подходов. Причины, по которым обращаются к эвристическим методам, в первую очередь связаны с недостатком априорной информации о процессах, происходящих в анализируемой системе. Также к таковым причинам можно отнести большую размерность вектора $\mathbf{x}$ и сложность структуры множества G. В данном случае трудности, возникающие в результате необходимости применения неформальных процедур анализа зачастую являются определяющими. Успешное решение задач системного анализа требует использования на каждом этапе исследования неформальных рассуждений. Ввиду этого проверка качества решения, его соответствие исходной цели исследования превращается в важнейшую теоретическую проблему.

1.4. Понятие сложной системы

Определение системы

Объектом изучения системного анализа являются сложные системы. Понятие системы стало широко использоваться в XX в. Длительное время оно применялось в самом общем смысле. Не было строгого формализованного определения данного понятия. По мере развития дисциплин кибернетического направления и особенно в связи с развитием и внедрением в различные сферы человеческой деятельности вычислительной техники появилась необходимость формализовать понятие сложной системы, попытаться дать его строгое определение.

В повседневной жизни термин система используют в тех случаях, когда хотят охарактеризовать объект как нечто целое, сложное, о чем невозможно сразу дать представление. Предполагается, что для характеристики системы необходимо рассмотреть различные аспекты ее функционирования, проанализировать различные ее свойства. Отметим сразу, что в литературе встречается большое количество определений сложной системы. Все они отражают те или иные важные стороны данного объекта. Приведем ряд определений и проанализируем их. В «Философском словаре» *система* определяется как «совокупность элементов, находящихся в определенных отношениях и связях между собой и образующих некоторое целостное единство». Ю.И. Дегтярев [10] определяет систему следующим образом: «*Системой* называется упорядоченная совокупность материальных объектов (элементов), объединенных какими-либо связями (механическими, информационными), предназначенных для достижения определенной цели и достигающих ее наилучшим (по возможности) образом». В данном определении выделяется три основных компонента системы – элементы, связи и операции. Важной особенностью системы является то, что она создается или функционирует (если это естественная, а не искусственная система) для достижения определенной цели. То есть в результате динамического поведения системы решаются какие-то определенные задачи, которые в конечном итоге приводят к достижению глобальной цели функционирования или развития системы. Авторы монографии [1] определяют систему следующим образом : «*система есть средство достижения цели*» и «*система есть совокупность взаимосвязанных элементов, обособленная от среды и взаимодействующая с ней как целое*». Естественно, что эти два определения необходимо рассматривать в совокупности, так как они дополняют друг друга и в каждом из них акцент делается на определенные свойства системы.

Наибольший вклад в формализацию представлений о сложных системах был сделан в связи с развитием автоматизированных систем управления. Авторы работ по теории систем [11] применительно к техническим системам понятие системы формулируют в виде следующих определений.

Под автоматизированной системой понимается программно-аппаратный комплекс, выполненный на базе средств измерительной и вычислительной техники, предназначенный для решения задач управления на основе получения и использования моделей объекта управления. В данном определении констатируется, что автоматизированная система является искусственной системой, создаваемой человеком. Для таких систем конечное состояние или цель функционирования задается заранее, а их поведение направлено на достижение поставленной цели. Цель автоматизированной системы состоит в решении выделенного набора задач автоматизации управления, как правило, поведением технического объекта.

Автоматизированная система – это совокупность частей (технических средств, математических методов, коллектива исполнителей), образующая организационное комплексное единое целое и обеспечивающая решение требуемого набора задач автоматизации с заданной точностью в пределах ограничений во времени и стоимости. В данном определении уточняется состав элементов, из которых строится система. Также отмечается, что разработка и функционирование системы должны производиться с учетом некоторых ограничений. Иными словами к системе предъявляются определенные требования оптимальности.

Логичным кажется не искать в литературе всеобъемлющего определения сложной системы, а указать на основные свойства системы, которые всесторонне характеризуют ее и так или иначе присутствуют в различных формулировках определений. Первая существенная особенность системы состоит в том, что система обладает новыми свойствами по сравнению с элементами, из которых она состоит. При этом система есть не просто механический набор элементов, а целенаправленное их соединение в виде определенных структур и взаимосвязей. Система есть организационное единство элементов. Нарушение взаимосвязей приведет к разрушению системы.

Вторая особенность систем состоит в том, что они обладают свойствами оптимальности. Системы проектируются с учетом критериев оптимальности и функционируют согласно построенным заранее оптимальным планам. Следующая черта, которая отражается в определении системы, – это цель или назначение системы. Системы создаются для достижения какой-либо цели, для решения определенных задач. Не существует систем, не предназначенных ни для чего, не решающих никаких задач. Любая система имеет свое предназначение.

Приведенные определения, тем не менее, не дают однозначного толкования, что считать системой, а что нет. Не устанавливают однозначных границ систем. И действительно, система – понятие относительное. На одном уровне иерархии элемент системы сам является системой, на другом уровне система есть элемент более крупной системы. Поэтому определения системы должны дополняться классификациями и уточнениями.

Классификация систем

Подходы к классификации системы могут быть самыми разными:

- по виду отображаемого объекта – технические, биологические, социальные и т. п.;
- по характеру поведения – детерминированные, вероятностные, игровые;
- по типу целеустремленности – открытые и закрытые;
- по сложности структуры и поведения – простые и сложные;
- по виду научного направления, используемого для их моделирования – математические, физические, химические и др.;
- по степени организованности – хорошо организованные, плохо организованные и самоорганизующиеся.

Рассмотрим некоторые из представленных видов классификации. *Детерминированной* называется система, состояние которой в будущем однозначно определяется ее состоянием в настоящий момент времени и законами, описывающими переходы элементов и системы из одних состояний в другие. Составные части в детерминированной системе взаимодействуют точно известным образом. Примером детерминированной системы может служить механический арифмометр. Установка соответствующих чисел на валике и задание порядка вычисления однозначно определяют результат работы устройства. То же самое можно сказать о калькуляторе, если считать его абсолютно надежным.

Вероятностные или *стохастические* системы – это системы, поведение которых описывается законами теории вероятностей. Для вероятностной системы знание текущего состояния и особенностей взаимной связи элементов недостаточно для предсказания будущего поведения системы со всей определенностью. Для такой системы имеется ряд направлений возможных переходов из одних состояний в другие, т. е. имеется группа сценариев преобразования состояний системы, и каждому сценарию поставлена в соответствие своя вероятность. Примером стохастической системы может служить мастерская по ремонту электронной и радиотехники. Срок выполнения заказа по ремонту конкретного изделия зависит от количества аппаратуры, поступившей в ремонт до поступления рассматриваемого изделия, от характера повреждений каждого из находящихся в очереди объектов, от количества и квалификации обслуживающего персонала и т. п.

Игровой является система, осуществляющая разумный выбор своего поведения в будущем. В основе выбора лежат оценки ситуаций и предполагаемых способов действий, выбираемых на основе заранее сформированных критериев, а также с учетом соображений неформального характера. Руководствоваться этими соображениями может только человек. Примером игровой системы может служить организация, выполняющая некоторые работы и выступающая в качестве исполнителя. Исполнитель вступает в отношения с заказчиком. Интересы исполнителя и заказчика противоположные. Исполнитель старается продать свою работу как можно выгоднее. Заказчик, наоборот, пытается сбить цену и соблюсти свои интересы. В данном торге между ними проявляется игровая ситуация.

Классификация по данному признаку условна, как и многое другое, касающееся характеристики сложных систем. Она допускает разные толкования принадлежности той или иной системы к сформированным классам. Так в детерминированной системе можно найти элементы стохастичности. С другой стороны, детерминированную систему можно считать частным случаем стохастической системы, если положить вероятности переходов из состояния в состояние соответственно равными нулю (перехода нет) и единице (переход имеет место). Точно также стохастическую систему можно рассматривать как частный случай игровой, когда идет игра с природой.

Следующий признак классификации: открытые и закрытые системы. По данному признаку классификации системы характеризуются различной степенью взаимодействия с внешней средой. *Открытые* системы обладают особенностью обмениваться с внешней средой массой, энергией, информацией. *Замкнутые* (или *закрытые*) системы изолированы от внешней среды. Предполагается, что разница между открытыми и замкнутыми системами определяется с точностью до принятой чувствительности модели.

По степени сложности системы подразделяются на простые, сложные и очень сложные. *Простые* системы характеризуются небольшим количеством возможных состояний, их поведение легко описывается в рамках той или иной математической модели. *Сложные* системы отличаются разнообразием внутренних связей, но допускают их описание. Причем набор методов, привлекаемых для описания сложных систем, как правило, многообразен, т. е. для построения математической модели сложной системы применяются различные подходы и разные разделы математики. *Очень сложные* системы характеризуются большой разветвленностью связей и своеобразностью отношений между элементами. Многообразие связей и отношений таково, что нет возможности все их выявить и проанализировать. Простыми системами можно считать лентопротяжные механизмы, механические передачи, системы слежения за целью и т.д. Сложными системами являются электронно-вычислительная машина, система управления и защиты энергоблока, система электроснабжения промышленного объекта и пр. Очень сложными являются социотехнические системы, такие как автоматизированные системы управления крупным предприятием, экспертные системы с функциями поддержки и принятия управленческих решений.

Классификация по признаку организованности систем впервые была предложена В.В. Налимовым [12]. Под *хорошо организованной* системой понимается система, у которой определены все элементы, их взаимосвязь, правила объединения в более крупные компоненты, связи между всеми компонентами и целями системы, ради достижения которых создается или функционирует система. При этом подразумевается, что все элементы системы с их взаимосвязями между собой, а также с целями системы можно отобразить в виде аналитических зависимостей. При формулировании задачи принятия решения для хорошо организованной системы проблемная ситуация описывается в виде математического выражения, критерия эффективности, критерия функционирования системы, который может быть представлен сложным уравнением, системой уравнений, сложными математическими моделями, включающими в себя и уравнения, и неравенства, и т. п. Важно, что решение задачи при представлении ее в виде хорошо организованной системы осуществляется аналитическими методами с использованием моделей формализованного представления системы. Примером хорошо организованной системы может служить сложное электронное устройство. Описание его работы производят с помощью системы уравнений, учитывающих условия функционирования, в том числе наличие шумов, нестабильность электропитания и т.д.

При представлении объекта в виде *плохо организованной* системы не ставится задача определить все учитываемые компоненты, их свойства и связи между собой, а также с целями системы. Для плохо организованной системы формируется набор макропараметров и функциональных закономерностей, которые будут ее характеризовать. Определение этих параметров и восстановление функциональных зависимостей осуществляется на основании некоторой выборочной информации, характеризующей исследуемый объект или процесс. Далее полученные оценки характеристик распространяют на поведение системы в целом. При этом предполагается, что полученный результат обладает ограниченной достоверностью и его можно использовать с некоторыми оговорками. Так, например, если результат получен на основании статистических наблюдений за функционированием системы на ограниченном интервале времени, т. е. на основании выборочных наблюдений, то его можно использовать с некоторой доверительной вероятностью. Примером применения подхода к отображению объектов в виде плохо организованной системы можно считать оценивание характеристик надежности системы с множеством компонентов. В данном случае характеристики надежности группы однотипных элементов определяются на основании выборочной информации, полученной в результате наблюдений за их работой на ограниченном отрезке времени при определенных уровнях воздействующих факторов. Затем полученные оценки распространяются на весь период эксплуатации объекта. Данные оценки используются при проведении расчетов характеристик надежности всей системы.

Самоорганизующиеся системы – это системы, обладающие свойством адаптации к изменению условий внешней среды, способные изменять структуру при взаимодействии системы со средой, сохраняя при этом свойства целостности, системы, способные формировать возможные варианты поведения и выбирать из них наилучшие. Эти особенности обусловлены наличием в структуре системы активных элементов, которые, с одной стороны, обеспечивают возможность адаптации, приспособления системы к новым условиям существования, с другой стороны, вносят элемент неопределенности в поведение системы, чем затрудняют проведение анализа системы, построение ее модели, формальное ее описание и, в конечном счете, затрудняют управление такими системами. Примерами самоорганизующихся систем могут служить биологические системы, предприятия и их система управления, городские структуры управления и т.д.

1.5. Характеристика задач системного анализа

Системный анализ в настоящее время вынесен на самое острие научных исследований. Он призван дать научный аппарат для анализа и изучения сложных систем. Лидирующая роль системного анализа обусловлена тем, что развитие науки привело к постановке тех задач, которые призван решать системный анализ. Особенность текущего этапа состоит в том, что системный анализ, еще не успев сформироваться в полноценную научную дисциплину, вынужден существовать и развиваться в условиях, когда общество начинает ощущать потребность в применении еще недостаточно разработанных и апробированных методов и результатов и не в состоянии отложить решение связанных с ними задач на завтра. В этом источник как силы, так и слабости системного анализа: силы – потому, что он постоянно ощущает воздействие потребности практики, вынужден непрерывно расширять круг объектов исследования и не имеет возможности абстрагироваться от реальных потребностей общества; слабости – потому, что нередко применение «сырых», недостаточно проработанных методов системных исследований ведет к принятию скороспелых решений, пренебрежению реальными трудностями.

Рассмотрим основные задачи, на решение которых направлены усилия специалистов и которые нуждаются в дальнейшей разработке. Во-первых, следует отметить задачи исследования системы взаимодействий анализируемых объектов с окружающей средой. Решение данной задачи предполагает:

• проведение границы между исследуемой системой и окружающей средой, предопределяющей предельную глубину влияния рассматриваемых взаимодействий, которыми ограничивается рассмотрение;

• определение реальных ресурсов такого взаимодействия;

• рассмотрение взаимодействий исследуемой системы с системой более высокого уровня.

Задачи следующего типа связаны с конструированием альтернатив этого взаимодействия, альтернатив развития системы во времени и в пространстве. Важное направление развития методов системного анализа связано с попытками создания новых возможностей конструирования оригинальных альтернатив решения, неожиданных стратегий, непривычных представлений и скрытых структур. Другими словами, речь здесь идет о разработке методов и средств усиления индуктивных возможностей человеческого мышления в отличие от его дедуктивных возможностей, на усиление которых, по сути дела, направлена разработка формальных логических средств. Исследования в этом направлении начаты лишь совсем недавно, и единый концептуальный аппарат в них пока отсутствует. Тем не менее, и здесь можно выделить несколько важных направлений – таких, как разработка формального аппарата индуктивной логики, методов морфологического анализа и других структурно-синтаксических методов конструирования новых

альтернатив, методов синтектики и организации группового взаимодействия при решении творческих задач, а также изучение основных парадигм поискового мышления.

Задачи третьего типа заключаются в конструировании множества имитационных моделей, описывающих влияние того или иного взаимодействия на поведение объекта исследования. Отметим, что в системных исследованиях не преследуется цель создания некоей супермодели. Речь идет о разработке частных моделей, каждая из которых решает свои специфические вопросы.

Даже после того как подобные имитационные модели созданы и исследованы, вопрос о сведении различных аспектов поведения системы в некую единую схему остается открытым. Однако решить его можно и нужно не посредством построения супермодели, а анализируя реакции на наблюдаемое поведение других взаимодействующих объектов, т. е. путем исследования поведения объектов-аналогов и перенесения результатов этих исследований на объект системного анализа. Такое исследование дает основание для содержательного понимания ситуаций взаимодействия и структуры взаимосвязей, определяющих место исследуемой системы в структуре суперсистемы, компонентом которой она является.

Задачи четвертого типа связаны с конструированием моделей принятия решений. Всякое системное исследование связано с исследованием различных альтернатив развития системы. Задача системных аналитиков выбрать и обосновать наилучшую альтернативу развития. На этапе выработки и принятия решений необходимо учитывать взаимодействие системы с ее подсистемами, сочетать цели системы с целями подсистем, выделять глобальные и второстепенные цели.

Наиболее развитая и в то же время наиболее специфическая область научного творчества связана с развитием теории принятия решений и формированием целевых структур, программ и планов. Здесь не ощущается недостатка и в работах, и в активно работающих исследователях. Однако и в данном случае слишком многие результаты находятся на уровне неподтвержденного изобретательства и разночтений в понимании как существа стоящих задач, так и средств их решения. Исследования в этой области включают:

а) построение теории оценки эффективности принятых решений или сформированных планов и программ;

б) решение проблемы многокритериальности в оценках альтернатив решения или планирования;

в) исследования проблемы неопределенности, особенно связанной не с факторами статистического характера, а с неопределенностью экспертных суждений и преднамеренно создаваемой неопределенностью, связанной с упрощением представлений о поведении системы;

г) разработка проблемы агрегирования индивидуальных предпочтений на решениях, затрагивающих интересы нескольких сторон, которые влияют на поведение системы;

д) изучение специфических особенностей социально-экономических критериев эффективности;

е) создание методов проверки логической согласованности целевых структур и планов и установления необходимого баланса между предопределенностью программы действий и ее подготовленностью к перестройке при поступлении новой информации как о внешних событиях, так и изменении представлений о выполнении этой программы.

Для последнего направления требуется новое осознание реальных функций целевых структур, планов, программ и определение тех, которые они *должны* выполнять, а также связей между ними.

Рассмотренные задачи системного анализа не охватывают полного перечня задач. Здесь перечислены те, которые представляют наибольшую сложность при их решении. Следует отметить, что все задачи системных исследований тесно взаимосвязаны друг с другом, не могут быть изолированы и решаться отдельно как по времени, так и по составу исполнителей. Более того, чтобы решать все эти задачи, исследователь должен обладать широким кругозором и владеть богатым арсеналом методов и средств научного исследования.

1.6. Особенности задач системного анализа

Конечной целью системного анализа является разрешение проблемной ситуации, возникшей перед объектом проводимого системного исследования (обычно это конкретная организация, коллектив, предприятие, отдельный регион, социальная структура и т. п.). Системный анализ занимается изучением проблемной ситуации, выяснением ее причин, выработкой вариантов ее устранения, принятием решения и организацией дальнейшего функционирования системы, разрешающего проблемную ситуацию. Начальным этапом любого системного исследования является изучение объекта проводимого системного анализа с последующей его формализацией. На этом этапе возникают задачи, в корне отличающие методологию системных исследований от методологии других дисциплин, а именно, в системном анализе решается двуединая задача. С одной стороны, необходимо формализовать объект системного исследования, с другой стороны, формализации подлежит

процесс исследования системы, процесс постановки и решения проблемы. Приведем пример из теории проектирования систем. Современная теория автоматизированного проектирования сложных систем может рассматриваться как одна из частей системных исследований. Согласно ей проблема проектирования сложных систем имеет два аспекта. Во-первых, требуется осуществить формализованное описание объекта проектирования. Причем на этом этапе решаются задачи формализованного описания как статической составляющей системы (в основном формализации подлежит ее структурная организация), так и ее поведение во времени (динамические аспекты, которые отражают ее функционирование). Во-вторых, требуется формализовать процесс проектирования. Составными частями процесса проектирования являются методы формирования различных проектных решений, методы их инженерного анализа и методы принятия решений по выбору наилучших вариантов реализации системы.

Важное место в процедурах системного анализа занимает проблема принятия решения. В качестве особенности задач, встающих перед системными аналитиками, необходимо отметить требование оптимальности принимаемых решений. В настоящее время приходится решать задачи оптимального управления сложными системами, оптимального проектирования систем, включающих в себя большое количество элементов и подсистем. Развитие техники достигло такого уровня, при котором создание просто работоспособной конструкции само по себе уже не всегда удовлетворяет ведущие отрасли промышленности. Необходимо в ходе проектирования обеспечить наилучшие показатели по ряду характеристик новых изделий, например, добиться максимального быстродействия, минимальных габаритов, стоимости и т. п. при сохранении всех остальных требований в заданных пределах. Таким образом, практика предъявляет требования разработки не просто работоспособного изделия, объекта, системы, а создания оптимального проекта. Аналогичные рассуждения справедливы и для других видов деятельности. При организации функционирования предприятия формулируются требования по максимизации эффективности его деятельности, надежности работы оборудования, оптимизации стратегий обслуживания систем, распределения ресурсов и т.п.

В различных областях практической деятельности (технике, экономике, социальных науках, психологии) возникают ситуации, когда требуется принимать решения, для которых не удается полностью учесть предопределяющие их условия. Принятие решения в таком случае будет происходить в условиях неопределенности, которая имеет различную природу. Один из простейших видов неопределенности – неопределенность исходной информации, проявляющаяся в различных аспектах. В первую очередь, отметим такой аспект, как воздействие на систему неизвестных факторов. Приведем примеры, поясняющие данный тип неопределенности. Проектируется дамба, которая должна защитить населенные пункты от селевых потоков. Ни моменты наступления неблагоприятных событий, ни размеры их заранее неизвестны. Тем не менее, строить защитные сооружения необходимо и необходимо принимать решения об их размерах. Причем лицо, принимающее решение, должно понимать уровень ответственности, которая на него ложится. Строительство слишком массивных конструкций потребует необоснованно больших материальных затрат. С другой стороны, экономия в этом вопросе в случае наступления паводков или селевых потоков может повлечь за собой несоизмеримые экономические убытки, а нередко и человеческие жертвы.

Другой пример приведем из области организации функционирования предприятия легкой промышленности. Предприятие планирует ассортимент товаров на будущий календарный период. Задача предприятия состоит в максимизации прибыли после реализации произведенного товара. Однако заранее неизвестно, какой товар будет пользоваться максимальным спросом, так как спрос определяется многими факторами, такими как соотношение цены и качества товара, моды, уровня жизни населения и прочими факторами. В условиях неопределенности многих факторов руководство предприятия должно разработать план работы.

Неопределенность, обусловленная неизвестными факторами, также бывает разных видов. Наиболее простой вид такого рода неопределенности – *стохастическая неопределенность*. Она имеет место в тех случаях, когда неизвестные факторы представляют собой случайные величины или случайные функции, статистические характеристики которых могут быть определены на основании анализа прошлого опыта функционирования объекта системных исследований. Пример, поясняющий стохастическую неопределенность, следующий. На предприятии планируются восстановительные мероприятия с целью поддержания оборудования на высоком уровне надежности. К таким мероприятиям относятся плановые профилактические работы, контрольные проверки исправности функционирования, ремонты. Частота и длительность соответствующих мероприятий зависит от надежности оборудования, для которого данные мероприятия разрабатываются. Наработки оборудования до отказа, длительности ремонтов, профилактик, проверок – величины случайные, в общем случае неизвестные. Однако характеристики случайных величин, входящих в задачу, могут быть получены, если организовать сбор соответствующей статистической информации.

Еще раз подчеркнем, что стохастическая неопределенность – одна из самых простых типов неопределенности. Задача исследователя заключается в определении вероятностных характеристик случайных факторов и постановке задачи принятия решения в форме статистической оптимизации. Гораздо хуже обстоит дело, когда неизвестные факторы не могут быть изучены и описаны статистическими методами. Это бывает в двух случаях: 1) когда распределение вероятностей для неизвестных факторов в принципе существует, но к моменту принятия решения не может быть получено; 2) когда распределение вероятностей для неизвестных факторов вообще не существует. Приведем пример, иллюстрирующий первый тип. Пусть планируется система профилактического обслуживания оборудования на вновь строящемся предприятии. В отличие от предыдущего примера у лица, принимающего решение, нет статистических данных о наработках оборудования, поскольку оно еще не работало. А решение принимать надо. В этом случае можно назначить время проведения профилактических работ из разумных субъективных соображений, а по мере накопления информации о работе оборудования скорректировать данное решение, иными словами, в процессе функционирования проводить адаптацию решения с учетом опыта эксплуатации.

Второй тип может быть рассмотрен на следующем примере. При подготовке к поездке в район с резко меняющимися климатическими условиями возникает задача оптимизации гардероба (одежда, зонт, обувь), который необходимо иметь во время поездки. В северных районах Сибири в июне месяце температура воздуха может меняться от –10 до +30°С, при этом возможна ясная погода, дождь различной интенсивности (вплоть до ливневого) и снег. Никакие многолетние наблюдения за погодой в данном регионе не дают прогноза в конкретный период времени. Вероятностного распределения в данном случае просто не существует. В данной ситуации необходимо принимать решения, которые наверняка будут далеки от оптимальных. В таких условиях лучше перестраховаться и быть готовым к самому неблагоприятному стечению обстоятельств.

Следующий вид неопределенности – *неопределенность целей*. Формулирование цели при решении задач системного анализа является одной из ключевых процедур, потому что цель является объектом, определяющим постановку задачи системных исследований. Неопределенность цели является следствием из многокритериальности задач системного анализа. Назначение цели, выбор критерия, формализация цели почти всегда представляют собой трудную проблему. Задачи со многими критериями характерны для крупных технических, хозяйственных, экономических проектов. Скажем, при создании проекта нового транспортного средства конструкторы пытаются добиться, чтобы это средство обладало максимальными скоростями, высокой надежностью, высокими техническими характеристиками, и при этом ставят задачу минимизации затрат. Здесь, во-первых, видно, что при формулировании задачи используется несколько критериев, во-вторых, критерии противоречивы между собой. Строго говоря, задача в сформулированной постановке вообще не имеет решения, поскольку минимум затрат – это их полное отсутствие, т. е. при такой постановке затраты должны быть равны нулю. Но тогда выполнение всего проекта невозможно. Следовательно, необходимо очень тщательно анализировать выдвигаемые критерии и грамотно формулировать цель исследования. Один из возможных путей решения данной проблемы – это постановка задачи на условный экстремум, когда ряд критериев переводят в разряд ограничений. Также возможно переходить к сложным, комбинированным критериям. Формализация цели в системном анализе – труднейшая часть проблемы. Можно заметить, что в системных исследованиях главный момент – формулирование цели, которую должен преследовать проект. Цель становится самостоятельным объектом исследования.

И, наконец, следует отметить такой вид неопределенности как неопределенность, связанная с последующим влиянием результатов принятого решения на проблемную ситуацию. Дело в том, что решение, принимаемое в настоящий момент и реализуемое в некоторой системе, призвано повлиять на функционирование системы. Собственно для того оно и принимается, так как по идее системных аналитиков данное решение должно разрешить проблемную ситуацию. Однако поскольку решение принимается для сложной системы, то развитие системы во времени может иметь множество стратегий. И конечно же на этапе формирования решения и принятия управляющего воздействия аналитики могут не представлять себе полной картины развития ситуации. При принятии решения существуют различные рекомендации прогнозирования развития системы во времени. Один из таких подходов рекомендует прогнозировать некоторую «среднюю» динамику развития системы и принимать решения исходя из такой стратегии. Другой подход рекомендует при принятии решения исходить из возможности реализации самой неблагоприятной ситуации.

В качестве следующей особенности системного анализа отметим роль моделей как средства изучения систем, являющихся объектом системных исследований. Любые методы системного анализа опираются на математическое описание тех или иных фактов, явлений, процессов. Употребляя слово «модель», всегда имеют в виду некоторое описание, отражающее именно те особенности изучаемого процесса, которые и интересуют исследователя. Точность, качество описания определяются, прежде всего, соответствием модели тем требованиям, которые предъявляются к исследованию, соответствием получаемых с помощью модели результатов наблюдаемому ходу процесса. Если при разработке модели используется язык математики, говорят о математических моделях. Построение математической модели является основой всего системного анализа. Это центральный этап исследования или проектирования любой системы. От качества модели зависит успешность всего последующего анализа.

Построение моделей – процедура всегда неформальная, она очень сильно зависит от исследователя, его опыта, научной интуиции, всегда опирается на определенный экспериментальный материал. Модель должна достаточно адекватно отражать описываемое явление и, кроме того, быть удобной для использования. Поэтому определенные требования предъявляются к степени детализации модели. Форма представления модели должна определяться целями исследования.

Обсуждая особенности задач системного анализа, нельзя не остановиться еще на одной из них. Ранее отмечалась большая роль математических методов и процедур при проведении системных исследований. Построение моделей исследуемого объекта, формулирование целей и критериев исследования в большой степени базируются на использовании аналитических методов либо процедур формализации закономерностей развития системы и методов проведения исследований. Однако в системном анализе наряду с формализованными процедурами большое место занимают неформальные, эвристические методы исследования. Этому есть ряд причин. Первая состоит в следующем. При построении моделей систем может иметь место отсутствие или недостаток исходной информации для определения параметров модели. В этом случае проводится экспертный опрос специалистов с целью устранения неопределенности или, по крайней мере, ее уменьшения, т. е. опыт и знания специалистов могут быть использованы для назначения исходных параметров модели.

Еще одна причина применения эвристических методов состоит в следующем. Попытки формализовать процессы, протекающие в исследуемых системах, всегда связаны с формулированием определенных ограничений и упрощений. Здесь важно не перейти ту грань, за которой дальнейшее упрощение приведет к потере сути описываемых явлений. Иными словами, желание приспособить хорошо изученный математический аппарат для описания исследуемых явлений может исказить их суть и привести к неверным решениям. В этой ситуации требуется использовать научную интуицию исследователя, его опыт и умение сформулировать идею решения задачи, т. е. применяется подсознательное, внутреннее обоснование алгоритмов построения модели и методов их исследования, не поддающееся формальному анализу. Эвристические методы поиска решений формируются человеком или группой исследователей в процессе их творческой деятельности. Эвристика – это совокупность знаний, опыта, интеллекта, используемых для получения решений с помощью неформальных правил. Эвристические методы оказываются полезными и даже незаменимыми при исследованиях, имеющих нечисловую природу или отличающихся сложностью, неопределенностью, изменчивостью.

Наверняка при рассмотрении конкретных задач системного анализа можно будет выделить еще какие-то их особенности, но, по мнению автора, отмеченные здесь особенности являются общими для всех задач системных исследований.

1.7. Развитие систем или процессов. Прогнозирование и планирование

В системном анализе имеется большая группа задач, в которых требуется спрогнозировать процессы развития системы и принять решение, в результате которого система в будущем должна попасть в некоторое оптимальное состояние. Например, в экономических системах требуется спланировать ассортимент и объем выпускаемой продукции в некоторый будущий период времени с целью получения максимальной прибыли после ее реализации. При этом необходимо спрогнозировать потребность рынка в продукции соответствующего типа, спрос на данный вид продукции. Как уже было отмечено, потребность рынка определяется многими факторами, например, соотношением цены и качества товара, уровнем доходов населения, модой и т.д. При решении задач такого типа на помощь приходит прогностика – наука о способах и методах разработки прогнозов.

Прогнозом называется научно обоснованное суждение об ожидаемых состояниях системы, объекта или явления окружающей действительности.

Прогнозирование – это разработка прогнозов, состоящая в организации и проведении специальных исследований перспектив развития исследуемых объектов, систем или явлений. Научное прогнозирование

чаще всего применяется в социально-экономических и научно-технических областях человеческой деятельности. Процесс прогнозирования базируется на изучении объективных тенденций развития объекта исследования. Разработка общей проблемы предсказания должна основываться на изучении реальных закономерностей развития объекта исследования. Содержание и степень достоверности прогноза определяются информацией о поведении объекта исследования, накопленной до того времени, на который составляется прогноз, закономерностями, выявленными при функционировании объекта исследования, а также опытом, знаниями и научной интуицией специалистов, занимающихся данным видом деятельности.

Итак, прогнозирование является необходимым этапом, процессом при проведении перспективного планирования развития систем. Прогнозирование и планирование являются двумя фазами общего процесса управления. Прогнозирование – это генерирование информации о представляющихся возможными будущих состояниях или траекториях развития системы. **Планирование** есть процесс принятия управленческих решений на сравнительно длительные сроки. Принятие планового решения отделено от его реализации достаточно продолжительным периодом.

Прогнозирование и планирование – обращенные в будущее виды управленческой деятельности. Рассмотрим соотношение данных видов работ. Во-первых, прогноз помогает найти проблемные ситуации, решение которых может быть предусмотрено при выполнении планирования, тем самым способствуя уточнению процедур, выполняемых на этапе планирования. Во-вторых, прогноз позволяет охарактеризовать множество допустимых траекторий решения, вокруг одной из которых в дальнейшем будет происходить разработка плана. Благодаря процессу прогнозирования появляется возможность зафиксировать ряд ограничений, имеющих существенное значение при выполнении операции планирования. В-третьих, прогноз позволяет предвидеть некоторые косвенные последствия реализации плана, характеристика которых не является составной частью последнего. Тем самым предоставляется возможность оценить вариант планового решения или его отдельные компоненты по весьма разнообразному кругу факторов (технических, экономических, экологических, социальных) с точки зрения реалистичности, устойчивости, эффективности и т.д.

Одной из проблемных задач системного анализа является задача стратегического управления, цель которой состоит в выборе стратегии развития анализируемой системы, а также модификация общей структуры управления в соответствии с проблемами, вытекающими из необходимости адекватного разрешения объективных структурных противоречий между интересами различных звеньев системы. Стратегическое управление направлено на разрешение противоречий между различными звеньями системы. Задача прогнозирования состоит в разработке стратегий развития системы и количественной оценке возможностей реализации каждой из них.

Таким образом, первая фаза прогнозирования – выявление или формирование возможных вариантов развития прогнозируемого процесса или явления. На втором этапе производится оценка вероятности реализации отдельных вариантов развития с тем, чтобы на основании соответствующих вероятностных характеристик получить количественные оценки реализуемости возможных траекторий развития прогнозируемых процессов. Процедура вероятностной оценки реализуемости некоторой траектории развития характерна для технических систем. При составлении социально-экономических прогнозов результат прогнозирования сложно описать количественно с помощью вероятностных оценок. Поэтому результат оформляется в виде списка возможных вариантов развития исследуемых процессов. Этот список является перечнем вариантов, каждому из которых сопутствует его содержательное описание – основной результат прогноза. Хотя следует оговориться, что и для систем подобного типа можно организовать сбор статистической информации, на основании которой в дальнейшем построить модель развития системы согласно той или иной траектории. Однако при построении данных моделей возникает сложность, состоящая в необходимости учета поведения индивидуумов и коллективов людей, которые являются неотъемлемой частью систем подобного типа.

Важную роль в процедурах прогнозирования приобретает анализ – он должен определить возможные траектории развития системы, приводящие к заданному конечному состоянию. На первом этапе средствами планирования определяется конечная, желаемая цель развития системы, затем намечаются мероприятия, последовательность которых могла бы в случае их успешной реализации обеспечить достижение этой цели. Такая плановая процедура в той или иной степени сама может использовать результаты ранее выполненных прогностических исследований. Когда проект плана подготовлен, он используется в качестве исходной информации для последующего прогноза – как основа сценария развития системы. Цель данной прогностической процедуры состоит в анализе и характеристике разнообразных последствий реализации разработанного проекта плана. Последующий анализ результатов прогноза позволяет оценить реалистичность и качество проекта плана.

Известны три группы методов прогнозирования, предлагаемых для практического применения, – это методы экстраполяции, методы экспертных оценок и логического моделирования. Методы экстраполяции основаны на аппроксимации результатов, полученных при анализе развития исследуемых процессов, описании полученных данных с помощью математических моделей и дальнейшем расчете моделей для будущих моментов времени. Данные методы позволяют осуществлять поиск приемлемых оценок состояний системы в будущем, однако их применение обоснованно только для описания процессов эволюционного развития. Хотя хорошо известно, что процессы развития систем включают в себя как периоды эволюционного изменения, так и скачкообразные переходы от одних состояний к другим. Скачкообразные переходы обусловлены открытиями новых физических принципов, реализацией оригинальных технических решений, осуществлением крупных проектов. Это обстоятельство должно учитываться при проведении исследований, для чего используются различного рода приемы, позволяющие выделить ожидаемые скачки на общем фоне изменений, интересующих исследователя. Рекомендуемым методом прогнозирования скачкообразных изменений развития систем может быть метод экспертных оценок. При этом в качестве экспертов должны выступать высококвалифицированные в данной области знаний специалисты. Необходимая для формирования прогноза информация обобщается путем обработки мнений экспертов. В результате разрабатывается сценарий развития системы, а также возможные его варианты, учитывающие наличие скачков, предсказание которых особенно ценно для системных аналитиков.

Приведем примеры наличия скачкообразных переходов в развитии систем. Рассмотрим историю развития транспорта. Вплоть до конца девятнадцатого века основным средством передвижения был гужевой транспорт. Развитие промышленности и связанная с данным процессом урбанизация привели к необходимости развития транспортного хозяйства больших городов. В этот период времени данная проблема развивается эволюционно, за счет увеличения количества ямщиков с соответствующими средствами управления. Прогнозируется ситуация, согласно которой дальнейшее эволюционное развитие приведет к загрязнению городов продуктами жизнедеятельности лошадей. Но, как известно, данному прогнозу не суждено было сбыться, так как в начале двадцатого века начинается развитие автомобильного транспорта, который решил указанную проблему. В данном случае имеет место типичный скачок в развитии транспортных средств как одной из систем городского хозяйства.

Другой пример касается развития топливно-энергетической отрасли. Рост промышленности и повышение уровня технического вооружения бытовой сферы приводит к необходимости обеспечения увеличивающихся потребностей в энергетических ресурсах. Прогнозы, построенные на основе моделей эволюционного развития, показывают, что в рамках применяемых технологий с определенного момента времени необходимо все больше и больше вводить в действие новые энергетические мощности, причем, все большая часть ресурсов используется на удовлетворение потребностей самой отрасли, т. е. отрасль начинает работать сама на себя. Следовательно, налицо конфликтная снтуация, которая может быть разрешена путем кардинальных изменений технологии производства и потребления энергетических ресурсов. Именно в это время появляются энергосберегающие технологии, внедряются новые технологические процессы в промышленности, осуществляется миниатюризация средств управления, связи, радиотехнических аппаратов и т.п. Таким образом, скачкообразные процессы развития систем или явлений в большей степени и способны разрешать конфликтные снтуации, возникающие в системных исследованиях. Поэтому при формировании прогнозов необходимо привлекать специалистов-аналитиков высокой квалификации для решения возникающих проблем; кроме того, необходимо анализировать и перерабатывать большое количество информации, публикуемой в периодических изданиях, с целью предугадать возможность появления новых технологий в рассматриваемой области знаний.

Важную роль при составлении прогноза играет умение своевременно распознать приближающийся скачок и снизить тем самым риск принятия ошибочных решений, замедляющих реализацию развития систем по оптимальным траекториям. Формальных правил для реализации такого вида деятельности не существует, и способность распознавания, предугадывания появления новых технологий, путей развития той или иной системы относится к высшим формам эвристической деятельности. Системные аналитики, как правило, располагают обширной информацией, накопленной в различных областях знаний. Проблема состоит в том, чтобы конкретизировать накопленные знания и опыт и применить их к конкретной сфере деятельности, отразить перспективы развития вполне конкретной системы.

Рассмотрим, наконец, последнюю группу методов прогнозирования – методы логического моделирования. Данные методы предполагают построение моделей, в которых проводятся аналогии между различными по своей природе явлениями, анализируются взаимосвязи отдельных процессов, обобщаются данные о процессах различной физической природы, к которым применимы понятия теории подобия. Это позволяет предсказывать нестандартные ситуации в той или иной области деятельности, находить наилучшие решения, учитывать реальные перспективы совершенствования анализируемых систем на основе их подобия другим, более хорошо изученным, процессам.

В заключение необходимо отметить важные проблемы, с которыми сталкивается методология прогнозирования, а именно, оценку точности прогноза и сравнение существующих методов прогнозирования. Проблема состоит не только в разнообразии стратегий развития системы в будущем, но и в уникальности каждой из них. Лишь элементарные события, такие как отказы объектов, появление очередей в системах массового обслуживания, могут рассматриваться как потоки, имеющие средние характеристики (интенсивность потока, математическое ожидание времени пребывания в очереди и т. п.). Сложные события подвержены влиянию большого количества внешних и внутренних факторов и даже при повторениях отличаются от своих аналогов траекторией развития. Все это сказывается на конечных результатах исследований и затрудняет сравнительный анализ методов прогнозирования. С целью анализа точности прогноза необходимо проводить исследования прогностических моделей на чувствительность, оценивать неопределенность моделей, их значимость. Несмотря на ограниченную точность прогнозов и сравнительно небольшой выбор методов исследований прогнозирование является важным средством формирования информации о стратегиях развития исследуемой системы.

1.8. Типовые постановки задач системного анализа

Во второй половине XX в. при решении практических задач стали находить широкое применение математические методы. Они стали использоваться при перспективном и текущем планировании научно-исследовательских работ, проектировании различных объектов, управлении технологическими и производственными процессами, прогнозировании развития социальных и производственных систем, оптимизации маршрутов перевозки грузов. Особенно часто к математическим методам прибегают при решении задач оптимизации функционирования производственных систем, при распределении материальных и трудовых ресурсов и страховых запасов, при выборе местоположения предприятий, исследовании и оценке безопасности функционирования объектов повышенного риска. Постановки перечисленных задач носят оптимизационный характер, где в качестве критериев эффективности применяются различные целевые функции. Как уже было отмечено, особенностью постановок задач системного анализа является то обстоятельство, что наряду со строгим математическим аппаратом применяются эвристические методы, основанные на интуиции исследователя, его опыте в решении задач подобного типа. Рассмотрим некоторые постановки задач, являющиеся типовыми задачами системных исследований.

Задачи распределения ресурсов

Задачи распределения ресурсов возникают, когда существует определенный набор работ или операций, которые необходимо выполнить, а имеющихся в наличии ресурсов для выполнения каждой из них наилучшим образом не хватает. Способы распределения ограниченных ресурсов при выполнении различных операций в системе управления могут быть различными. Для того чтобы решить задачу распределения ресурсов, необходимо сформулировать некоторую систему предпочтений или решающее правило. Такое правило принятия решений по определению объема ресурсов, которые целесообразно выделить для каждого процесса, обычно разрабатывается с учетом оптимизации некоторой целевой функции при ограничениях на объем имеющихся ресурсов и временные характеристики.

В зависимости от условий задачи распределения ресурсов делятся на три класса.

1. Заданы и работы, и ресурсы. Требуется распределить ресурсы между работами таким образом, чтобы максимизировать некоторую меру эффективности (скажем, прибыль) или минимизировать ожидаемые затраты (издержки производства). Например, предприятию установлено производственное задание в рамках оговоренного срока. Известны мощности предприятия. При изготовлении продукции изделия проходят обработку на разных станках. Естественным является ограничение – одновременно на одном станке может обрабатываться только одна единица продукции. Мощности предприятия ограниченны и не позволяют для каждого изделия использовать наилучшую технологию. Требуется выбрать такие способы производства для каждой единицы продукции, чтобы выполнить задание с минимальными затратами.

2. Заданы только наличные ресурсы. Требуется определить, какой состав работ можно выполнить с учетом этих ресурсов, чтобы обеспечить максимум некоторой меры эффективности. Приведем пример. Имеется предприятие с определенными производственными мощностями. Требуется произвести планирование ассортимента и объема выпуска продукции, которые позволили бы максимизировать доход предприятия.

3. Заданы только работы. Необходимо определить, какие ресурсы требуются для того, чтобы минимизировать суммарные издержки. Например, составлено расписание движения автобусов пригородного и междугороднего сообщения на летний период времени. Требуется определить необходимое количество водителей, кондукторов, контролеров и прочего обслуживающего персонала, чтобы выполнить план перевозок с минимальными эксплуатационными затратами.

Рассмотрим более детально две постановки задач, в которых возникает необходимость распределения ресурсов. Задачи в той постановке, в которой они будут сформулированы, решаются при проектировании технических объектов, в том числе автоматизированных систем обработки информации и управления.

Первая задача – **задача составления титульного списка.**

На начальном этапе разработки автоматизированной системы обработки информации и управления встает проблема выбора комплекса задач, подлежащих автоматизации. В техническое задание на систему включают как задачи с известными алгоритмами решения, так и задачи, требующие выполнения специальных исследований, проведения научных разработок. Известно, что проектирование и ввод системы в эксплуатацию осуществляются поэтапно. На первом этапе автоматизации в перечень задач, подлежащих автоматизации, включаются только те задачи, которые имеют алгоритмы для решения. Проблемные задачи должны составлять самостоятельную группу. В проектную разработку их включают только после четкой постановки и определения методов решения и оценки эффективности. Эти задачи подлежат автоматизации на последующих этапах, возможно, уже после ввода системы в эксплуатацию. Так формулируется перечень задач, решаемых на первом этапе автоматизации.

После составления перечня задач, включаемых в первый этап разработки, необходимо оценить требуемый состав ресурсов на их разработку и требуемое время для их внедрения. Если для разработки и внедрения задач первого этапа имеется достаточное количество ресурсов, и время, требуемое на их разработку, не превышает заданного срока ввода первой очереди в эксплуатацию, то занимаются распределением ресурсов по задачам, подлежащим автоматизации. Если же время, требуемое на разработку задач, превышает заданный срок ввода первой очереди в эксплуатацию, возникает проблема составления титульного списка, т. е. возникает необходимость ограничения перечня задач, автоматизируемых на первом этапе. Проблема выбора комплекса задач из сформированного перечня в условиях дефицита времени и ресурсов на разработку всего перечня задач, выполняемых на первом этапе автоматизации, называется задачей составления титульного списка. Таким образом, формулировка задачи будет выглядеть так: требуется сформировать перечень задач, подлежащих автоматизации (титульный список), с учетом имеющихся материальных, временных, трудовых и прочих ресурсов. Данная задача относится ко второму классу задач, когда заданы ресурсы и необходимо сформировать состав работ.

Рассмотрим вторую задачу, возникающую при проектировании систем, – **задачу определения оптимальной очередности разработки.**

Задача определения оптимальной очередности разработки встает перед проектировщиками на следующем этапе проектирования после составления титульного списка задач, подлежащих автоматизации. Суть задачи состоит в распределении ресурсов, выделяемых на разработку системы, между задачами и упорядочении процесса разработки задач во времени. При определении очередности разработки необходимо учитывать одно важное обстоятельство, а именно, зависимость задач друг от друга, т. е. тот факт, что для некоторых задач не может начаться разработка, пока не закончено проектирование задач, от результатов внедрения которых они зависят. Иными словами, необходимо учитывать ситуации, когда задачи связаны между собой, например, по информации, т. е. выходная информация одних задач является входной для других. В качестве модели разработки такого рода проектов используется сетевая модель комплекса операций, так как сетевые модели позволяют отразить взаимосвязи операций проекта. Формализованная постановка данной задачи будет выглядеть следующим образом: необходимо оптимизировать некоторый функционал при выполнении ограничений на потребление ресурсов, выделенных на разработку проекта в размере, не превышающем заданного объема в заданном временном интервале. В качестве оптимизируемого функционала чаще всего используются экономические критерии. Задача в такой постановке относится к первому классу задач.

Задачи управления запасами

Одной из сфер практической деятельности, в которой успешно применяются методы системного анализа, является сфера управления запасами. Система управления запасами представляет особый интерес для системного аналитика. Это объясняется, во-первых, сложностью задач, которые приходится решать в этой сфере управления, а во-вторых, общностью постановки задачи, которая находит применение в системах различного типа, например, производственных системах, пред-

приятиях торговли и сбыта, при обосновании количества запасных изделий и т. п. Первые системы управления запасами были разработаны применительно к обоснованию необходимой потребности в запасных частях предприятий крупных компаний. Задачи управления запасами обладают одной особенностью – с увеличением запасов увеличиваются расходы, связанные с их хранением, но уменьшаются потери от возможного их дефицита. Следовательно, одна из задач управления запасами заключается в минимизации суммы ожидаемых затрат, связанных с хранением запасов, и потерь, обусловленных их отсутствием в случае необходимости.

Основная проблема, возникающая при решении задачи управления запасами, состоит в создании эффективной и надежной системы управления. Общая система управления запасами представляет собой три уровня решения данной задачи.

Первый уровень предусматривает обработку, ведение учета и хранение информации о запасах. Основные сведения, необходимые для успешного функционирования системы управления запасами, – это информация об уровне имеющихся запасов и запасов, которые будут созданы за счет размещения заказов, а также о заделе по заказам потребителя, если такие заказы имеются. Учет информации относительно любой имеющейся в запасах продукции включает в себя такие показатели как ее стоимость, единица измерения, источник получения, время упреждения и т. п. На первый взгляд количество различных данных о производимых или необходимых операциях невелико, однако в условиях функционирования крупных промышленных предприятий объем поступающей информации оказывается существенным. Значительно повысить эффективность обработки и хранения информации удается путем автоматизации процесса обработки данных.

Второй уровень рассматриваемой системы предполагает разработку правил принятия решения, на основе которых устанавливаются срок и размер заказа, необходимого для пополнения запасов. Каждое правило принятия решений является частной задачей оптимизации управления запасами. Задачи, которые приходится решать пользователю в практических условиях, могут отличаться по своему характеру, так как конкретные системы имеют различные параметры. Поэтому системному аналитику необходимо знать условия, при которых одно правило принятия решений оказывается предпочтительнее другого. С формальной точки зрения эти решения являются оптимальными, тем не менее, для успешного применения необходимо учитывать мнение соответствующих сотрудников органов управления, имеющих опыт работы и соответствующую квалификацию в данном вопросе. Следует иметь в виду, что при разработке правил принятия решений более важным является соответствующий учет реальной обстановки, чем достижение максимального показателя эффективности функционирования системы.

Третий уровень системы позволяет на основе разработанных правил принятия решений построить модель системы управления запасами и в соответствии с этой моделью определить стратегию функционирования системы на длительную перспективу. Организация контроля за работой системы дает возможность органам управления осуществлять управление системой в целом, а не запасами конкретной продукции и в случае необходимости вмешиваться в работу системы. Правила принятия решений по восстановлению уровня первоначального запаса и сохранению уровня резервного запаса разрабатываются с помощью методов прогнозирования. Точность прогноза определяется уровнем отклонения результатов прогноза от фактических данных. Каждое правило принятия решений содержит, по крайней мере, одну управляющую переменную, значение которой обычно задается органами управления. Это делается с целью обеспечения возможности оказывать влияние на перераспределение материальных ресурсов между затратами, связанными с запасами, и эксплуатационными затратами. Для ряда систем актуальными являются вопросы разработки правил принятия решений по созданию запасов в условиях наличия ограничений на ресурсы.

Таким образом, можно сделать вывод, что задача управления запасами есть комплексная задача, составными частями которой являются ведение информационной базы, построение моделей управления запасами, оптимизация объема создаваемого запаса и времени его пополнения и прочие вопросы.

Организация обслуживания оборудования

К задачам организации обслуживания оборудования относятся задачи назначения времени проведения проверок исправности функционирования оборудования, проведения профилактического обслуживания, выбор оптимального числа запасных изделий и приборов для оборудования, находящегося в эксплуатации. Любое оборудование в процессе работы изнашивается, устаревает и поэтому нуждается в организации контроля за исправностью его функционирования, а также в проведении ремонтных, восстановительных работ. Под техническим обслуживанием систем понимается совокупность мероприятий, которые служат поддержанию и восстановлению рабочих свойств систем. Данные мероприятия подразделяют следующим образом:

- текущее обслуживание;
- контроль работоспособности и диагностика отказов;
- ремонтно-восстановительные работы.

Предупредительные (регламентные) работы проводятся в системах, которые еще не утратили работоспособность, т. е. еще функционируют. Действия, необходимые для восстановления работоспособности после того, как система отказала, относятся к ремонтно-восстановительным работам. Обычно в теории систем предупредительные работы называют профилактическими, а восстановительные – аварийным восстановлением. В свою очередь, мероприятия по ремонту, в случае, если система после его окончания по своим качествам эквивалентна новой, иазывают полным восстановлением.

Характер функционирования и обслуживания систем существенно зависит от процессов, происходящих внутри системы и обуславливающих тип отказов. Различают внезапные и постепенные отказы. Внезапный отказ практически мгновенно переводит систему из работоспособного состояния в состояние отказа. О постепенных отказах говорят в тех случаях, когда удовлетворительное функционирование системы сохраняется в некоторой допустимой области определяющих параметров, которые, в свою очередь, зависят от времени. Наблюдения за вектором параметров позволяют спрогнозировать состояния работоспособности и неработоспособности системы. Если имеют место исключительно внезапные отказы, различают только работоспособное и неработоспособное состояния.

На практике при организации функционирования систем об их состоянии можно узнать только с помощью контроля. Контрольные проверки являются неотъемлемой частью мероприятий по восстановлению работоспособности. Поскольку, с одной стороны, отказ системы приводит или может привести к экономическим потерям, а с другой стороны, контроль также сопряжен с затратами, возникает задача оптимального с точки зрения общих затрат планирования проверок. Постановка задачи в этом случае будет следующая. Определить сроки проведения контрольных проверок по обнаружению неисправностей, при которых суммарные затраты на проведение контроля и потери от простоя оборудования из-за несвоевременного обнаружения и замены вышедших из строя элементов минимизируются.

Ремонтно-восстановительные мероприятия – более масштабные по своему содержанию работы. Они связаны с проведением комплексной проверки работоспособности систем, заменой отказавших или достигших установленного ресурса элементов, регулировкой отдельных параметров и прочими работами. При планировании профилактических и восстановительных мероприятий так же, как и в задаче планирования контрольных проверок необходимо учитывать, что несвоевременное проведение профилактических работ может привести к отказам системы и связанным с ними материальным потерям. С другой стороны, проведение профилактических работ также сопряжено с определенными материальными издержками, и поэтому их необоснованно частое проведение снижает эффективность функционирования системы. Таким образом, в данном случае также имеет место оптимизационная задача. Ее формулировка может выглядеть следующим образом: определить сроки проведения профилактических работ и замены оборудования, при которых суммарные ожидаемые затраты по ремонту и замене, а также потери, связанные с ухудшением технологических характеристик работы оборудования ввиду его старения, минимизируются на всем интервале эксплуатации системы.

И, наконец, существует еще одна задача, относящаяся к задачам организации обслуживания систем, – обеспечение запасными частями или элементами. Она формулируется следующим образом: какое количество запасных элементов следует иметь для того, чтобы быть уверенным в том, что система с вероятностью α будет бесперебойно функционировать в течение времени t.

Все сформулированные задачи относятся к числу оптимизационных. Однако описываемые в данном разделе задачи отличаются по постановке от задач предыдущих разделов. Дело в том, что для задач распределения ресурсов и управления запасами использовались, в основном, детерминистические модели. Для задач организации обслуживания необходимо применять вероятностные модели, поскольку в качестве управляющего параметра в данных задачах используется наработка объектов до отказа, а это, как известно, величина случайная. Поэтому при построении моделей такого типа необходимо проводить большую работу по сбору информации о функционировании системы и ее элементов, о наработках элементов до отказа, временах восстановления их работоспособности, стратегиях обслуживания и т.д. Таким образом, задачи организации обслуживания систем являются комплексными задачами, в которых требуется организовать и вести наблюдение за эксплуатационными параметрами системы и ее элементов, осуществлять обработку информации с целью определения параметров модели, формировать модель, решать оптимизационную задачу, принимать решения и внедрять их в практику управления эксплуатацией систем. В программе эксплуатации сложной системы помимо указанных работ необходимо также предусмотреть ее взаимодействие с операторами и другими системами, участвующими в решении общих задач, учет

различных режимов работы, перестройку структуры системы при возникновении нарушений, организацию ремонтов, модернизацию и продление ресурса системы.

Задачи массового обслуживания

Задачи массового обслуживания рассматривают вопросы образования и функционирования очередей. Для образования очереди необходим поток требований, которые ожидают поступления в обслуживающее устройство (обслуживающий канал). Под очередью подразумевают линейную цепочку выстроившихся один за другим объектов, нуждающихся в том или ином обслуживании. Объекты, подлежащие обслуживанию, называются заявками или требованиями на обслуживание. Объекты, осуществляющие обслуживание, называются обслуживающими каналами. В различных областях практической деятельности возникают ситуации, когда клиенты вынуждены ожидать своей очереди на обслуживание. Это относится к таким сферам деятельности как средства связи (телефония, почта, телеграф), транспортные системы, предприятия обслуживания (театры, рестораны, ремонтные мастерские), системы медицинского обслуживания, производственные процессы, процессы управления большими системами и т.д.

Приведем примеры систем, для которых характерно наличие заявок на обслуживание и обслуживающих каналов. Самый распространенный, известный всем из повседневной жизни пример – возникновение очереди у касс (железнодорожных, в кинотеатре, магазине и пр.). Второй пример – поступление в ремонтную мастерскую требующей технического обслуживания телерадиоаппаратуры, когда в зависимости от численности ремонтного персонала мастерская может обслужить только ограниченное число приборов, нуждающихся в ремонте, а большая часть аппаратуры лежит на полках в ожидании обслуживания. Третий пример – образование очереди пациентов на прием к врачу, когда даже при наличии определенной системы предварительной записи по целому ряду причин возможно образование очереди. Было бы весьма желательно ожидать врача согласно расписанию приема, которое приемлемо как для врача, так и для пациента. Дело в том, что нередко потеря времени для пациента связана с бо́льшим экономическим и социальным ущербом, чем потеря времени для врача. В этой ситуации, если расписание тщательно составлено на основе определенного исследования, это может только укрепить престиж лечебного заведения. По крайней мере, это свидетельствует об уважении, с которым клиника относится к пациенту и его времени. Таким образом, можно сделать вывод о необходимости проведения исследований при организации функционирования систем массового обслуживания.

Как те, кто образует очередь, так и те, к кому она образована, должны знать и понимать природу массового обслуживания и предвидеть его тонкости. Особенно это замечание относится к системе обслуживания. Специалисты, занимающиеся организацией работы системы массового обслуживания, должны знать, как лучше спланировать работу системы, чтобы минимизировать длину очереди и потери времени ожидающих.

Классификация целевых назначений теории массового обслуживания основана на выделении в ее структуре различных классов задач и соответственно областей применения получаемых результатов. Выделяют следующие классы задач: 1) задачи анализа поведения системы; 2) статистические задачи; 3) операционные задачи. Дадим им краткую характеристику.

Задачи анализа поведения системы. Цель рассмотрения задач такого рода заключается в том, чтобы с помощью математических моделей, описывающих свойства реальных систем, выявить операционные характеристики, определяющие поведение этих систем в процессе функционирования.

Статистические задачи. Статистическое исследование является неотъемлемой частью разработки математической модели реальной системы. В общем виде модель может существовать сама по себе, но приведение ее в количественное соответствие с конкретной системой достигается путем статистического анализа эмпирических данных, оценивания фигурирующих в модели параметров и проверки исходных гипотез. Параметры системы должны быть ассоциированы с процессом поступления требований и механизмом обслуживания.

Операционные задачи. Существование при решении задач массового обслуживания операционной направленности позволяет считать эту теорию одним из разделов системного анализа. Некоторые из операционных задач по своей природе относятся к разряду статистических, другие возникают при проектировании сложных систем, управлении реальными системами и оценке их эффективности. При постановке операционных задач следует различать описательный и нормативный подходы. В первом случае описание системы через ее операционные характеристики (характеристики поведения) используется для принятия решений относительно режима функционирования данной системы. С другой стороны, при нормативном подходе путем математического моделирования протекающих в системе процессов устанавливаются нормативные требования по обеспечению эффективной работы системы. Например, с учетом стоимостных и физических ограничений можно установить процедуры обслуживания, которые оптимизируют надлежащим образом построенную целевую функцию.

Имеет место еще одно обстоятельство, согласно которому задачи массового обслуживания относят к задачам системного анализа. Оно состоит в следующем. Функционирование систем массового обслуживания связано с людьми. Исследование причин очередей и принятие определенных мер нельзя полностью отделить от рассмотрения человеческого фактора и его влияния на работу системы. Нельзя принимать меры без учета того обстоятельства, что именно люди, использующие средства обслуживания, оказывают влияние на результат конечного анализа.

Таким образом, можно сделать вывод о том, что задачи массового обслуживания обладают рядом особенностей, согласно которым их можно отнести к задачам системного анализа. Эти задачи включают в себя необходимость проведения статистического анализа эмпирических данных, комплексных исследований операционных характеристик, учет человеческого фактора при организации работы систем массового обслуживания и принятии управленческих решений и, наконец, разработки оптимизационных критериев задачи массового обслуживания.

Планирование работ над проектом. Проектирование систем

Любой проект – это комплекс взаимосвязанных работ, для выполнения которых выделяются соответствующие ресурсы и устанавливаются определенные сроки. В последнее время особое внимание уделяется проблеме эффективного руководства проектными работами. Решение этой проблемы требует участия специалистов в самых различных областях науки и техники. Проектирование сложной системы – работа, требующая участия коллектива ученых и инженеров, причем работа такого коллектива должна планироваться с учетом деятельности других коллективов, подрядчиков и т.д.

Эффективное руководство коллективом, работающим над проектом, и его окружением требует учета и оптимизации психологических, экономических, организационных и других факторов. К настоящему времени является доказанным, что качество проекта существенно зависит от четкости формулировки целей проекта и ограничений, а также от системы поощрений коллектива исполнителей. Выбор проекта и распределение ресурсов включает определение того, какие из соответствующих идей следует принять за основу. Процесс выбора тесно связан с планированием работ над проектом, т. е. с календарным планированием выполнения заданий и задач, предусмотренных реализацией проекта. Именно в этом заключается основной вклад системного анализа в данной области. Планирование работ над проектом предусматривает также детальное определение организационных взаимоотношений в проектной организации.

В настоящее время методы системного анализа в сфере управления проектами используются, в основном, для решения трех проблем: выбора проекта, планирования выполнения работ над проектом и руководства проектом.

Общий процесс выбора можно рассматривать как процесс последовательного заполнения портфеля заказов. Лицу, принимающему решения, необходимо знать, какие средства можно расходовать на каждый из нескольких возможных проектов в каждый из периодов времени. В конце каждого периода времени портфель заказов меняется с учетом проектов, которые в этот момент имеются. Множество имеющихся проектов состоит из проектов, которые в текущий момент выполняются, и проектов, которые находятся в резерве. Определение таких оптимальных проектов исключительно важно для организации, поскольку решения по портфелю заказов влекут за собой большие организационные обязательства, выполнение которых может быть связано со значительными затратами. В такой ситуации важно знать, что будут выбраны наилучшие проекты, и средства для их выполнения будут распределены наилучшим из всех возможных способов. В простейшем виде постановка задачи выбора проектов аналогична задаче составления титульного списка, рассмотренной ранее. Однако оптимизация решений по выбору проектов в реальной ситуации исключительно сложна, поскольку для этих решений характерно как распределение ресурсов, так и распределение ответственности за выбор проектов. При обосновании выбора того или иного проекта необходимо сформировать систему ценностей, которая будет положена в основу принятия решения. Система же ценностей не всегда сводится только к экономическим критериям. Чаще всего это многофакторная характеристика, которая может содержать неформализуемые показатели.

Планирование выполнения работ представляет собой задачу упорядочения работ проекта и распределения выделенных ресурсов между работами с учетом отведенного на выполнение работ времени. Формулировка данной задачи аналогична задаче распределения ресурсов, рассмотренной в соответствующем разделе. Необходимо отметить, что при решении данной задачи широкое применение находят методы сетевого планирования и управления. При выполнении проекта могут возникнуть ситуации, когда необходимо производить перераспределение ресурсов

между проектными процедурами, осуществлять выравнивание потребления ресурсов, находить компромиссные решения относительно времени выполнения проекта и стоимости соответствующих работ.

Руководство проектом в процессе его осуществления состоит в сопоставлении текущих расходов на данный момент разработки проекта с запланированными средствами и корректировке проекта в случае больших отклонений для данного достигнутого состояния. Управление процессом проектирования тесно связано с этапом планирования, на котором определяются затраты и сроки выполнения операций проекта, и этапом выбора проекта и распределения ресурсов, на котором перераспределяются фонды между проектами в соответствии с соотношением состояния выполнения проекта и затратами, а также изменением этого соотношения во времени.

Задача анализа риска и безопасности использования новых технологий

В настоящее время происходит быстрое развитие качественно новых технологий, таких как добыча нефти на морском шельфе, ядерная энергетика, производство и транспортировка сжиженного газа, новые химические производства. Их функционирование связано с производством и транспортировкой новых материалов, большого количества энергии. При этом критерии экономической эффективности влияют на увеличение единичной мощности используемого оборудования, усиление концентрации производства, приближение новых промышленных объектов к потребителям продукции. Совокупность указанных факторов приводит к тому, что последствия аварийных ситуаций на этих производствах при неблагоприятных условиях могут оказаться катастрофическими. Хотя вероятность их возникновения очень низка, сами они и их последствия могут повлечь за собой крупные экономические потери, представлять серьезную угрозу экологии целых регионов и, что самое главное, жизни значительного количества людей.

Меры, направленные на увеличение безопасности используемого оборудования и уменьшение вероятности неблагоприятных событий и их последствий, требуют больших дополнительных затрат и не могут, как правило, абсолютно исключить возможность аварий. В связи с этим возникает необходимость исследования широкого круга задач, связанных с принятием обоснованных, рациональных решений с учетом факторов риска.

Имевшие место крупномасштабные технологические аварии значительно усилили внимание к данной проблеме как со стороны специа-

4 — 4355

листов, так и со стороны общественности и средств массовой информации. Вопросы использования новых технологий, выбора площадок для размещения потенциально опасных производств оказались в центре политической жизни многих стран. Проводимые научные исследования в этой области способствовали концентрации внимания на вопросах безопасности и оценки риска и в традиционных областях человеческой деятельности, таких как горное дело, металлургическое производство, автомобильное движение, строительство нефтепроводов, дамб и т. п. Одновременно активизировалось изучение проблем защиты окружающей среды и здравоохранения. В связи с этим исследования вопросов риска и безопасности стали охватывать все аспекты человеческого существования, всей окружающей человека среды обитания и постепенно оформились в новое научное направление, получившее название «анализ риска».

Основные проблемы анализа риска

Исследования по анализу риска включают в себя широкий спектр взаимосвязанных проблем. При этом основными являются: измерение риска; определение допустимого уровня риска; меры по предотвращению аварий; управление в условиях аварийных ситуаций.

Рассмотрим кратко их основные особенности.

Измерение риска. Первая проблема, возникающая при проведении любого аналитического исследования, в том числе и в области анализа риска, является проблемой измерений. Для ее решения, прежде всего, необходимо конкретизировать само понятие «риск». Трудность заключается в том, что выразить риск через один обобщенный показатель невозможно. Ожидаемое число жертв за год, вероятность для индивидуума стать жертвой той или иной технологии в течение года, вероятность опасных последствий для определенных групп людей (например, для детей), вероятность аварий с одновременным большим числом жертв и другие показатели по-разному характеризуют конкретную ситуацию и по-разному должны учитываться при принятии решений. Оценивая возможный вред от использования какой-либо технологии, необходимо принимать во внимание не только смертельные исходы, но и любой другой ущерб, который она может нанести здоровью людей (профессиональные заболевания и т. п.). Таким образом, само понятие риска многоаспектно; поэтому необходима разработка системы показателей, адекватно характеризующих величину риска в различных ситуациях.

Решение задачи анализа риска происходит, как правило, в условиях недостаточности или частичного отсутствия необходимой информации, особенно когда речь идет о новых технологиях. В этих случаях для получения количественных оценок показателей риска используют экспертные оценки, методы моделирования, строят «деревья отказов», пытаясь смоделировать возможные причины сбоев в сложных технических системах. В случае новых технологий, когда отсутствует сколько-нибудь надежная статистика, на всех этапах анализа используют в том или ином виде экспертные оценки вероятностных событий. В то же время известно, что человек является «плохим статистиком». Решение вероятностных задач представляет для него сложную проблему, с которой он не всегда хорошо справляется. Пытаясь решить ее, он часто прибегает к различным эвристикам, которые в ряде случаев ведут к существенным и устойчивым ошибкам. Следовательно, в процессе измерения риска необходимо учитывать реальные возможности получения надежных экспертных оценок.

Следует также заметить, что при измерении риска необходимо учитывать показатели, относящиеся к различным моментам времени. Например, при принятии решений о месте расположения нового технического производства необходимо учесть как оценки, характеризующие прямые потери, связанные со строительством, так и оценки, связанные с воздействием на окружающую среду в процессе его эксплуатации.

Одним из наиболее распространенных способов измерения риска, надежности сложных технических систем является построение «деревьев отказов». При этом определяются возможные поломки или отказы в системе и прослеживаются причинно-следственные цепочки вплоть до события, к которому эти отказы могут привести. Использование вероятностных оценок таких отказов дает возможность оценить в количественной форме вероятности соответствующих аварийных событий.

Определение допустимого уровня риска. Задача определения допустимого уровня риска, определение стандартов безопасности обслуживающего персонала и населения, является универсальной проблемой. Кажется естественной возможность установления единого допустимого уровня риска для различных технологий. Однако экономические соображения ставят под сомнение целесообразность такого единого показателя. Действительно, если техническое решение лишь незначительно уступает нормативному с точки зрения безопасности, но обходится значительно дешевле, то разумнее не добиваться достижения нормативного уровня безопасности ценой непомерно больших затрат, а использовать сэкономленные деньги в других областях с бóльшей эффективностью.

С экономической точки зрения логично потребовать, чтобы дополнительные затраты, направленные на эквивалентное снижение риска в различных областях человеческой деятельности, были бы одинаковы. Однако и это требование оказывается неосуществимым. Анализ существующих уровней риска, сопоставление затрат на спасение одной человеческой жизни при реализации различных программ безопасности показывают, что в действительности реальные уровни риска, которые считаются традиционно приемлемыми, значительно варьируются в различных областях. Так, считается необходимым добиваться большего уровня безопасности при эксплуатации АЭС, чем при использовании автомобильного транспорта. Удельные затраты на эквивалентное увеличение безопасности при осуществлении специальных программ в США, например, измеряются от нескольких десятков тысяч долларов до нескольких миллионов. Этот, кажущийся на первый взгляд парадокс, можно попытаться объяснить неразработанностью проблемы оценки риска, несовершенством организационных механизмов принятия решений и т.п. Однако многочисленные исследования свидетельствуют о том, что основная причина указанных различий состоит в особенностях субъективного восприятия риска. Люди по-разному воспринимают риск и соответственно по-разному оценивают величину допустимого уровня риска в зависимости от ряда сопутствующих ему обстоятельств; большое значение имеет так называемая степень обязательности при использовании той или иной технологии, или, другими словами, возможность для каждого человека принимать индивидуальное решение относительно ее использования (в случае использования автомобиля такая возможность имеется в отличие от ситуации, когда принято решение о постройке АЭС или другого промышленного объекта). Известно, что чем больше степень обязательности в использовании определенной технологии, тем меньший уровень риска считается допустимым. Такое же влияние на оценку допустимого уровня риска оказывают степень контролируемости ситуаций, т.е. возможность индивидуума влиять на происходящие события, а также степень новизны технологии. На уровень допустимого риска влияют оценки преимуществ, получаемых при реализации того или иного проекта, и оценки масштаба возможных неблагоприятных последствий, а также распределение их во времени и в пространстве между различными группами людей.

Всего известно несколько десятков факторов, определяющих восприятие и оценку риска в конкретных условиях, причем число их зависит от детальности рассмотрения проблемы. Все это позволяет понять, почему отдельные лица и общество в целом подходят фактически с разными мерками к оценке допустимого уровня риска в отдельных ситуациях.

Таким образом, определение допустимого уровня риска может решаться лишь как конкретная задача принятия решений с учетом экономических, психологических, социальных и других факторов, включая факторы риска и безопасности. При этом из множества альтернативных вариантов (различные виды технологий, проектные решения, место расположения производств), различающихся своими оценками по критериям, необходимо выбрать тот, который наилучшим способом сочетает в себе эти различные качества. Уровень риска, соответствующий выбранному решению, и может считаться рациональным в рассматриваемой задаче.

Меры по предотвращению аварий. Авария сложного объекта может быть следствием различных причин, которые допустимо условно разделить на две основные группы: технические причины, обусловленные недостатками в используемых технологических схемах или дефектами оборудования, и причины, связанные с неадекватным, несвоевременным, некачественным или ненадежным исполнением своих функций человеком-оператором или, иными словами, так называемым «человеческим фактором». Анализ многочисленных аварий показывает, что значительная их часть обусловлена неправильным поведением операторов и другого обслуживающего персонала. В связи с этим в настоящее время наряду с решением задач по повышению надежности оборудования все больше внимания уделяется особенностям поведения операторов сложных технических систем в чрезвычайных ситуациях.

Проведенные исследования обнаружили удивительную схожесть «сценариев» многих крупных аварий. Развитие ситуации, как правило, начинается с накопления ряда отклонений в поведении объекта. Затем следует какое-либо инициирующее событие, сопровождаемое неправильным управляющим воздействием со стороны оператора, которое и приводит к чрезвычайной ситуации. При этом именно ошибка оператора, как правило, значительно усугубляет последствия аварийной ситуации.

Подобные ошибки являются следствием нескольких причин. Во-первых, ни один специалист не обладает исчерпывающими знаниями об особенностях функционирования сложного объекта. Во-вторых, в процессе работы оператора происходит привыкание его как к нормальному функционированию управляемого им объекта, так и к небольшим отклонениям. Поэтому с течением времени оператор допускает все большие отклонения, с которыми он уже не может совладать, когда система выходит из-под контроля. Другими словами, наблюдается традиционный способ обучения методом «проб и ошибок», который оказывается недопустимо дорогим при освоении новых технологий. В связи с этим сегодня все большее внимание уделяется вопросам специальной подготовки операторов, разработке специальных тренажеров, систем поддержки принятия решений, способных оказать оператору эффективную помощь. Особое значение приобретает проблема адекватного информационного обеспечения оператора с учетом его возможностей по переработке больших объемов информации.

Одновременно с этим предпринимаются значительные усилия и в области повышения надежности и безопасности используемого оборудования. Для этого дублируются наименее надежные и наиболее критические с точки зрения безопасности элементы технологических систем, производится замена опасных веществ на менее токсичные соединения и т. п.

Предотвращение аварий тесно связано с проблемой допустимого уровня риска, установлением стандартов безопасности технических систем. Здесь необходимо учитывать не только особенности функционирования, но и масштабы тиражирования этих систем. При увеличении их числа происходит соответствующее увеличение вероятности наступления аварий даже при ничтожно малой вероятности отказа одной отдельной системы. Одним из способов регулирования в данном случае является введение динамичных, пересматриваемых с течением времени стандартов. При этом в процессе роста масштабов производства и потребления должно происходить ужесточение соответствующих стандартов.

Управление в условиях аварийных ситуаций. Поскольку никакие мероприятия по повышению безопасности не могут дать гарантии от аварийных снтуаций, необходимо предусматривать меры на случай их наступления. Такие меры позволяют, как правило, значительно уменьшить масштабы аварии и ее последствия. Они должны базироваться на анализе специальных сценариев чрезвычайных ситуаций и включать в себя широкий круг вопросов по организации аварийных работ как на объекте, так и в прилегающих районах, территория и население которых подверглись неблагоприятному воздействию. При этом большую роль играет заблаговременное распределение ответственности и обязанностей различных организаций, определение порядка их взаимодействия.

Известно, что при наступлении аварийной ситуации огромную роль играет фактор времени принятия решений. Поэтому необходимо обеспечить рациональное распределение принимаемых решений между центральными государственными органами управления и местными властями.

Разработка мероприятий по управлению в условиях аварийных ситуаций должна учитывать имеющийся опыт в этой области. Для этого необходим специальный банк данных об авариях, который позволял бы проводить их систематическое изучение. Подобный банк способствовал бы своевременному решению всех вопросов, касающихся измерения риска и управления в аварийных ситуациях.

Таким образом, особенности задачи анализа риска позволяют рассматривать ее как задачу системного анализа, заключающуюся в принятии сложного многокритериального коллективного решения, требующего исследования широкого круга вопросов и проведения комплексного анализа и оценки технических, экономических, социальных и даже политических факторов риска. При этом основной ее особенностью является доминирование социально-психологических аспектов, вовлеченность в ее решение групп людей со своими оценками и предпочтениями.

Проведение вероятностного анализа безопасности объектов повышенного риска

Наиболее отработаны методики и процедуры проведения анализа риска применительно к исследованию безопасности атомных электростанций (АЭС) и других ядерно-опасных объектов. Применительно к данным объектам разработана специальная методология, называемая вероятностным анализом безопасности (ВАБ). Рассмотрим основные цели и характеристику процедур, выполняемых при решении данной задачи.

Методы ВАБ стали стандартным инструментом при анализе безопасности атомных электростанций. Основным достоинством ВАБ является возможность углубленного качественного и количественного исследования проекта АЭС с точки зрения его внутренних свойств и воздействий со стороны окружающей среды, включая выявление факторов, вносящих наибольший вклад в риск, а также сравнения различных возможностей уменьшения риска. ВАБ дает возможность построить согласованную интегральную модель поведения станции с точки зрения безопасности. Соответственно ВАБ позволяет иметь основу для принятия решений в области безопасности. На общей основе, а именно, путем количественного сравнения оценок риска можно проводить сопоставления вариантов предлагаемых изменений или альтернативных решений в совершенно разных проектах и технических областях атомной энергетики. Более того, ВАБ представляет собой концептуальный и математический инструмент для проведения численных оценок риска в целом, связанного с атомными станциями и ядерными установками. ВАБ позволяет также численно оценить неопределенности таких оценок.

ВАБ ставит перед собой цель наиболее полно выявить аварийные последовательности, которые могут возникнуть в результате широкого спектра исходных событий, при этом используется систематизированный и реалистичный подход к определению частоты и последствий аварий. Существенным достижением ВАБ является возможность численно оценить неопределенность анализа безопасности с учетом мнений и выводов экспертов.

Общая процедура выполнения ВАБ-1 может быть организована в виде шести основных этапов.

Управление и организация. Этот этап включает в себя мероприятия и действия, необходимые для организации и управления исследованием:

- определение целей и объема исследований;
- определение схемы управления ВАБ;
- выбор методологии и разработки общих процедур анализа;
- отбор персонала и организация группы, которая будет проводить ВАБ, ее подготовка;
- разработка плана-графика, определение объема и получение финансирования;
- разработка процедур обеспечения качества и анализа работы.

Определение источников радиоактивности и событий, инициирующих аварии. На этом этапе устанавливаются потенциальные источники выхода радиоактивности в окружающую среду, определяются возможные состояния АЭС, которые необходимо рассмотреть, присущие станции функции безопасности, а также определяются события, инициирующие аварии, при которых может потребоваться выполнение этих функций, и системы, их реализующие. Устанавливаются связи между исходными событиями (ИС), функциями безопасности и системами, а также проводится разделение их на группы (группирование). На этом шаге группа аналитиков знакомится с объектом и используемой методологией и собирает большую часть исходной информации, на основании которой будет проводиться последующая работа.

Моделирование аварийных последовательностей. Третий этап относится к построению модели, которая описывает процесс возникновения аварии и реакцию станции. Эта модель состоит, главным образом, из комбинаций событий, включающих в себя исходные события, отказы систем и ошибки персонала, которые ведут к нежелательным последствиям. Эти комбинации событий называются аварийными последовательностями (АП), и цель этого этапа – определить их. Разрабатываются подробные модели анализа отказов систем и ошибок персонала. На этом этапе также проводится качественный анализ с целью учета в моделях зависимостей между событиями.

Получение и анализ данных и определение параметров. На этом этапе процедуры собирается и формируется вся информация, необходимая для количественного анализа модели, построенной на третьем этапе. В частности, определяются основные элементы модели станции и параметры, значения которых требуется найти. Необходимые для проведения этих оценок данные соответствующим образом собираются и обрабатываются. Искомые параметры могут быть разделены на три основные категории: частоты исходных событий, показатели надежности элементов и вероятности ошибок персонала. Также оцениваются параметры, необходимые для моделирования возможных зависимостей между различными элементами модели (исходными событиями, выходом из строя оборудования и ошибками персонала).

Количественный анализ аварийных последовательностей. Модель, построенная на третьем этапе, анализируется количественно с использованием результатов четвертого этапа. Результатами этого этапа являются оценки частот аварийных последовательностей. Обычно они сопровождаются анализом неопределенностей. Проводится исследование чувствительности результатов к важным допущениям и представляются показатели относительной значимости различных составляющих для результатов расчетов.

Документирование анализа, представление и интерпретация результатов. Результаты анализа тщательно документируются на каждом этапе и представляются так, чтобы наилучшим образом удовлетворить требованиям конечного пользователя материала. Необходимо дать анализ полученных результатов с точки зрения целей, поставленных перед ВАБ.

Вероятностный анализ безопасности является одним из наиболее действенных и эффективных инструментов, помогающих в принятии решений по безопасности и при управлении риском на атомных станциях.

Конкретными целями и соответствующими направлениями использования ВАБ являются следующие.

Определение доминирующих аварийных последовательностей. Здесь выявляются те комбинации исходных событий, отказов оборудования и ошибок персонала, которые могут привести к нежелательным последствиям со значительной частотой.

Определение систем, элементов и действий персонала, важных для безопасности. Проводится анализ результатов ВАБ с точки зрения выявления относительной значимости отдельных аварийных последовательностей для безопасности АС. Рассчитывается оценка относительной значимости различных систем, оборудования и эксплуатационных процедур. Оценивается неопределенность результатов расчетов.

Анализ важных системных и человеко-машинных зависимостей. Оцениваются зависимости между действиями персонала и системами, которые влияют на безопасность. Они включают в себя исходные события, вызванные общей причиной, отказы по общей причине, множественно зависимые ошибки персонала и другие факторы, которые уменьшают предусмотренную степень резервирования систем на АЭС, а следовательно, ее безопасность.

Выявление новых проблем безопасности и их оценка. В качестве результата ВАБ могут быть выявлены новые важные вопросы, относящиеся к безопасности конкретной АЭС.

Анализ тяжелых аварий. Результаты ВАБ могут помочь определить важные аварийные последовательности, которые должны рассматриваться как проектные аварии, а также другие аварийные последовательности, ведущие к запроектным авариям, для которых может потребоваться дальнейший анализ.

Решения по модернизации отдельной АС или группы АС. ВАБ может использоваться для количественного анализа относительной значимости конкретных изменений на эксплуатируемых АС. Результаты ВАБ позволяют провести ранжирование предлагаемых реконструкций и модернизации с точки зрения выявления относительной эффективности предлагаемых мероприятий для безопасности.

Сопоставление проектных решений. Результаты ВАБ на стадии проекта могут использоваться для оценки различных вариантов проектных решений.

Установление приоритетов в регулирующей деятельности и исследованиях по безопасности. Выводы ВАБ могут содействовать установлению приоритетов в регулирующих требованиях и в исследованиях по безопасности.

Таким образом, выполнение ВАБ предполагает проведение целого комплекса взаимосвязанных работ, направленных на изучение всех аспектов проблемы безопасности. При этом решаются такие специфичные для задач системных исследований вопросы как глубокий анализ структуры объекта, статистическое исследование объекта анализа, проведение операционных исследований. При этом статистическому

анализу подвергаются как технические объекты (системы защиты, объекты, обеспечивающие безопасное функционирование энергоблока, системы нормальной эксплуатации), так и люди, операторы энергоустановок, выполняющие соответствующие функции по управлению энергоблоком. В данном учебнике большое внимание уделяется анализу «человеческого фактора», его роли в обеспечении безопасного функционирования атомных станций. При операционном исследовании проводится комплексный анализ характеристик надежности элементов и систем энергоблока, оцениваются их ресурсные характеристики, выполняется оценка риска при эксплуатации АС. Глубина исследования обеспечивается также путем проведения анализа неопределенности, чувствительности, значимости моделей систем и процессов развития аварийных ситуаций.

Обсуждение

Таким образом, рассмотрено несколько типовых задач, с которыми потенциально может столкнуться системный аналитик в своей практической деятельности. Для решения задачи распределения ресурсов применяются методы линейного программирования, задачи управления запасами решаются с помощью более сложных моделей, в основном, методами нелинейного программирования; задачи организации обслуживания и задачи массового обслуживания решаются с помощью методов статистического анализа, дифференциальных уравнений и теории восстановления; задачи управления проектами и вероятностного анализа безопасности решаются с помощью методов статистического анализа, эвристических процедур, методов теории восстановления и других математических и эвристических методов. Математические методы решения экстремальных задач составляют основу аппарата системного анализа, но сама теория системного анализа никак не может быть сведена только лишь к решению экстремальных задач. Более того, как уже отмечалось ранее, системный анализ не является чисто математической дисциплиной. Главные сложности анализа, как правило, заключаются не в преодолении математических трудностей.

Первый шаг любой задачи – это ее формализация, описание с помощью языка математики. От того, насколько успешно формализована задача, зависит вся судьба системных исследований. Простое описание делает анализ простым, но если оно не будет в достаточной степени адекватно реальности, то результат исследования, основанного на таких моделях, будет иметь сомнительную достоверность. С другой стороны, переусложненная задача, учитывающая разнообразные детали исследуемых процессов и с большими подробностями описывающая реальность, может привести к большим затратам времени, необходимого для исследований, и высокая точность моделей может оказаться неоправданной, а результат неприемлемым по времени его получения. Таким образом, системный аналитик должен руководствоваться своим опытом, уметь вникать в содержание задачи, ясно понимать цели всего исследования.

Еще более сложные проблемы возникают при попытке формирования критерия оценки качества результатов проводимых исследований, при сравнении различных вариантов стратегий развития систем. Как уже было сказано, на данном этапе исследователи сталкиваются с такими проблемами как многокритериальность задачи, неопределенность целей и т.д. Преодолеть возникающие неопределенности формальными методами невозможно. На данном этапе необходимо проводить дополнительные исследования с целью проверки всевозможных гипотез.

И, наконец, одна из сложностей, возникающих при проведении системных исследований – это учет так называемого «человеческого фактора», т.е. учет поведения активных и пассивных участников реально функционирующей системы.

Глава 2

ХАРАКТЕРИСТИКА ЭТАПОВ СИСТЕМНОГО АНАЛИЗА

2.1. Процедуры системного анализа

В предыдущей главе были сформулированы три этапа проведения системного анализа. Эти этапы являются основой решения любой задачи проведения системных исследований. Суть их состоит в том, что необходимо построить модель исследуемой системы, т.е. дать формализованное описание изучаемого объекта, сформулировать критерий решения задачи системного анализа, т.е. поставить задачу исследования и далее решить поставленную задачу. Указанные три этапа проведения системного анализа являются укрупненной схемой решения задачи. В действительности задачи системного анализа являются достаточно сложными, поэтому перечисление этапов не может быть самоцелью. Практикующему системотехнику требуется представить методику выполнения каждого из этапов. В данной главе сосредоточим внимание на последовательности операций выполнения системного анализа. Специалисты по системному анализу в своих работах приводят различные схемы его проведения. Фактически эти процедуры или операции не отличаются от тех, которые присущи любому научному исследованию. Задача данной главы заключается в том, чтобы представить для каждого этапа конструктивную схему действий, в наибольшей степени отвечающую поставленным задачам. Поэтому помимо описания процедур проведения системного анализа рассмотрим вопросы их практического выполнения. Таким образом, основное внимание сосредоточим на методике проведения системного анализа. Использование правильной методики гарантирует исследователю, что он не будет искать решения неверно поставленной задачи. Грамотное проведение системного анализа предупреждает также и возможность неверного решения правильно поставленной задачи. Если исследовательская группа руководствовалась правильной методикой, то разработанные модели адекватны изучаемой проблеме и допустимы с точки зрения реализации вычислительного процесса, выполняются ограничения на выделяемые средства и сроки исполнения работ, а внедрение результатов системного анализа осуществляется квалифицированно и эффективно. В этом случае работы по системному анализу завершаются достижением цели.

Отметим также, что методика проведения системного анализа и руководящие принципы не являются универсальными – каждое исследование имеет свои особенности и требует от исполнителей интуиции, инициативы и воображения, чтобы правильно определить цели проекта и добиться успеха в их достижении. Перейдем к формулированию последовательности работ по проведению системного анализа. Как уже было отмечено, специалисты по системному анализу приводят различные схемы его выполнения, которые представляются в виде алгоритмов. То обстоятельство, что системный анализ оперирует не только формализованными, но и неформализованными процедурами, не означает, что нельзя говорить о его алгоритмах. Неоднократно имели место попытки создать достаточно общий, универсальный алгоритм системного анализа. Тщательное рассмотрение имеющихся в литературе алгоритмов показывает, что у них большая степень общности в целом и различия в частностях, деталях. Постараемся изложить основные процедуры алгоритма проведения системного анализа, которые являются обобщением последовательности этапов проведения такого анализа, сформулированных рядом авторов [1, 13, 14], и отражают его общие закономерности. При этом нельзя утверждать, что предлагаемая схема работ по проведению системного анализа является универсальной. Как замечают авторы монографии [1], алгоритм является прагматической моделью деятельности. Было бы неправильно утверждать, что один алгоритм является более правильным, чем другой, что реализация одного из них является системным анализом, а другого – нет. Выбрав конкретный алгоритм выполнения работ по системному анализу, необходимо следовать предписаниям именно данного алгоритма. Если бы был выбран другой алгоритм, то работы велись бы согласно схеме действий, предписываемых моделью другого алгоритма. Следует заметить, что различные алгоритмы системного анализа могут быть взаимозависимыми, например, ряд этапов может совпадать. Однако при этом в них может уделяться большее внимание различным вопросам. Соотношение алгоритмов проведения системного анализа такое же, как алгоритмов программирования. Одна и та же, скажем вычислительная, задача может быть решена различными способами. Существуют различные численные методы реализации одних и тех же процедур, разная квалификация исполнителей, опыт работы, предпочтения в использовании тех или иных процедур, в конце концов, существуют разные языки программирования. Естественно, что разные программисты реализуют одну и ту же вычислительную задачу с помощью разных программ. Одна программа будет изящна, другая старомодна, но все они будут решать одну и ту же задачу. Точно также системный аналитик может использовать тот или иной алгоритм системных исследований. Важно, чтобы все они позволяли решать задачи системного анализа и приводили к достижению поставленной цели.

Перечислим основные процедуры системного анализа:

• изучение структуры системы, анализ ее компонентов, выявление взаимосвязей между отдельными элементами;

• сбор данных о функционировании системы, исследование информационных потоков, наблюдения и эксперименты над анализируемой системой;

• построение моделей;

• проверка адекватности моделей, анализ неопределенности и чувствительности;

• исследование ресурсных возможностей;

• определение целей системного анализа;

• формирование критериев;

• генерирование альтернатив;

• реализация выбора и принятие решений;

• внедрение результатов анализа.

Перейдем теперь к изложению сути работ, выполняемых на каждом из перечисленных этапов.

2.2. Анализ структуры системы

Любая задача системного анализа начинается с построения модели исследуемой системы. Для решения задачи построения модели необходимо вначале произвести изучение структуры системы, выполнить анализ ее компонентов, выявить взаимосвязи между отдельными элементами. Чтобы обоснованно проводить анализ структуры системы, необходимо рассмотреть ряд понятий и определений, характеризующих строение и функционирование системы.

Основные понятия и определения

В качестве первого понятия охарактеризуем элемент. Под элементом принято понимать простейшую неделимую часть системы. Понятие элемента условно, так как зависит опять же от уровня иерархии рассмотрения объектов в структуре системы. Принято считать, что

элемент – это предел членения системы с точки зрения решения конкретной задачи и поставленной цели. Рассмотрим примеры элементов. В системе управления и защиты энергоблоков атомных станций при одном уровне иерархии рассмотрения системы в качестве элементов можно выделить блок питания, датчики (камеры нейтронные компенсирующие), устройства отображения информации (электронные показывающие приборы), устройства, вырабатывающие сигналы срабатывания аварийной защиты (по превышению уровня мощности, по превышению скорости нарастания мощности и т.п.), устройства автоматического регулирования и прочие блоки и устройства. В свою очередь, каждый из приведенных блоков и устройств может быть расчленен на более мелкие составляющие. Так в их структуре можно выделить резисторы, конденсаторы, диоды, транзисторы. В ряде устройств используются процессоры, элементы памяти, устройства ввода и вывода информации, которые также можно расчленить на составляющие элементы.

В качестве следующей структурной компоненты рассмотрим подсистему. **Подсистема** – совокупность взаимосвязанных элементов, обладающая свойствами системы (в частности, свойством целостности), способная выполнять относительно независимые функции, подцели, направленные на достижение общей цели системы. Отличие подсистемы от простой группы элементов состоит в том, что для подсистемы формулируются подцели ее функционирования. Продолжим рассмотрение системы управления и защиты энергоблоков атомных станций. В данной системе выделяют ряд подсистем: автоматического регулирования, ручного регулирования, аварийной защиты, аварийной и предупредительной сигнализации и т.д. Каждая из указанных подсистем выполняет конкретную функцию. Так, например, подсистема автоматического регулирования выполняет функцию поддержания заданного уровня мощности энергоустановки. Подсистема аварийной защиты принудительно прекращает работу реакторной установки в случае возникновения аварийной ситуации. Таким образом, каждая из подсистем выполняет свою конкретную функцию.

Если же части системы не обладают свойством целостности и способностью выполнять независимые функции, а представляют собой совокупности однородных элементов, то такие части принято называть **компонентами**.

Структура отражает определенные взаимосвязи, взаиморасположение составных частей системы, ее устройство, строение. При описании системы недостаточно перечислить элементы, из которых она состоит. Требуется отобразить систему путем расчленения ее на подсистемы, компоненты и элементы и показать, каким путем обеспечивается в объекте выполнение поставленной цели. Для выполнения такой процедуры и вводят понятие структуры. Таким образом, структура отражает наиболее существенные взаимоотношения между элементами и их группами, которые мало меняются при изменениях в системе и обеспечивают существование системы и ее основных свойств. Структура характеризует организованность системы, устойчивую упорядоченность ее элементов и связей. *Структура системы* – состав системы и схема связей между ее элементами. Понятие структуры можно определить следующим образом. Совокупность отношений, заданных на множестве подсистем и элементов, образующих некоторую систему, называется *структурой* этой системы.

Следующее понятие, которое будем рассматривать, – **связь**. Данное понятие входит в любое определение системы наряду с понятием элемент и обеспечивает возникновение и сохранение структуры и целостных свойств системы. Понятие связь характеризует одновременно и статическое строение системы, и динамическое ее поведение. Связь определяют как ограничение степени свободы элементов. Связь характеризуется направлением, силой и характером. По первому признаку связи делят на направленные и ненаправленные. По второму признаку различают сильные и слабые связи. По характеру выделяют связи подчинения, равноправные, генетические, связи управления. Различают также связи по направленности процессов – прямые и обратные. Обратные связи могут быть положительными, сохраняющими тенденции происходящих в системе изменений того или иного параметра, и отрицательными – противодействующими тенденциям изменения выходного параметра. Обратная связь направлена на сохранение, стабилизацию требуемого значения параметра. Обратная связь является основой приспособления систем к изменяющимся условиям существования, основой саморегулирования и развития систем.

Следующее понятие – цель системы – важное понятие, лежащее в основе развития систем. **Цели системы** – желательные состояния или результаты поведения системы. *Глобальная цель системы* – конечное состояние, к которому стремится система в силу своей структурной организации. Цель можно также определить следующим образом: «*цель* – это субъективный образ (абстрактная модель) не существующего, но желаемого состояния среды, которое решило бы возникшую проблему» [16]. В практических применениях цель – это идеальное устремление, которое позволяет коллективу увидеть перспективы или реальные возможности, обеспечивающие своевременное завершение очередного этапа на пути к идеальным устремлениям. Цель достигается путем решения задач. *Задачи системы* – цели, которые желательно достичь к определенному моменту времени в пределах определенного периода функционирования системы.

Для описания системы создается ее модель. **Модель** – это отражение структуры системы, ее элементов и взаимосвязей, направленное на отображение определенной группы свойств. Создание модели системы позволяет предсказывать ее поведение в определенном диапазоне условий.

Формы представления структур

Структурные представления являются средством исследования систем. Одну и ту же систему можно представить различными структурами, необходимый выбор которых обусловлен содержанием исследований, проводимых на данном этапе. Принятый способ описания структур – графическое изображение. В таком графе элементы, компоненты, подсистемы и прочие объекты системы отображаются в виде вершин графа; связи между объектами представляют в виде дуг. Рассмотрим основные способы представления структур.

Сетевые структуры представляют собой отображение взаимосвязи объектов между собой. Их применяют для представления организационных структур, для изображения структурных схем систем, для представления информационного обеспечения и т.д. С помощью сетевых структур отображаются пространственные взаимосвязи между элементами, как правило, одного иерархического уровня. Примером сетевой структуры может служить структурная схема ЭВМ (рис. 2.1). На рисунке стрелками показаны связи между элементами.

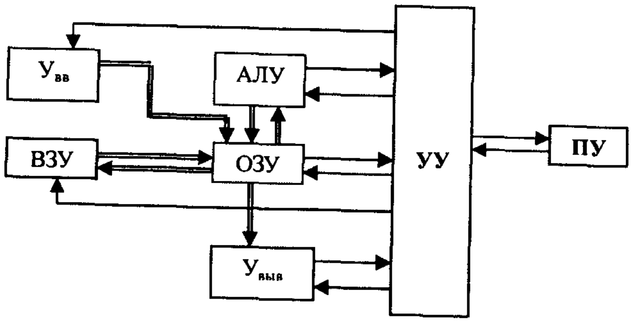

Рис. 2.1. Упрощенная структурная схема ЭВМ:

$У_{вв}$ – устройство ввода информации; ВЗУ – внешнее запоминающее устройство; АЛУ – арифметико-логическое устройство; ОЗУ – оперативное запоминающее устройство; УУ – устройство управления; ПУ – пульт управления; $У_{выв}$ – устройство вывода информации

Различают следующие виды сетевых структур. *Линейные структуры* со строго упорядоченным взаимоотношением элементов «один к одному». Примером линейной структуры может служить схема одного из каналов (любого) аварийной защиты энергоблока ядерной энергетической установки (ЯЭУ). Каналы строятся по принципу линейного соединения группы устройств (рис. 2.2).

Рис. 2.2. Схема канала аварийной защиты энергоблока ЯЭУ:

КНК – камера нейтронная компенсирующая (датчик нейтронного потока); БП – блок питания; УЗС – устройство защиты по скорости нарастания мощности (вторичный прибор)

Древовидная структура представляет собой объединение многих линейных подструктур. Примером может служить схема подсистемы аварийной защиты энергоблока ЯЭУ. Подсистема состоит из группы однотипных каналов, каждый из которых дублирует работу других (рис. 2.3).

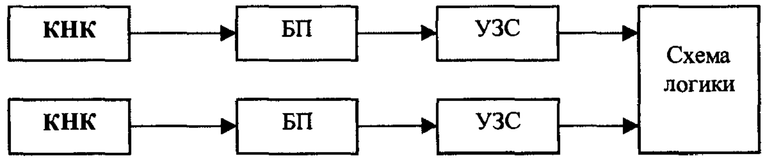

Рис. 2.3. Схема подсистемы аварийной защиты энергоблока ЯЭУ

Кольцевая структура (циклическая) имеет замкнутые контуры в соответствующих графах. С помощью циклических структур изображаются схемы циркуляции информации в системах. Обобщенная сетевая структура характеризуется многочисленными межэлементными связями.

Иерархические структуры представляют собой декомпозицию системы в пространстве и применяются, прежде всего, для описания подчиненности элементов в структурах управления. Термин *иерархия* означает соподчиненность, порядок подчинения низших по должности лиц высшим. В настоящее время концепция иерархии распространена на любой согласованный по подчиненности порядок объектов. В иерархических структурах важно лишь выделение уровней соподчиненности, а между уровнями и между компонентами в пределах уровня, в принципе, могут быть любые взаимоотношения. Примером применения иерархической структуры может служить изображение схемы ЭВМ с детализацией на каждом новом уровне иерархии (рис. 2.4).

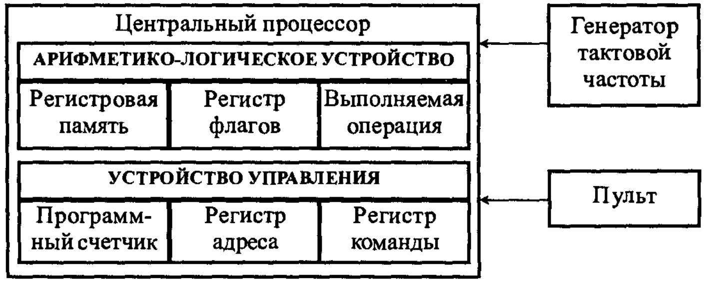

Рис. 2.4. Фрагмент схемы ЭВМ с детализацией описания отдельных уровней

Таким образом, рассмотрены основные понятия, с помощью которых осуществляется решение задачи анализа системы. Для решения задачи в том виде, в котором она была сформулирована в п. 2.1, на данном этапе необходимо произвести изучение структуры системы, анализ ее компонентов, выявление взаимосвязей между отдельными элементами, т.е. осуществить структурную декомпозицию системы. Основное содержание процедур, выполняемых на данном этапе, состоит в том, чтобы подготовить информацию к проведению работ по построению модели системы.

2.3. Сбор данных о функционировании системы. Исследование информационных потоков

Решение задачи, сформулированной в предыдущем параграфе, а именно, изучение структуры системы, анализ ее компонентов, выявление взаимосвязей между отдельными элементами, преследует цель отразить статическое состояние системы. Однако при проведении системного анализа исследователя интересуют вопросы, касающиеся изучения свойств системы. Свойства системы реализуются в процессе ее функционирования, т.е. в процессе динамического поведения системы. Чтобы построить модель системы, которая имела бы возможность отражать свойства и характеристики системы, реализующиеся в процессе ее функционирования во времени, необходимо помимо структуры системы знать ее параметры, поэтому следующим этапом работ при проведении системного анализа является сбор данных о функционировании системы и исследование информационных потоков.

Основное содержание данного этапа состоит в идентификации параметров системы с целью последующего включения их в модель. Этот этап связан с определением числовых значений параметров системы в режиме ее функционирования. Параметры системы подразделяются на внутренние и внешние.

Внешние параметры системы – характеристики функционирования системы, служащие показателями качества ее работы как единого целого. В качестве примера внешних параметров можно привести параметры автоматизированной системы:

• общая производительность системы по обработке данных;
• объем передаваемой информации;
• достоверность выходной экспериментальной информации;
• точность получения результатов (для информации, заданной количественно);
• количественные характеристики надежности системы;
• объем используемой в системе аппаратуры (объем памяти, количество преобразователей формы информации, количество внешних устройств и т.д.);
• время задержки с момента поступления в систему исходных данных до момента выдачи окончательных результатов (во время решения определенной задачи);
• стоимость системы (с учетом разработки математического обеспечения);
• показатели удобства системы в эксплуатации (наличие элементов «психологического комфорта») и др.

Внутренние параметры системы – характеристики, показывающие особенности технических решений, принятых при организации системы в целом и отдельных технических средств, входящих в состав системы, а также в совокупности влияющие на значения внешних параметров системы. Примерами внутренних параметров автоматизированных систем являются:

• вид и характеристики сигналов для представления информации в системе, в каналах связи – при обмене информацией между отдельными звеньями системы;
• способ кодирования информации;
• вид приоритетности при приеме и обработке информации от различных источников;
• способ организации программы-диспетчера;
• быстродействие отдельных элементов и т.д.

Ограничения, накладываемые на внутренние параметры, связаны с конечностью распространения электромагнитных волн, с конечностью возможных типов используемых блоков и элементов аппаратуры системы, с дискретностью числа ячеек запоминающих устройств и т.д. Разделение на внешние и внутренние параметры весьма условно. Обычно к внешним относят те параметры, на которые задаются ограничения, определяемые назначением системы или вызванные условиями ее функционирования.

При описании параметров системы определению подлежат идентификатор параметра, единицы измерения, диапазон изменения, качественные характеристики (однозначный – многозначный, регулируемый – не регулируемый), место применения в модели.

Параметры отражают свойства системы. Одни из них определить достаточно просто, например, такие параметры как объем памяти, количество внешних устройств, стоимость системы, способ кодирования информации, вид приоритетности при приеме и обработке информации от различных источников можно получить на основании изучения документации на систему. Другие же параметры определяются опосредованно, на основании обработки информации, полученной в результате наблюдений за работой системы. К таким параметрам относятся параметры, характеризующие надежность системы, качество функционирования, точность получения количественных результатов и т.п.

Наблюдения с целью сбора данных могут проводиться в процессе функционирования системы либо же для сбора данных организуются специальные экспериментальные исследования. В первом случае говорят, что данные получены в результате пассивного эксперимента. Во втором случае имеет место активный эксперимент. Активный эксперимент проводится по специально составленному плану с использованием методов планирования эксперимента. При этом предусматривается возможность изменения входных параметров, влияющих на процесс функционирования системы. Исследуется изменение выходных параметров системы в зависимости от уровней входных параметров. Результаты испытаний фиксируются с помощью измерений, т.е. изображения результатов опыта в виде символов, номеров или чисел. Измерение – это алгоритмическая операция, которая данному наблюдаемому состоянию системы или процесса ставит в соответствие определенное обозначение: число, номер или символ. Такое соответствие обеспечивает то, что результаты измерений содержат информацию об исследуемой системе. Требуемая информация в виде оценок параметров получается путем преобразования результатов измерения или, как еще говорят, с помощью обработки экспериментальных данных.

Современное понимание измерений существенно шире только количественных измерений. Есть наблюдаемые явления, в принципе не допускающие числовой меры, но которые можно фиксировать в «слабых» шкалах, и эти результаты учитывают в моделях, получая качественные, но вполне научные выводы. Расплывчатость некоторых наблюдений признана их неотъемлемым свойством, которому придана строгая математическая форма, и разработан формальный аппарат работы с такими наблюдениями. Погрешности измерений также являются неотъемлемым, естественным и неизбежным свойством процесса измерения. Причиной этого является наличие неопределенностей в процессе наблюдения, квантование результатов измерения, шумы аппаратуры и т.д. Широкое распространение получили статистические измерения, т.е. оценивание функционалов распределения вероятностей по зафиксированным наблюдениям значений реализации случайного процесса. Этот класс измерений имеет особое значение потому, что большое число входных и выходных показателей реализуется в виде случайных величин.

Анализ информационных потоков

В случае, когда анализируют социотехнические системы (организационные, человеко-машинные, автоматизированные) помимо определения параметров системы, для построения модели важное значение приобретают вопросы исследования информационных потоков, циркулирующих в системе. Анализ информационных потоков позволяет выявить схему работы объектов управления, обеспечивает информационное отображение объекта управления, взаимосвязь между его элементами, структуру и динамику информационных потоков. Изучаются формы документов и недокументированных сообщений. В процессе изучения информационных потоков анализируются следующие группы документов:

1) официальные положения и инструкции, регламентирующие функции подразделений и определяющие сроки и процедуры обработки информации и принятия решений;

2) входные документы, источники которых находятся вне системы;

3) систематически обновляемые записи в виде картотек или книг, используемые в процессе работы;

4) промежуточные документы, получаемые и используемые в процессе обработки данных;

5) выходные документы.

Анализ информационных потоков осуществляется с помощью специально разработанных методов: графического, метода с использованием сетевой модели, графоаналитического и метода с использованием графов типа «дерево». *Графический* метод применяется для описания потоков информации, главным образом, на макроуровне, когда решается задача анализа общей схемы работы объектов управления. Здесь отношения между элементами потока, в виде которых выступают документы, изображают структурно-информационно-временной схемой. На схеме приводятся краткие пояснения, описывающие движение информации и материальных потоков.

Метод с *использованием сетевой модели* состоит в следующем. В качестве события сетевой модели фигурирует определенный документ. Если документ представляет собой результат выполнения какой-либо работы, он является конечным, если же он будет использоваться в дальнейшем ходе выполнения работ, такой документ будет начальным. Под работой понимается определенная задача или функция, выполняемая элементом органа управления.

Графоаналитический метод основан на анализе матрицы смежности информационного графа. В данном случае исходными для анализа информационных потоков являются данные о парных отношениях между наборами информационных элементов, формализуемые в виде матрицы смежности. Под информационными элементами понимают различные типы входных, промежуточных и выходных данных. Матрица смежности – квадратная бинарная матрица с количеством строк (и столбцов), равным количеству информационных элементов. В каждой позиции матрицы смежности записывают единицу, если между соответствующими элементами матрицы существует отношение, т.е. информация одного документа используется при формировании другого, и в соответствующей позиции ставится ноль, если отношения нет. Далее матрице смежности ставится в соответствие граф информационных взаимосвязей. Множеством вершин графа является множество информационных элементов, дуги отражают взаимосвязи между элементами. Дуга присутствует, если в матрице смежности отношение между элементами отмечено единицей, и отсутствует в противном случае. Анализ графа позволяет выявить информационную зависимость между входными, промежуточными и выходными документами, характер зависимости, установить направление движения информации. Графоаналитический метод является развитием метода с использованием сетевой модели и позволяет проводить более детальный анализ информационных потоков.

Метод *с использованием графов типа «дерево»* применяют для описания системы потоков информации. Строится граф взаимосвязи показателей и так называемые графы расчетов, описывающие преобразование информации в процессе формирования отдельных показателей. При построении дерева взаимосвязи показателей ребра ориентируют с учетом иерархии от исходных к результирующим. Такой подход позволяет строить графы с более высокой степенью укрупнения. Полученный комплекс графов отражает процесс движения и преобразования информации в системе.

2.4. Построение моделей систем

На основании изложенного в предыдущих двух параграфах решаются задачи изучения структуры системы, выявления параметров, характеризующих функционирование системы и влияющих на эффективность и качество ее работы, анализа информационных потоков, циркулирующих в системе. Данные этапы являются предварительными этапами работы по построению модели системы; цель этих этапов – выявление основных структурных элементов, динамических и информационных компонентов системы. После выяснения этих вопросов переходят к решению основной задачи – построению модели системы.

Моделью называют некий объект, который в определенных условиях может заменять оригинал, воспроизводя интересующие свойства и характеристики оригинала. Модели бывают материальные и абстрактные. Разновидностью абстрактных моделей являются математические модели. Они и будут объектом дальнейшего рассмотрения.

Построение математической модели системы есть процесс формализации определенных сторон существования, жизнедеятельности системы, ее поведения с точки зрения конкретной решаемой задачи. Различают статические и динамические модели. Статическая модель отражает конкретное состояние объекта. Примером статической модели является структурная схема системы. Динамическая модель описывает процесс изменения состояний системы. При решении задач системного анализа цели исследования заключаются в изучении характеристик системы, прогнозировании путей развития системы, сравнении вариантов развития и т.п., т.е. интересуются, в основном, вопросами динамического поведения систем. Следовательно, можно сказать, что динамические модели находят более широкое применение, чем статические.

Следующий вопрос, на котором следует остановиться при обсуждении подходов к построению математической модели, – это целевое предназначение модели. Перед тем как приступать к созданию математической модели необходимо уяснить существо решаемой задачи, для которой создается данная модель. Ошибочным будет разработка модели системы, описывающая все стороны, все аспекты существования и развития системы. Такая модель будет излишне громоздка и скорее всего не пригодна для проведения каких-либо серьезных исследований. Модель всегда должна быть конкретной, нацеленной на решение поставленной задачи. Для оценки характеристик надежности системы необходимо строить модель надежностную, для решения задач прогнозирования развития производственных процессов – производственную модель, для решения экономических задач – экономическую модель. Если перед системными аналитиками ставится задача исследования ряда аспектов, то целесообразнее создавать несколько моделей, а не пытаться разрабатывать одну всеобъемлющую модель. Правда, в этом случае необходимо, чтобы разные модели, отражающие различные аспекты существования и развития системы, были взаимосвязаны по входным и выходным параметрам и характеристикам системы. Такая взаимосвязь достигается путем проведения итеративных расчетов на моделях, т.е. осуществляется последовательный расчет моделей. Те параметры, которые известны до проведения расчетов, задаются в качестве входных в каждой из моделей, где их присутствие необходимо. Недостающие параметры получают расчетным путем и последовательно включают в модели от первой к последующим по мере проведения расчетов. На начальном этапе эти параметры заменяют оценками, принадлежащими области определения параметра. По мере получения результатов модели должны уточняться и процесс расчетов по уточненным моделям должен повторяться. В этом заключается итеративность процесса. Расчеты прекращаются, когда исследователь отмечает сходимость процессов уточнения параметров.

Рассмотрим теперь типы математических моделей. Выделяют два класса моделей: аналитические и имитационные. В аналитических моделях поведение сложной системы записывается в виде некоторых функциональных соотношений или логических условий. Наиболее полное исследование удается провести в том случае, когда получены явные зависимости, связывающие искомые величины с параметрами сложной системы и начальными условиями ее изучения. Однако это удается выполнить только для сравнительно простых систем. Для сложных систем исследователю приходится идти на упрощение реальных явлений, дающее возможность описать их поведение и представить взаимодействия между компонентами сложной системы. Это позволяет изучить хотя бы некоторые общие свойства сложной системы, например, оценить устойчивость системы, характеристики надежности и т.п. Для построения математических моделей имеется мощный математический аппарат (функциональный анализ, исследование операций, теория вероятностей, математическая статистика, теория массового обслуживания и т.д.). Наличие математического аппарата и относительная быстрота и легкость получения информации о поведении сложной системы способствовало повсеместному и успешному распространению аналитических моделей при анализе характеристик сложных систем.

Когда явления в сложной системе настолько сложны и многообразны, что аналитическая модель становится слишком грубым приближением к действительности, системный аналитик вынужден использовать имитационное моделирование. В имитационной модели поведение компонентов сложной системы описывается набором алгоритмов, которые затем реализуют ситуации, возникающие в реальной системе. Моделирующие алгоритмы позволяют по исходным данным, содержащим сведения о начальном состоянии сложной системы, и фактическим значениям параметров системы отобразить реальные явления в системе и получить сведения о возможном поведении сложной системы для данной конкретной ситуации. На основании этой информация аналитик может принять соответствующие решения. Отмечается, что предсказательные возможности имитационного моделирования значительно меньше, чем у аналитических моделей.

Вопрос о том, какой модели следует отдать предпочтение при проведении исследований характеристик системы, не является очевидным. Аналитическая модель имеет некоторые преимущества по сравнению с имитационной моделью. Во-первых, аналитическая модель дает решение поставленной задачи в законченной форме. Во-вторых, применение аналитической модели обеспечивает глубину анализа. С помощью аналитических моделей можно проводить исследование характеристик в некоторой области определения параметров, в которой модель адекватна описываемым явлениям или процессам. Применение аналитических моделей позволяет получить решение в виде функциональной зависимости исследуемых характеристик от параметров модели. Имитационная модель за один цикл ее применения производит расчет характеристик в одной точке. Для получения функциональной зависимости выходной характеристики от параметров модели необходимо провести многократные расчеты на имитационной модели.

С другой стороны, построить аналитическую модель для сложной системы очень трудно. При таком построении требуется принимать существенные упрощающие предположения, которые могут привести к тому, что построенная модель будет неадекватна описываемым процессам или явлениям. В этом смысле имитационные модели имеют преимущества, так как они могут быть построены в самых общих предположениях о функционировании системы. Следовательно, имитационные модели могут быть более адекватны. К недостаткам аналитических моделей относится также и то, что простая модификация проекта или изменение предположений о функционировании элементов структуры может потребовать коренной перестройки модели, в то время как у имитационной модели потребуется изменить лишь входную информацию.

Рассмотрим простой пример. Пусть необходимо оценить характеристики надежности системы, структура которой известна. Если проводить расчеты в предположении об отсутствии восстановительных мероприятий после отказов элементов, то аналитическая модель для такого расчета строится с использованием логико-вероятностного подхода. Если изменить предположения и считать, что после отказа элементов осуществляется восстановление и потоки отказов и восстановлений пуассоновские, то для расчета надежности используются уравнения Колмогорова–Чепмена. Если же будем предполагать восстановление элементов, но потоки отказов или восстановлений описывать не пуассоновским, а каким-нибудь другим распределением, то для построения моделей расчета надежности необходимо использовать аппарат теории восстановления, т.е. для решения одной и той же задачи при смене предположений о характере функционирования системы для построения аналитической модели приходится полностью менять теоретический аппарат. В имитационной модели в этом случае меняются лишь входные данные. Таким образом, на основании сказанного нельзя однозначно решить, какая модель лучше. Обе модели являются полезным инструментом исследования и об их соответствии решаемым проблемам надо судить в контексте конкретного применения. В задачах системного анализа целесообразно проводить комбинированные исследования, использующие как аналитические, так и имитационные модели.

2.5. Проверка адекватности моделей, анализ неопределенности и чувствительности

После того как модель построена, необходимо удостовериться в ее качестве. С этой целью выполняют ряд операций, а именно, – проверку адекватности модели процессу, объекту или явлению, для которых она построена, проверку непротиворечивости модели, неопределенности, чувствительности, реалистичности и работоспособности. Рассмотрим существо каждой из проводимых работ.

Проверка адекватности моделей

Важный вопрос, который интересует исследователя после того, как построена модель исследуемого явления или процесса, – это проверка адекватности модели. Проверить адекватность модели – это значит установить, насколько хорошо модель описывает реальные процессы, происходящие в системе, насколько качественно она будет прогнозировать развитие данных процессов. Проверка адекватности модели проводится на основании некоторой экспериментальной информации, полученной на этапе функционирования системы или при проведении специального эксперимента, в ходе которого наблюдаются интересующие системного аналитика процессы. Проверка адекватности модели заключается в доказательстве факта, что точность результатов, полученных по модели, будет не хуже точности расчетов, произведенных на основании экспериментальных данных. Если иметь в виду целевое предназначение моделируемого объекта, то под адекватностью модели нужно понимать степень ее соответствия этому предназначению. В качестве примера, иллюстрирующего необходимость решения вопроса об адекватном описании результатов наблюдений соответствующими моделями, рассмотрим регрессионную модель, с помощью которой описали поведение некоторого процесса. Рассмотрим два рисунка (рис. 2.5, *а* и *б*) с одинаковым расположением экспериментальных точек и, следовательно, одинаковым разбросом относительно линии регрессии. Эти рисунки различаются тем, что модели, изображенные на них, построены на основании разного количества экспериментальных данных. В связи с этим имеем различный средний разброс в экспери-

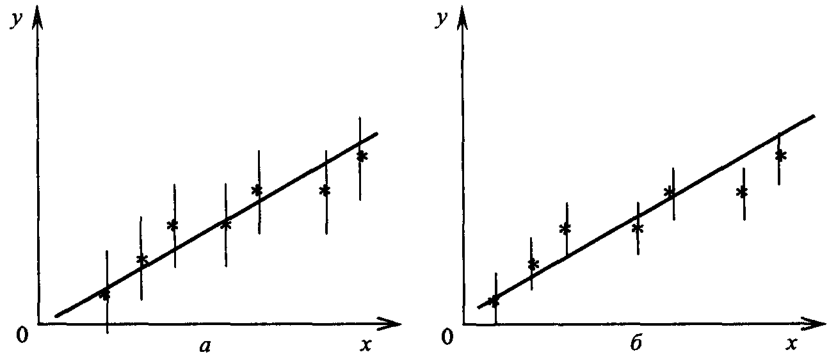

Рис. 2.5. Проверка адекватности модели:

а – объем экспериментальных данных мал; *б* – объем экспериментальных данных велик ментальных точках факторного пространства. Разброс в точках показан отрезками прямых, численно равных величине доверительного интервала, построенного для функции отклика.

Линейная модель регрессии адекватна в первом случае (рис. 2.5. *а*), так как разброс в точках того же порядка, что и разброс относительно линии регрессии. Во втором случае (рис. 2.5. *б*) не все отрезки прямых, численно равных величине доверительного интервала, накрывают линию регрессии. Следовательно, в этом случае требуется более сложная модель, чтобы точность ее предсказания была сравнима с точностью экспериментальных данных.

В первом случае модель обладает удовлетворительными точностными характеристиками по сравнению с экспериментальной информацией, на основании которой она построена. Во втором случае точность предсказания модели хуже точности экспериментальных данных. Таким образом, модель адекватна экспериментальным данным только в первом случае.

Непротиворечивость модели

Целью данного этапа является проверка предположения: дает ли модель не противоречащие логике результаты при вариации величин важнейших параметров, особенно в тех случаях, когда их значения близки к экстремальным. Чтобы ответить на этот вопрос, необходимо проанализировать характер реакции модели на изменения соответствующих входных параметров. Для проверки непротиворечивости модели, в первую очередь, анализируют, какие результаты дает модель при нулевых значениях входных параметров, в том числе в нулевой момент времени, далее исследуется состояние модели на границе области определения входных параметров, например, в точке бесконечность, если она входит в область определения.

Можно привести такой пример. При решении задачи анализа надежности сложной системы вычисляют значение коэффициента готовности. Асимптотические модели для описания коэффициента готовности хорошо известны, они приведены в соответствующей литературе, например в [22]. В случае расчета коэффициента готовности системы в моменты времени, сравнимые с наработками элементов системы до отказа, приходится строить неасимптотические модели. Они гораздо сложнее асимптотических. Для проверки непротиворечивости неасимптотических моделей можно использовать известные результаты, полученные для описания асимптотического поведения коэффициента готовности. Для этого необходимо в неасимптотической модели задать

время, равное бесконечности, и провести сравнение результатов расчета с асимптотической моделью. Если результаты совпали, это может служить подтверждением того, что модель дает результаты, не противоречащие известным ранее.

Анализ неопределенности модели

Поскольку модель системы только стремится отобразить реальность, неизбежно существование упрощений, допущений и идеализаций сложных процессов и явлений, происходящих в системе. Следствием этих упрощений и идеализаций будут неопределенности в итоговых результатах, получаемых в процессе применения модели. Природа возникновения неопределенностей многогранна. Выделяют следующие источники неопределенностей в соответствующих моделях: обусловленные неполнотой моделей, неадекватностью моделей и неопределенностью исходных параметров.

Неопределенности, обусловленные *неполнотой моделей*, возникают из-за того, что при построении моделей системный аналитик не предусмотрел некоторые стороны развития моделируемых процессов, происходящих в системе. Иными словами, при разработке модели системы не были учтены отдельные особенности существования и развития систем. Это может быть сделано сознательно, когда аналитик считает, что данные особенности системы не играют большой роли и ими можно пренебречь. Иногда это происходит в результате недостаточной проработанности вопросов, связанных с изучением структуры и динамического поведения систем. В результате имеем недостаток полноты модели, который приводит к неопределенности в результатах и выводах и который трудно проанализировать и определить количественно.

Второй тип неопределенностей связан с *неадекватностью моделей*. Даже в тех случаях, когда в модели учтены все особенности существования и развития систем, последовательность событий и логические особенности функционирования систем, заложенные в модель, не точно отражают реальность. Существуют неопределенности, вызванные неадекватностью концептуальных и математических моделей, числовой аппроксимацией, ошибками в вычислительных программах и ограничениями вычислительного процесса. Эти неопределенности рассматриваются как часть анализа неопределенности моделей; для оценки их относительной значимости проводятся исследования чувствительности результатов моделирования.

Третий тип неопределенностей – *неопределенность исходных параметров*. Параметры различных моделей точно не известны. Причиной этого является недостаточность данных, используемых при статистическом оценивании входных параметров, невозможность точного описания поведения персонала, работающего в составе анализируемой системы, наличие допущений, принятых при составлении модели. Эта третья категория неопределенностей при современном состоянии методологии может быть наиболее успешно охарактеризована численно.

Анализ чувствительности модели

Анализом чувствительности модели называют процедуру оценки влияния допусков входных параметров на ее выходные характеристики. Проводят анализ чувствительности следующим образом: задают отклонение входного параметра в правую и левую стороны от его среднего значения и фиксируют, как при этом изменяются выходные значения характеристик модели. В качестве величины отклонения обычно принимают среднее квадратическое отклонение. Практическая сторона анализа чувствительности модели к изменению входных параметров состоит в том, что устанавливается степень зависимости выходных параметров от входных характеристик. Эту степень влияния затем можно проранжировать и выявить наиболее значимые входные параметры. Если в ходе проверки модели на чувствительность к изменению входных параметров установлено, что ряд параметров приводят к незначительным изменениям выходных характеристик, сравнимых с точностью проведения расчетов на модели, то данные входные параметры можно вывести из модели. Таким образом, анализ чувствительности модели может привести к упрощению модели и исключению из нее незначимых факторов.

Реалистичность

Установить реалистичность модели, значит ответить на вопрос: соответствует ли модель тем частным случаям, для которых уже имеются фактические данные. Одним из способов проверки реалистичности модели может служить метод прогнозирования назад, т.е. в модели задаются требуемые входные параметры и производится расчет некоторого события, которое уже имело место, или же рассчитываются характеристики системы на время, которое система отработала, и оценки этих характеристик можно получить по реальным данным. Если результаты расчета на модели дают хорошее совпадение с практикой, то можно считать, что модель реалистична.

Работоспособность

Цель анализа работоспособности модели – выяснить, насколько модель практична и удобна в эксплуатации. Во-первых, модель должна обеспечивать результат за разумное время. Если модель используется в процессе выработки и принятия решения, то необходимо, чтобы расчеты можно было выполнить в пределах сроков, установленных для подготовки соответствующих решений. Если это условие не выполняется, смысл модели пропадает, так как теряется ее предназначение. Во-вторых, трудозатраты и ресурсы, требуемые для эксплуатации модели, должны укладываться в установленные лимиты машинного времени и фонда зарплаты. Должно выполняться условие практической целесообразности.

Следующий аспект проверки модели связан с анализом допущений и предположений, принятых при построении модели. На этом этапе проверки работоспособности оценивается качество модели, ее свойства в условиях воздействия реальных внешних возмущений и параметров. Суть данной процедуры состоит в том, чтобы удостовериться, что при составлении модели «не выплеснули ребенка вместе с водой». Принятие некоторых допущений и ограничений может привести к тому, что модель не будет отражать сути происходящих явлений и процессов. Следует отметить, что эта задача взаимосвязана с задачей проверки адекватности модели.

2.6. Исследование ресурсных возможностей

Для того, чтобы модель начала давать результаты, чтобы она заработала, необходимы затраты ресурсов. Модель нужно не только воплотить в надлежащем виде, но и обеспечить возможность получения решения нужного качества и к нужному моменту времени. Не требует пояснений то обстоятельство, что даже самое обоснованное решение становится ненужным, если оно появляется после того, как истекли сроки, выделенные для принятия решения, т.е. когда необходимость в нем уже отсутствует. Поэтому при реализации моделей необходимы ресурсы, которые позволяют обеспечить выполнение условий качества и своевременности. Принципиальное значение имеет вопрос, в какой степени обеспечено ресурсами управление ходом выполнения задач системного анализа.

Рассмотрим основные виды ресурсов, используемых при реализации задач системного анализа. Выделяют энергетические, материальные, временные и информационные ресурсы. Характеристику ресурсов начнем с *энергетических ресурсов*. Обычно энергетические затраты на реализацию модели значительно меньше, чем затраты энергии, потребляемые самой системой, для которой разработана модель. Поэтому в обычной ситуации энергетическими ресурсами, как правило, пренебрегают. Однако актуальность обеспечения энергетическими ресурсами возникает в тех случаях, когда модельные исследования проводятся на объектах, работающих в относительно автономных условиях. Примерами таких объектов являются научно-исследовательские морские суда, летательные аппараты, космические станции. Энергетические возможности таких объектов ограничены. Следовательно, прежде чем приступить к организации модельных исследований в таких условиях, требуется обосновать обеспеченность готовящихся исследований с точки зрения достаточности энергетических ресурсов.

Следующий вид ресурсов – *материальные ресурсы*, которые представляют собой достаточно обширную категорию. Сюда можно отнести и людские ресурсы, требуемые для реализации моделей, и ресурсы обеспеченности проводимых исследований необходимым оборудованием, приборами и инструментами, и канцелярские товары и принадлежности, и т.п. В случае решения задачи путем моделирования на ЭВМ в качестве материальных ресурсов выступают объем памяти и машинное время. Указанные ресурсы ограничивают возможности решения задач большой размерности в реальном масштабе времени. С подобными проблемами приходится сталкиваться при решении задач в экономических, социальных, метеорологических, организационно-управленческих, сложных технических системах. В случае нехватки ресурсов для решения подобных задач необходимо проводить реконструкцию модели. Самое тривиальное действие, которое можно применить в данном случае, это провести декомпозицию модели системы на совокупность связанных моделей меньшей размерности.

Временные ресурсы. Практика решения задач системного анализа такова. Заказчик работ заключает с системными аналитиками, которые выступают в роли исполнителей работ, договор. В данном договоре оговариваются сроки выполнения работ по проведению системного анализа. Как правило, эти сроки являются ограничивающим временным фактором на выполняемые работы. Таким образом, исследования, проводимые с помощью моделей, должны по времени укладываться в рамки, оговоренные договором.

Наконец, *информационные ресурсы*. Количество и качество информации, используемой при построении моделей систем, различно. Если при построении модели используется достоверная информация в достаточно представительном объеме, это является одним из условий построения хорошей модели. Качество и полнота информации, представленной в модели, обеспечивает принятие обоснованных решений и является гарантией успешного управления. В свою очередь, ограниченность информации приводит к значительной неопределенности результатов, получаемых в ходе расчетов на модели. Решения, принимаемые на основе таких моделей, будут обладать слабой степенью обоснованности.

Таким образом, при построении и реализации моделей следует уделять внимание обеспечению процесса использования моделей всеми видами ресурсов. Даже самая качественная модель в смысле адекватности описания происходящих в системе процессов может на практике оказаться бесполезной, если она не обеспечена в надлежащем объеме всеми видами ресурсов, необходимых для ее успешного применения.

2.7. Определение целей системного анализа

Формулирование проблемы

Для традиционных наук начальный этап работы заключается в постановке формальной задачи, которую надо решать. В исследовании сложной системы это промежуточный результат, которому предшествует длительная работа по структурированию исходной проблемы. Начальный пункт определения целей в системном анализе связан с формулированием проблемы. Здесь следует отметить следующую особенность задач системного анализа. Необходимость системного анализа возникает тогда, когда заказчик уже сформулировал свою проблему, т.е. проблема не только существует, но и требует решения. Однако системный аналитик должен отдавать себе отчет в том, что сформулированная заказчиком проблема представляет собой приблизительный рабочий вариант. Причины, по которым исходную формулировку проблемы необходимо считать в качестве первого приближения, состоят в следующем. Система, для которой формулируется цель проведения системного анализа, не является изолированной. Она связана с другими системами, входит как часть в состав некоторой надсистемы, например, автоматизированная система управления отделом или цехом на предприятии является структурной единицей АСУ всего предприятия; АСУ предприятия имеет связи с отраслевой системой; сама система, в свою очередь, состоит из подсистем. И поэтому, формулируя проблему для рассматриваемой системы, необходимо учитывать, как решение данной проблемы отразится на системах, с которыми связана данная система. Неизбежно планируемые изменения затронут и подсистемы, входящие в состав данной системы, и надсистему, содержащую данную систему. Таким образом, к любой реальной проблеме следует относиться не как к отдельно взятой, а как к объекту из числа взаимосвязанных проблем.

Другая причина того, что к сформулированной заказчиком проблеме следует относиться как к первоначальному рабочему варианту, состоит в том, что она (проблема) является его рабочей моделью, его взглядом на проблемную ситуацию. В реальной жизни необходимо учитывать позиции всех заинтересованных сторон. Учет мнений всех заинтересованных сторон приводит к дополнениям, уточнениям первоначального варианта описанной проблемы. Следовательно, системное исследование проблемы должно начинаться с ее расширения до системы проблем, связанных с исследуемой, без учета которых она не может быть решена. Это расширение должно происходить как с учетом связей данной системы с над- и подсистемами, так и с точки зрения углубления данной проблемы, ее детализации.

Для формулирования системы проблем необходимо сформировать перечень заинтересованных лиц, так или иначе связанных с работами по системному анализу. В данный перечень следует включать, во-первых, клиента, который ставит проблему, заказывает и оплачивает системный анализ. Именно заказчик формулирует исходную проблему системного анализа. Далее включаются лица, принимающие решения, от полномочий которых зависит решение проблемы. Необходимо учитывать мнения активных участников решения проблемы, поскольку на них лягут основные работы по реализации принятых решений. Следующий контингент – пассивные участники, те, на ком скажутся последствия решения проблемы. На этапе формулирования проблемы необходимо учитывать, к каким изменениям приведут внедрения мероприятий проведенного системного анализа, и как это отразится на пассивных участниках. И, наконец, требуется включать в перечень самого системного аналитика и его сотрудников, главным образом, для того, чтобы предусмотреть возможность минимизации его влияния на остальных заинтересованных лиц.

При формулировании системы проблем системный аналитик должен следовать некоторым рекомендациям. Во-первых, за основу должно браться мнение заказчика. Как правило, в качестве такового выступает руководитель организации, для которой проводится системный анализ. Именно он, как было отмечено выше, генерирует исходную формулировку проблемы. Далее системный аналитик, ознакомившись со

сформулированной проблемой, должен уяснить задачи, которые были поставлены перед руководителем, ограничения и обстоятельства, влияющие на поведение руководителя, противоречивые цели, между которыми он старается найти компромисс. Насколько это возможно, следует выяснить личные качества руководителя, его склонности и предубеждения. Далее системный аналитик должен изучить организацию, для которой проводится системный анализ. Необходимо тщательно ознакомиться с существующей иерархией управления, функциями различных групп, а также предыдущими исследованиями соответствующих вопросов, если таковые проводились. Аналитик должен воздерживаться от высказывания своего предвзятого мнения о проблеме и от попыток втиснуть ее в рамки своих прежних представлений ради того, чтобы использовать желательный для себя подход к ее решению. Наконец, аналитик не должен оставлять непроверенными утверждения и замечания руководителя. Как уже отмечалось, проблему, сформулированную руководителем, необходимо, во-первых, расширять до комплекса проблем, согласованных с над- и подсистемами, и, во-вторых, согласовывать ее со всеми заинтересованными лицами.

Следует также отметить, что каждая из заинтересованных сторон имеет свое видение проблемы, отношение к ней. Поэтому при формулировании комплекса проблем необходимо учитывать, какие изменения и почему хочет внести та или другая сторона. Кроме того, проблему необходимо рассматривать всесторонне, в том числе и во временном, историческом плане. Требуется предвидеть, как сформулированные проблемы могут измениться с течением времени или в связи с тем, что исследование заинтересует руководителей другого уровня. Формулируя комплекс проблем, системный аналитик должен дать развернутую картину того, кто заинтересован в том или ином решении.

Определение целей

После того, как сформулирована проблема, которую требуется преодолеть в ходе выполнения системного анализа, переходят к определению цели. Определить цель системного анализа – это означает ответить на вопрос, что надо сделать для снятия проблемы. Сформулировать цель – значит указать направление, в котором следует двигаться, чтобы разрешить существующую проблему, показать пути, которые уводят от существующей проблемной ситуации.

Формулируя цель, требуется всегда отдавать отчет в том, что она имеет активную роль в управлении. В определении цели было отражено, что цель – это желаемый результат развития системы. Таким образом, сформулированная цель системного анализа будет определять весь дальнейший комплекс работ. Следовательно, цели должны быть реалистичны. Задание реалистичных целей направит всю деятельность по выполнению системного анализа на получение определенного полезного результата. Важно также отметить, что представление о цели зависит от стадии познания объекта, и по мере развития представлений о нем цель может быть переформулирована. Изменение целей во времени может происходить не только по форме, в силу все лучшего понимания сути явлений, происходящих в исследуемой системе, но и по содержанию, вследствие изменения объективных условий и субъективных установок, влияющих на выбор целей. Сроки изменения представлений о целях, старения целей различны и зависят от уровня иерархии рассмотрения объекта. Цели более высоких уровней долговечнее. Динамичность целей должна учитываться в системном анализе.

При формулировании цели нужно учитывать, что на цель оказывают влияние как внешние по отношению к системе факторы, так и внутренние. При этом внутренние факторы являются такими же объективно влияющими на процесс формирования цели факторами, как и внешние.

Далее следует отметить, что даже на самом верхнем уровне иерархии системы имеет место множественность целей. Анализируя проблему, необходимо учитывать цели всех заинтересованных сторон. Среди множества целей желательно попытаться найти или сформировать глобальную цель. Если этого сделать не удается, следует проранжировать цели в порядке их предпочтения для снятия проблемы в анализируемой системе.

Исследование целей заинтересованных в проблеме лиц должно предусматривать возможность их уточнения, расширения или даже замены. Это обстоятельство является основной причиной итеративности системного анализа.

На выбор целей субъекта решающее влияние оказывает та система ценностей, которой он придерживается, поэтому при формировании целей необходимым этапом работ является выявление системы ценностей, которой придерживается лицо, принимающее решение. Так, например, различают технократическую и гуманистическую системы ценностей. Согласно первой системе природа провозглашается как источник неисчерпаемых ресурсов, человек – царь природы. Всем известен тезис: «Мы не можем ждать милостей от природы. Взять их у нее наша задача». Гуманистическая система ценностей говорит о том, что природные ресурсы ограничены, что человек должен жить в гармонии с природой и т.д. Подробный анализ этих систем ценностей приведен в [18]. Практика развития человеческого общества показывает, что следование технократической системе ценностей приводит к пагубным последствиям. С другой стороны, полный отказ от технократических ценностей тоже не имеет оправдания. Необходимо не противопоставлять эти системы, а разумно дополнять их и формулировать цели развития системы с учетом обеих систем ценностей.

2.8. Формирование критериев

Критерий – это способ сравнения альтернатив. Необходимо различать понятия критерий и критериальная функция. Критерием качества альтернативы может служить любой ее признак, значение которого можно зафиксировать в порядковой или более сильной шкале. После того как критерий сформирован, т.е. найдена характеристика, которая будет положена в основу сравнения альтернатив, появляется возможность ставить задачи выбора и оптимизации.

Задача формирования критериев решается непосредственно после того, как сформулированы цели системного анализа. Ситуация становится понятной, если к критериям относиться как к количественным моделям качественных целей. Задача системного аналитика состоит в том, чтобы формализовать проблемную ситуацию, возникающую в ходе системного анализа. Этой цели как раз и служит этап формирования критериев. Сформированные критерии в некотором смысле должны заменять цели. От критериев требуется как можно большее сходство с целями, чтобы оптимизация по критериям соответствовала максимальному приближению к целям. Выполняя данный этап, необходимо сознавать, что критерии не могут полностью совпадать с целями. Одной из причин этого является то, что критерии и цели формулируются в разных шкалах: цели в номинальных, критерии в более сильных, допускающих упорядочение. Критерий является отображением ценностей, воплощенных в целях, на параметры альтернатив, допускающие упорядочение. Определение значения критерия для данной альтернативы является косвенным измерением степени ее пригодности как средства достижения цели.

Обсуждая вопрос формирования критериев, следует сказать, что это достаточно трудная и серьезная задача. Редко бывает так, что решение лежит на поверхности. Зачастую для формирования хорошего критерия, адекватно отражающего цель системного анализа, приходится прибегать к неформализуемым процедурам. Неформализуемые, творческие, эвристические этапы играют важную роль в процессе формирования критериев. При решении задач системного анализа, возникает ситуация, когда невозможно предложить один критерий, адекватно отражающий цель исследования: даже одну цель редко удается выразить одним критерием, хотя к этому необходимо стремиться. Критерий, как и всякая модель, лишь приближенно отображает цель; адекватность одного критерия может оказаться недостаточной. Поэтому решение может состоять не обязательно в поиске более адекватного критерия, оно может выражаться в использовании нескольких критериев, описывающих одну цель по-разному и дополняющих друг друга. Еще более усложняется задача в случае, когда сформулировано несколько целей системного анализа, отражающих разные системы ценностей. В этом случае исследователь тем более вынужден формировать несколько критериев и в последующем решать многокритериальную задачу. Таким образом, можно отметить, что многокритериальность является способом повышения адекватности описания цели. Однако введение многокритериальности в задачах системного анализа не должно быть самоцелью. Качество постановки задачи заключается не только и не столько в количестве критериев, сколько в том, чтобы они достаточно адекватно описывали цель системного анализа. Критерии должны описывать по возможности все важные аспекты цели, но при этом желательно минимизировать число необходимых критериев.

Формирование критериев отражает цель, которую ставит заказчик. Но при постановке и решении задач системного анализа необходимо учитывать не только цели, на решение которых он направлен, но и возможности, которыми обладают стороны для решения поставленных задач и которые позволяют снять выявленные проблемы. В первую очередь, необходимо учитывать ресурсы, имеющиеся у сторон. К ресурсам следует отнести денежные ресурсы, которые заказчик согласен выделить системным аналитикам для решения поставленной задачи; ресурсы исполнителя – людские ресурсы, ресурсы вычислительные (наличие вычислительной техники, ее количество и т.д.), материальные ресурсы, требуемые для решения задач (например, наличие канцелярских товаров, транспорта, ресурсов связи); временны́е ресурсы (сроки решения задач системного анализа, как правило, оговариваются). При формулировке задачи системного анализа необходимо также учитывать интересы окружающей среды. Хоть окружающая среда и играет пассивную роль, необходимо учитывать, что любая система существует внутри нее, взаимодействует с ней. Поэтому при постановке задачи системного анализа необходимо следовать принципу не навредить, не предпринимать ничего, что противоречило бы законам природы. Чтобы удовлетворить условиям непревышения количества имеющихся ресурсов, в постановку задачи системного анализа вводят ограничения.

Между целевыми критериями и ограничениями имеются сходство и различия. Общее заключается в том, что и критерий, и ограничения являются математической формулировкой некоторых условий. В некоторых задачах оптимизации они могут выступать равноправно. Однако на этапе формирования целевой критерий открывает возможности для генерирования новых альтернатив в поисках лучшей из них, а ограничение заведомо уменьшает их число, запрещая некоторые из них. Одними целевыми критериями можно жертвовать ради других, ограничения же исключить нельзя, они должны четко соблюдаться. При формулировании задач системного анализа встречаются случаи, когда ограничения задаются завышенными. Это может привести к нереальности достижения целей системного анализа. В этом случае необходимо ставить вопрос об ослаблении ограничений. Приведем пример. Слишком высокие требования к характеристикам надежности системы могут привести к необходимости чрезвычайных дополнительных финансовых вложений. А это, в свою очередь, может привести к неэффективности разработки и эксплуатации объекта, для которого проводится анализ. Таким образом, формулируя ограничения, необходимо руководствоваться соображениями здравого смысла. В качестве приема, позволяющего найти наилучшие соотношения между критериями и ограничениями, можно порекомендовать использование итерационных процедур. После проведения определенных вычислений и установления факта завышенности требований, сформулированных в ограничениях, можно эти требования ослабить и попытаться решить задачу заново.

В заключение данного параграфа перечислим основные критерии, наиболее часто встречающиеся в анализе сложных технических систем. Это экономические критерии – прибыль, рентабельность, себестоимость; технико-экономические – производительность, надежность, долговечность; технологические – выход продукта, характеристики качества и пр.

2.9. Генерирование альтернатив

Следующим этапом системного анализа является создание множества возможных способов достижения сформулированной цели. Иными словами, на данном этапе необходимо сгенерировать множество альтернатив, из которых затем будет осуществляться выбор наилучшего пути развития системы. Данный этап системного анализа является очень важным и трудным. Важность его заключается в том, что конечная цель системного анализа состоит в выборе наилучшей альтернативы на заданном множестве и в обосновании этого выбора. Если в сформированное множество альтернатив не попала наилучшая, то никакие самые совершенные методы анализа не помогут ее вычислить. Трудность этапа обусловлена необходимостью генерации достаточно полного множества альтернатив, включающего в себя, на первый взгляд, даже самые нереализуемые.

Генерирование альтернатив, т.е. идей о возможных способах достижения цели, является настоящим творческим процессом. Существует ряд рекомендаций о возможных подходах к выполнению рассматриваемой процедуры. Необходимо сгенерировать как можно бо́льшее число альтернатив. Имеются следующие способы генерации [1]:

а) поиск альтернатив в патентной и журнальной литературе;

б) привлечение нескольких экспертов, имеющих разную подготовку и опыт;

в) увеличение числа альтернатив за счет их комбинации, образования промежуточных вариантов между предложенными ранее;

г) модификация имеющейся альтернативы, т.е. формирование альтернатив, лишь частично отличающихся от известной;

д) включение альтернатив, противоположных предложенным, в том числе и «нулевой» альтернативы (не делать ничего, т.е. рассмотреть последствия развития событий без вмешательства системотехников);

е) интервьюирование заинтересованных лиц и более широкие анкетные опросы;

ж) включение в рассмотрение даже тех альтернатив, которые на первый взгляд кажутся надуманными;

з) генерирование альтернатив, рассчитанных на различные интервалы времени (долгосрочные, краткосрочные, экстренные).

При выполнении работы по генерированию альтернатив важно создать благоприятные условия для сотрудников, выполняющих данный вид деятельности. Большое значение имеют психологические факторы, влияющие на интенсивность творческой деятельности, поэтому необходимо стремиться к созданию благоприятного климата на рабочем месте сотрудников.

Существует еще одна опасность, возникающая при выполнении работ по формированию множества альтернатив, о которой необходимо сказать. Если специально стремиться к тому, чтобы на начальной стадии было получено как можно больше альтернатив, т.е. стараться сделать множество альтернатив как можно более полным, то для некоторых проблем их количество может достичь многих десятков. Для подробного изучения каждой из них потребуются неприемлемо большие затраты времени и средств. Поэтому в данном случае необходимо провести предварительный анализ альтернатив и постараться сузить множество на ранних этапах анализа. На этом этапе анализа применяют качественные методы сравнения альтернатив, не прибегая к более точным количественным методам. Тем самым осуществляется грубое отсеивание.

Приведем теперь методы, используемые в системном анализе, для проведения работы по формированию множества альтернатив.

Методы коллективной генерации идей

Методы коллективной генерации идей известны также, как методы мозгового штурма или мозговой атаки. Данный метод является методом систематической тренировки творческого мышления, нацеленным на открытие новых идей и достижение согласия группы людей на основе интуитивного мышления. Техника мозгового штурма состоит в следующем. Собирается группа лиц, отобранных для генерации альтернатив. Главный принцип отбора заключается в подборе специалистов разных профессий, опыта работы и квалификации. Данная группа обсуждает проблему, причем заранее оговаривается, что приветствуются любые идеи, возникшие как индивидуально, так и по ассоциации при выслушивании предложений других участников. Приветствуются даже идеи, лишь незначительно улучшающие высказывания предыдущих выступающих. При обсуждении придерживаются ряда правил:

• необходимо обеспечить как можно бо́льшую свободу мышления участников мозгового штурма и высказывания ими новых идей;

• допускается высказывание любых идей, даже если вначале они кажутся сомнительными и абсурдными;

• не допускается критика, не объявляется ложной и не прекращается обсуждение ни одной идеи;

• приветствуется высказывание как можно большего числа идей, особенно нетривиальных.

Разработка сценариев

В некоторых проблемах искомое решение должно описывать реальное поведение объекта в будущем, определять реальный ход событий. В таких случаях альтернативами являются различные последовательности действий и вытекающих из них событий, которые могут произойти с системой в будущем. Эти последовательности имеют общее начальное состояние и различные траектории движения развития системы. Это различие и приводит к проблеме выбора. Такие гипотетические альтернативные описания поведения системы в будущем называются **сценариями**. Сценарии-альтернативы – это логически обоснованные модели поведения проблемосодержащей системы в будущем, которые после принятия решения можно рассматривать как прогноз изменения состояний системы. Разработка сценариев относится к типичным неформализуемым процедурам. Для составления сценариев привлекаются специалисты, которые должны знать общие закономерности развития систем. При составлении сценариев проводят анализ внутренних и внешних факторов, влияющих на развитие системы, определяют источники этих факторов, целенаправленно анализируют высказывания ведущих специалистов в научных публикациях по рассматриваемой тематике. Сценарий является предварительной информацией, на основе которой проводится дальнейшая работа по прогнозированию развития системы. Сценарий помогает составить представление о проблеме; затем приступают к более тщательным, как правило, количественным процедурам анализа.

Морфологические методы

Основная идея морфологических методов состоит в систематическом переборе всех мыслимых вариантов решения проблемы или развития системы путем комбинирования выделенных элементов или их признаков. Системный аналитик определяет все мыслимые параметры, от которых может зависеть решение проблемы и представляет их в виде матриц-строк. Затем в этой матрице определяются все возможные сочетания параметров по одному из каждой строки. Полученные таким образом варианты подвергаются оценке и анализу с целью выбора наилучшего варианта решения проблемной ситуации. Методологию морфологического анализа можно проиллюстрировать на примере, приведенном в [19], касающемся разработки системы телевизионной связи. Рассмотрим табл. 2.1, которая порождает 384 различные возможные системы телевизионной связи. Современному телевизионному вещанию соответствует только одна альтернатива. Таким образом, анализируя данную таблицу, можно сказать, что у телевидения широкие возможности для дальнейшего развития.

Деловые игры

Деловыми играми называется имитационное моделирование реальных ситуаций. В процессе моделирования участники игры ведут себя таким образом, будто они в реальности выполняют порученную им роль.

Таблица 2.1

Независимая переменная	Значение переменной
Цвет изображения	Черно-белое Одноцветное Двухцветное Трехцветное .. Семицветное
Размерности изображения	Плоское изображение Объемное изображение
Градация яркости	Непрерывные Дискретные
Звуковое сопровождение	Без звука Монофонический звук Стереофонический звук
Передача запахов	Без передачи запахов С сопровождением запахов
Наличие обратной связи	Без обратной связи С обратной связью

Реальная ситуация в данном случае заменяется некоторой моделью. Чаще всего деловые игры используются для обучения, однако их с успехом применяют и для экспериментального генерирования альтернатив, особенно в слабо формализованных ситуациях. Важная роль в деловых играх отводится руководителю игры, тому, кто управляет моделью, регистрирует ход игры и обобщает ее результаты.

Методы экспертного анализа

Методы экспертного анализа разрабатывались для решения задачи структурирования и системной организации процесса получения и кодирования данных и знаний, источником которых является человек-эксперт. Методы экспертного анализа применяются для решения слабоформализованных задач. Суть методов состоит в подборе группы экспертов, являющихся специалистами в рассматриваемой области знаний. Перед ними формулируется задача, скажем, изложить свое мнение по проблеме, требующей решения, предложить пути развития системы, обосновать траекторию изменения состояний системы в будущем и т.п. После получения ответов появляется как бы коллективное мнение, коллективный взгляд на решаемую проблему. В результате обработки экспертных ответов получают наиболее вероятный прогноз по развитию системы.

Метод «Дельфи»

Метод «Дельфи» – итеративная процедура при проведении мозговой атаки, которая должна снизить влияние психологических факторов при проведении обсуждений проблемы и повысить объективность результатов. В отличие от традиционного подхода к достижению согласованности мнений экспертов путем открытой дискуссии метод «Дельфи» предполагает полный отказ от коллективных обсуждений. Это делается для того, чтобы уменьшить влияние таких психологических факторов как присоединение к мнению наиболее авторитетного специалиста, нежелание отказаться от публично выраженного мнения, следование за мнением большинства. В методе «Дельфи» прямые дебаты заменены тщательно разработанной программой последовательных индивидуальных опросов, проводимых в форме анкетирования. Ответы экспертов обобщаются и вместе с новой дополнительной информацией поступают в распоряжение экспертов, после чего они уточняют свои первоначальные ответы. Такая процедура повторяется несколько раз до получения приемлемой сходимости совокупности высказанных мнений.

Методы типа дерева целей

Метод типа дерева целей или дерева направлений прогнозирования подразумевает использование иерархической структуры, полученной путем разделения общей цели на подцели, а их, в свою очередь, на более детальные составляющие – новые подцели, функции и т.д. Древовидные иерархические структуры используются при исследовании вопросов совершенствования организационных систем.

Таким образом, рассмотрены методы, которые находят применение при решении задачи генерирования альтернатив. Важным моментом при решении данного вопроса является итеративность. Суть ее состоит в том, что на любой последующей стадии системного анализа должна быть возможность порождения новой альтернативы и включения ее в состав анализируемых. При рассмотрении слабоструктурированных проблем в качестве метода анализа используют следующий подход. Берут за основу одну подходящую альтернативу и производят ее пошаговое улучшение.

2.10. Реализация выбора и принятия решений

Целевое предназначение всего системного анализа состоит в том, чтобы в результате осуществить выбор. Выбор или принятие решения есть суть поставленной задачи системного анализа, конечный итог всей работы. Заказчик формулирует перед системным аналитиком проблему. Его интересуют прагматичные вопросы, например, сформулировать мероприятия, которые гарантировали бы быстрое развитие предприятия с обеспечением максимальной прибыли, или же предложить наилучшее решение по обеспечению стабильного электроснабжения некоторого региона. Системный аналитик должен ответить на вопрос: «Что лучше – строить новую электростанцию или провести модернизацию действующей, но выработавшей свой ресурс? Какова будет надежность электростанции после проведения работ по модернизации? Будет ли на допустимом уровне риск от ее эксплуатации?». Заказчика, в общем-то, не интересует, каким способом будет выработано то или иное решение. Для него важно, чтобы оно было обосновано и отвечало на поставленный вопрос.

Все описанные ранее этапы работ являлись предварительными, направленными на изучение проблемной ситуации. Для того, чтобы обоснованно подойти к решению задачи выбора анализируется система и строится ее модель, изучаются цели, которые ставит перед собой (и, естественно, системными аналитиками) заказчик, исследуются возможные пути развития системы, т.е. генерируются альтернативы. После столь тщательной проработки проблемной ситуации наступает завершающий этап – этап принятия решения. Процедура принятия решения представляет собой действие над множеством альтернатив, в результате которого получается подмножество выбранных альтернатив. Желательно, чтобы это была одна альтернатива. Сужение множества альтернатив возможно, если есть способ сравнения альтернатив между собой и определения наиболее предпочтительных. Для того чтобы имелась возможность сравнивать альтернативы, необходимо выработать критерий предпочтения. Проблема выбора сама по себе достаточно сложна. Она допускает существенно различающиеся математические постановки задач. Отметим основные сложности, возникающие при решении задач выбора и принятия решений:

- множество альтернатив может быть конечным, счетным или бесконечным;
- оценка альтернативы может осуществляться по одному или по нескольким критериям;

• критерии могут иметь количественное выражение или допускать только качественную оценку;

• режим выбора может быть однократным или повторяющимся, допускающим обучение на опыте;

• последствия выбора могут быть точно известны, иметь вероятностный характер или иметь неоднозначный исход, не допускающий введение вероятностей.

Различные сочетания перечисленных вариантов приводят к многообразным задачам выбора. Для решения задач выбора предлагаются различные подходы, наиболее распространенный из которых – критериальный подход. Основным предположением критериального подхода является следующее: каждую отдельно взятую альтернативу можно оценить конкретным числом – значением критерия. Критерии, на основе которых осуществляется выбор, имеют различные названия – критерий качества, целевая функция, функция предпочтений, функция полезности и т.д. Объединяет их то, что все они служат решению одной задачи – задачи выбора.

Сравнение альтернатив сводится к сравнению результатов расчетов соответствующих критериев. Если далее предположить, что выбор любой альтернативы приводит к однозначно определяемым последствиям и заданный критерий численно выражает оценку этих последствий, то наилучшей альтернативой является та, которая обладает наибольшим значением критерия. Задача поиска наилучшей альтернативы, простая по постановке, часто оказывается сложной для решения, поскольку метод ее решения определяется размерностью и типом множества альтернатив, а также видом критериальной функции. Однако на практике сложность отыскания наилучшей альтернативы многократно возрастает, так как оценивание вариантов приходится проводить на основании нескольких критериев, качественно различающихся между собой. Если в результате сравнения по нескольким критериям получилось, что одна альтернатива обладает наилучшими значениями по всем критериям, то выбор не представляет затруднений, именно эта альтернатива и будет наилучшей. Однако такая ситуация встречается лишь в теории. На практике дело обстоит куда как сложнее. В данной ситуации приходим к необходимости решения многокритериальных задач. Подходы к решению таких задач известны – это метод сведения многокритериальной задачи к однокритериальной, метод условной максимизации, поиск альтернативы с заданными свойствами, нахождение паретовского множества альтернатив. Выбор альтернативы на основании критериального подхода предполагает, что выполненными являются несколько условий: известен критерий, задан способ сравнения вариантов и метод нахождения лучшего из них. Однако этого оказывается недостаточно. При решении задач выбора необходимо учитывать условия, при которых осуществляется выбор, и ограничения задачи, так как их изменение может привести к изменению решения при одном и том же критерии.

Оптимизационный подход нашел широкое применение в задачах системного анализа. Это обусловлено тем, что понятие оптимальности получило строгое и точное представление в математических теориях. Оптимизационный подход прочно вошел в практику проектирования и эксплуатации технических систем, сыграл важную роль в формировании современных системных представлений, широко используется в административном управлении. Нахождение оптимальных вариантов особенно важно для оценки состояния современной техники и определения перспектив ее развития. Знание параметров оптимальной альтернативы позволяет составить представление о принципиально не улучшаемых возможностях технических объектов. Сравнение с оптимальными параметрами помогает решить вопрос о целесообразности дальнейших усилий по улучшению показателей качества изделий. Однако у оптимизационного подхода есть свои ограничения, требующие внимательного и осторожного обращения с ним. Остановимся на особенностях, накладывающих ограничения на применение оптимизационного подхода и требующих учета при решении задач принятия решений.

1. Оптимальное решение часто оказывается чувствительным к изменению условий задачи. Следует учитывать, что иногда такие изменения могут привести к выбору существенно отличающихся альтернатив.

2. Обычно система, для которой принимается решение, входит в структуру более общей системы, т.е. является ее подсистемой, и решения оптимальные для этой подсистемы, могут входить в противоречие с целями надсистемы; т.е. возникает необходимость увязывать критерии подсистем с критериями надсистем.

3. Необходимо очень тщательно и скрупулезно подходить к выбору и обоснованию критерия. Критерий должен выбираться из анализа цели исследования; при этом надо помнить, что он характеризует цель лишь косвенно, иногда хуже, иногда лучше, но всегда приближенно.

4. Помимо критериев в оптимизационной задаче немаловажную роль играют ограничения. Анализ существа проблемной ситуации и качественное обоснование ограничений задачи имеют значительное влияние на принимаемое решение. Нередко даже небольшие изменения в ограничениях отражаются на принимаемом решении. Еще больший эффект получается, когда одни ограничения заменяются другими. Не задав всех необходимых ограничений, можно одновременно с оптимизацией основного критерия получить непредвиденные и нежелательные эффекты.

Таким образом, можно сделать вывод о том, что оптимизация – это мощное средство повышения эффективности, но использовать его необходимо осторожно, особенно при работе со сложными проблемными ситуациями. Проблема еще более обостряется, когда речь идет о принятии решений в организационных или социальных системах. Можно констатировать, что оптимизационные задачи, которые удается поставить при исследовании сложных систем, имеют обоснованный характер, если описывают хорошо структурированные системы, и являются заведомо приближенными, если относятся к системе в целом. Поэтому отметим, что оптимизационный подход является не единственным при решении задач выбора и принятия решений. Существуют другие методы, которые дополняют оптимизационный выбор. Одним из таких методов является экспертный. Он применяется в тех случаях, когда при исследовании сложных систем возникают проблемы, которые не удается представить в виде формальных математических задач. В таких случаях прибегают к услугам экспертов – лиц, чья интуиция и опыт могут уменьшить сложность проблемы. И наконец, для решения задач выбора в сложных проблемных ситуациях создаются специальные человеко-машинные, проблемно-ориентированные системы. Системы поддержки решений ориентированы не на автоматизацию функций лица, принимающего решение, а на предоставление ему помощи в проведении данной работы.

Подводя итог, отметим, что проблема выбора и принятия решений – центральная проблема системного анализа. Налицо сложности, которые возникают перед системным аналитиком. Но, с другой стороны, имеется развитый математический и эвристический аппарат, который является мощным оружием, помогающим обоснованно подходить к проблеме выбора

2.11. Внедрение результатов анализа

Системный анализ является прикладной наукой, его конечная цель – изменение существующей ситуации в соответствии с поставленными целями. Окончательное суждение о правильности и полезности системного анализа можно сделать лишь на основании результатов его практического применения. Конечный результат будет зависеть не только от того, насколько совершенны и теоретически обоснованы методы, применяемые при проведении анализа, но и от того, насколько грамотно и качественно реализованы полученные рекомендации.

В настоящее время вопросам внедрения результатов системного анализа в практику уделяется повышенное внимание. В этом направлении можно отметить работы Р. Акоффа [23] и П. Чеклэнда [24]. Следует заметить, что практика системных исследований и практика внедрения их результатов существенно различаются для систем разных типов. Согласно классификации, введенной П. Чеклэндом, системы делятся на три типа: естественные, искусственные и социотехнические. В системах первого типа связи образованы и действуют природным образом. Примерами таких систем могут служить экологические, физические, химические, биологические и т.п. системы. В системах второго типа связи образованы в результате человеческой деятельности. Примерами могут служить всевозможные технические системы. В системах третьего типа, помимо природных связей, важную роль играют межличностные связи. Такие связи обусловлены не природными свойствами объектов, а культурными традициями, воспитанием участвующих в системе субъектов, их характером и прочими особенностями.

Системный анализ применяется для исследования систем всех трех типов. В каждой из них есть свои особенности, требующие учета при организации работ по внедрению результатов. Наиболее велика доля слабоструктурированных проблем в системах третьего типа. Следовательно, наиболее сложна практика внедрения результатов системных исследований в этих системах.

При внедрении результатов системного анализа необходимо иметь в виду следующее обстоятельство. Работа осуществляется на клиента (заказчика), обладающего властью, достаточной для изменения системы теми способами, которые будут определены в результате системного анализа. В работе должны непосредственно участвовать все заинтересованные стороны. Заинтересованные стороны – это те, кто отвечает за решение проблемы, и те, кого эта проблема непосредственно касается. В результате внедрения системных исследований необходимо обеспечить улучшение работы организации заказчика с точки зрения хотя бы одной из заинтересованных сторон; при этом не допускаются ухудшения этой работы с точки зрения всех остальных участников проблемной ситуации.

Говоря о внедрении результатов системного анализа, важно отметить, что в реальной жизни ситуация, когда сначала проводят исследования, а затем их результаты внедряют в практику, встречается крайне редко, лишь в тех случаях, когда речь идет о простых системах. При исследовании социотехнических систем они изменяются с течением времени как сами по себе, так и под влиянием исследований. В процессе проведения системного анализа изменяются состояние проблемной ситуации, цели системы, персональный и количественный состав участников, соотношения между заинтересованными сторонами. Кроме того, следует заметить, что реализация принятых решений влияет на все факторы функционирования системы. Этапы исследования и внедрения в такого типа системах фактически сливаются, т.е. идет итеративный процесс. Проводимые исследования оказывают влияние на жизнедеятельность системы и это видоизменяет проблемную ситуацию, ставит новую задачу исследований. Новая проблемная ситуация стимулирует дальнейшее проведение системного анализа и т.д. Таким образом, проблема постепенно решается в ходе активного исследования.

Глава 3
ПОСТРОЕНИЕ МОДЕЛЕЙ СИСТЕМ

3.1. Понятие модели системы

Центральным понятием системного анализа является понятие системы. При описании процедуры проведения системного анализа было отмечено, что одной из составных частей этого процесса является формализация описания системы, т.е. построение ее модели.

Понятие модели системы играет важную роль в проведении системных исследований любой направленности. Модель – это искусственно создаваемый образ конкретного объекта, процесса или явления, в конечном счете, любой системы. Понятие модели связано с наличием какого-либо сходства между выбранными объектами, один из которых является оригиналом, а другой – его образом, выполняющим роль модели. Модели являются всегда упрощенным описанием системы. Модель – это отображение реальной системы (оригинала), имеющее определенное объективное соответствие ей и позволяющее прогнозировать и исследовать ее функциональные характеристики, т.е. характеристики, определяющие взаимодействие системы с внешней средой. При составлении модели отражают отдельные стороны функционирования системы, т.е. то специфичное, что направлено на решение поставленной целевой установки общей задачи системного анализа. Сходство двух объектов с точки зрения выполнения каких-либо функций, целей или задач позволяет утверждать, что между ними существует отношение оригинала и модели. В задачах системного исследования первоочередной интерес представляет сходство поведения модели и объекта, выраженное на каком-либо формальном языке и изучаемое путем преобразований соответствующих формул или высказываний. Так приходим к понятию математической модели, являющейся основой аналитических исследований и имитационных экспериментов на ЭВМ. Математические модели можно классифицировать таким же образом, как это было проделано в случае классификации систем. Остановимся на описании классов, имеющих принципиально различный характер в подходе к построению моделей, а именно, охарактеризуем следующие типы моделей: детерминированные, вероятностные и игровые модели.

Детерминированные модели описывают поведение систем с позиций полной определенности состояний системы в настоящем и будущем. Примерами таких моделей являются описания физических закономерностей, формулы, описывающие взаимодействие химических веществ, программы обработки деталей и т.д. Детерминированный подход находит применение при решении задач планирования транспортных перевозок, при составлении расписаний, планировании и распределении ресурсов, в задачах материально-технического снабжения, в планировании производства.

Вероятностные модели описывают поведение системы в условиях воздействия случайных факторов. Следовательно, такие модели оценивают будущие состояния системы с позиций вероятностей реализации тех или иных событий. Примерами вероятностных моделей являются описание времени ожидания, обслуживания или длины очереди в системах массового обслуживания, модели расчета надежности системы, модели определения риска от наступления нежелательного события и пр.

Игровые модели дают возможность изучать конфликтные ситуации, в которых каждая из конфликтующих сторон придерживается своих взглядов, и характер поведения каждой из них диктуется личными интересами. Примерами таких систем являются отношения двух или нескольких производителей одинакового товара. Их поведение на рынке обусловлено интересами каждой из сторон. Как правило, эти отношения имеют характер конкурентной борьбы.

Широкое применение математических моделей в задачах системного анализа обусловлено универсальностью подхода к анализу как систем в целом, так и явлений и процессов, происходящих в них, способностью отразить все разнообразие закономерностей их развития и поведения. При применении математического моделирования появляется возможность проведения глубокого анализа задачи, обнаружения ошибок и корректировки исходных постулатов. При этом затраты на проведение исследований существенно меньше по сравнению с аналогичными исследованиями на реальных объектах. Если к тому же учесть, что ряд исследований на реальных объектах провести нет возможности либо по причине физической нереализуемости, либо ввиду больших материальных затрат, либо ввиду нежелательных последствий, наступаемых в результате завершения исследований, то становится понятным, что исследование на математических моделях является чуть ли не единственным способом решения поставленных задач. Понятна нежелательность, мягко говоря, проведения натурных испытаний по установлению причин, приводящих к авариям на атомных электростанциях. Такие исследования проводят исключительно на моделях.

При составлении моделей проявляются знания, опыт, интуиция и квалификация системных аналитиков. Создание модели требует четких представлений о роли моделируемых систем в решении поставленной задачи системного анализа, об их особенностях, о предполагаемом использовании результатов системных исследований. Математические модели могут иметь вид формул, систем уравнений или неравенств, логических выражений, графических образов, отражающих зависимость между выходными параметрами, состояниями системы, входными параметрами и управляющими воздействиями. Анализируемая система может быть описана разными моделями, каждая из которых обладает характерными свойствами и пригодна для решения лишь определенного круга задач, относящихся к структуре и функционированию системы. Рассмотрим основные виды моделей систем и способы их построения.

3.2. Способы описания систем

Модель черного ящика

Наиболее простой, грубой формой описания системы является представление ее в виде черного ящика, которое имеет следующие особенности. Во-первых, такое представление не раскрывает внутренней структуры, внутреннего устройства системы. Оно лишь выделяет систему из окружающей среды, подчеркивает ее целостное единство. Во-вторых, такое представление говорит также о том, что система хотя и является обособленной, выделенной из среды, но, тем не менее, она не является изолированной от нее. Действительно, планируемая цель пред-

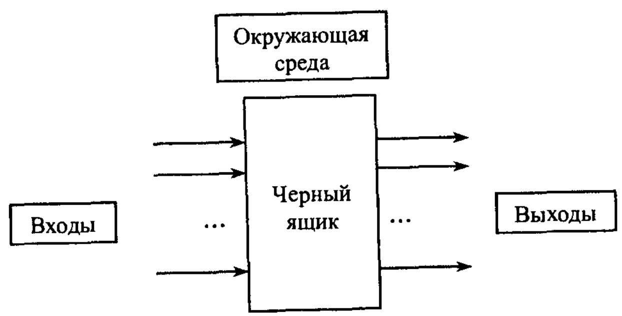

Рис. 3.1. Представление системы в виде черного ящика

полагает, что в конечном итоге будут произведены изменения в системе, которые будут оказывать воздействия на внешнюю среду. Любая система работает на какого-либо внешнего потребителя. Иными словами система связана со средой и с помощью этих связей воздействует на среду. Таким образом, можно заключить, что у системы есть *выходы*. Выходы системы отражают ее целевое предназначение. С другой стороны, система является средством, с помощью которого достигаются те или иные цели. Следовательно, должны существовать возможности воздействия на систему, управления системой. Эти связи направлены из среды в сторону системы. Такие воздействия называются *входами* системы. В результате такого представления получилась модель системы, которая называется черным ящиком (рис 3.1). Это название подчеркивает полное отсутствие сведений о внутреннем содержании модели. В данном представлении задаются только связи модели с внешней средой в виде входных и выходных воздействий. Такая модель, несмотря на внешнюю простоту, бывает полезной для решения определенного круга задач.

В модели черного ящика входы и выходы могут иметь качественное, словесное описание. Тогда и сама модель будет качественной. В реальных ситуациях для построения модели требуется количественное описание входов и выходов. В этом случае формируются множества входных – $\mathbf{X}$ и выходных – $\mathbf{Y}$ переменных.

В общем виде математическое описание исследуемой системы может быть выражено зависимостью

$$\{\mathbf{Y}\} = \Phi\big[\{\mathbf{X}\},\{\mathbf{Z}\},\{\mathbf{V}\}\big],$$

где $\{\mathbf{Y}\} = (Y_1, Y_2, \ldots, Y_l)$ – множество векторов выходных переменных системы. В качестве выходных переменных, как правило, используются критерии, отражающие цели исследования. Под критерием понимают целевые функции, параметры оптимизации и т.д. В общем случае множество входных переменных подразделяют на три класса: $\{\mathbf{X}\}=(X_1, X_2, \ldots, X_m)$ – множество векторов входных контролируемых управляемых независимых переменных (факторов), действующих на процессы; $\{\mathbf{Z}\}=(Z_1, Z_2, \ldots, Z_k)$ – множество векторов входных контролируемых, но неуправляемых независимых переменных; $\{\mathbf{V}\} = (V_1, V_2, \ldots, V_n)$ – множество векторов неконтролируемых возмущающих воздействий; Φ – оператор системы, определяющий связь между указанными величинами. Отметим, что в модели черного ящика оператор системы, определяющий связь между указанными величинами, не исследуется.

Модель черного ящика является начальным этапом изучения сложных систем. На первых этапах проведения системных исследований

необходимо задать, сформировать множество входных и выходных параметров системы. Задача формирования множества параметров применительно к рассмотрению сложных систем сама по себе является непростой задачей. Сложная система имеет множественные и разнообразные связи с внешней средой. Чтобы избежать ошибки на этом этапе, необходимо сформулировать одно правило, гласящее, что в модель следует отбирать только те входы и выходы, которые отражают целевое назначение модели. Дело в том, что реальная система взаимодействует с объектами окружающей среды неограниченным числом способов. Задача системного аналитика при построении модели состоит в том, чтобы из бесчисленного множества связей отобрать конечное их число для включения в список входов и выходов. Критерием отбора является существенность той или иной связи по отношению к цели, ради достижения которой строится модель.

Рассмотрим пример, иллюстрирующий процедуру отбора входов и выходов для достаточно простой системы. Пусть требуется разработать модель черного ящика для калькулятора. Основное целевое предназначение калькулятора – производить расчеты согласно заданной программе. Для того чтобы формировать задание для расчетов, калькулятору необходимы кнопки, причем двух типов. С одной стороны, с помощью кнопок будут вводиться цифры, с другой стороны, с помощью кнопок будут задаваться операции. Для отображения результатов расчетов необходимо табло. С точки зрения выполнения калькулятором основных функций отмеченных входов и выходов оказывается достаточно. Добавим теперь к целям использования калькулятора возможность автономной работы. Это повлечет за собой требование обеспечения его батареями питания. Выдвинув цель, состоящую в том, что калькулятор должен быть удобным и транспортабельным, получим новые требования на вес и габариты устройства. Выдвинув в качестве цели удобство и наглядность чтения результатов с табло, получим требования на характеристики табло. Рассматривая условия эксплуатации калькулятора,можно добавить требования надежности работы и прочности корпуса. Далее можно еще расширить круг учитываемых требований и добавить соответственно несколько выходов, таких как эстетичный внешний вид, соответствие цены покупательной способности потребителя. Далее можно говорить о возможности работы с калькулятором в условиях плохой видимости, например, в темноте. Тогда следует рассмотреть необходимость подсветки циферблата, различение кнопок на ощупь и т.д. Данный ряд требований можно продолжать, рассматривая химические, физические, социальные, экономические аспекты. Таким образом, в зависимости от выдвигаемых целей и формулируемых требований, предъявляемых к калькулятору, будет различный набор входов и выходов данной системы.

Следовательно, можно сделать вывод о том, что построение модели черного ящика не является тривиальной задачей. Вопрос о том, сколько именно и какие входы и выходы необходимо включать в модель, не имеет простого и однозначного ответа. С одной стороны, выполнение только основной цели недостаточно, необходимо учитывать дополнительные цели. С другой стороны, встает вопрос, сколько дополнительных целей необходимо учитывать, где требуется остановиться. Критерия не существует. Здесь выбор полностью ложится на исследователя и зависит от его опыта и компетенции.

Модель состава системы

В том случае, когда системного аналитика интересуют вопросы внутреннего устройства системы, модели черного ящика оказывается недостаточно. Для решения данного вопроса необходимо разрабатывать более детальные, более развитые модели. Одной из разновидностей таких моделей, раскрывающей внутреннее содержание системы, является модель состава системы.

Свойства системы, отображенные в модели черного ящика, целостность и обособленность, являются внешними свойствами. Внутреннее содержание системы в модели черного ящика не рассматривается. Но именно внутреннее содержание ящика оказывается неоднородным. Как было отмечено в п. 2.2, в структуре системы можно выделить различные элементы, подсистемы, компоненты системы, причем, обозначенные понятия условны. В зависимости от цели, для решения которой строится модель, один и тот же объект может быть определен и в качестве элемента, и в качестве подсистемы.

Рассмотрим пример, поясняющий данную мысль. Перед системным аналитиком сформулирована цель – проведение расчета характеристик надежности некоторой системы. Предположим, что систему можно разбить на подсистемы, подсистемы на блоки, а блоки, в свою очередь, на элементы. Вернемся к примеру п. 2.2 и будем анализировать систему управления и защиты энергоблока атомной станции. В данной системе был выделен ряд подсистем: автоматического регулирования, ручного регулирования, аварийной защиты, аварийной и предупредительной сигнализации и т.д. В структуре каждой подсистемы можно выделить блоки: блок питания, датчик (счетчик нейтронов), устройство отображения информации, вторичные приборы, исполнительные механизмы и т.д. В свою очередь, в каждом из блоков выделяют электронные

(транзисторы, диоды, конденсаторы, сопротивления), электромеханические (реле, лентопротяжные механизмы) элементы, кабели, тросы и пр. Задача аналитика состоит в выработке решения, на каком уровне разбиения объектов необходимо остановиться. Здесь требуется вернуться к цели и наметить пути ее достижения.

К реализации сформулированной цели можно подойти с разных позиций. Рассмотрим первый возможный подход. Система, для которой проводится анализ надежности, длительное время находилась в эксплуатации; имеется представительная статистика поведения объектов системы в процессе ее функционирования. В результате анализа надежности ставится задача максимально использовать опыт эксплуатации, так как именно в процессе эксплуатации реализуются свойства системы, и поэтому результаты расчета, полученные на основании статистической информации, наиболее реально отражают объективные закономерности. При анализе эксплуатационной информации выяснилось, что в службах предприятия на каждый блок имеются паспорта, в которых отмечается вся история поведения блока. Таким образом, делается вывод о том, что в процессе обработки информации можно получить характеристики надежности блоков и далее, используя эту информацию в качестве исходной, проводить дальнейшие расчеты надежности системы. Следовательно, при построении модели состава системы в качестве первичного элемента достаточно ограничиться блоками, для которых рассчитывается исходная информация.

Другой возможный случай. Система, для которой проводится анализ надежности, в эксплуатации не находилась. В качестве исходной информации для проведения расчетов имеется справочная информация об интенсивностях отказов элементов типа транзисторов, диодов, конденсаторов, сопротивлений, реле, лентопротяжных механизмов и т.д. В этом случае делается вывод о том, что при составлении модели состава системы необходимо проводить декомпозицию каждого блока на составляющие элементы. Полученная таким образом модель будет более детальной.

Таким образом, в зависимости от цели исследования, постановки задачи по достижению данной цели и исходной информации, имеющейся для решения задачи, одну и ту же систему следует представлять в виде различных частей, различных иерархий. Далее следует отметить, что условным является также разбиение на подсистемы. В той же системе управления и защиты одни и те же элементы, в зависимости от решаемой задачи, могут быть отнесены к разным подсистемам, т.е. границы между подсистемами условны. Проиллюстрируем это примером. В системе имеется группа стержней автоматического регулирования, которые в штатной ситуации осуществляют регулирование мощности энергоустановки. Следовательно, для нормального режима эксплуатации эти стержни следует отнести к подсистеме автоматического регулирования. Однако регламент работы предусматривает, что в случае возникновения аварийной ситуации данные стержни принимают участие в глушении реакторной установки, в переводе ее функционирования на подкритический уровень. Таким образом, при рассмотрении аварийных режимов стержни автоматического регулирования следует рассматривать как элементы подсистемы аварийной защиты.

И еще одна особенность, которую необходимо рассматривать при составлении модели состава системы, – это неоднозначность границ между системой и окружающей средой. При работе системы управления и защиты для ее нормального штатного функционирования необходимо задать соответствующие программы: какой уровень мощности считать номинальным, какие отклонения от данного уровня являются допустимыми, а какие требуют вмешательства системы, на какие ситуации следует реагировать как на аварийные и осуществлять при этом глушение реакторной установки. Данные программы формируются внешними по отношению к технической системе органами. Включать их в состав системы или нет, также зависит от цели, для решения которой строится модель состава системы.

Подводя итог, можно отметить, что границы между системой и внешней средой определяются целями построения модели и не имеют абсолютного характера. Таким образом, модель состава ограничивается снизу теми объектами, которые приняты в качестве элементов, а сверху – границей системы, определяемой целями анализа

Модель структуры системы

Следующий тип модели, который еще глубже характеризует внутреннюю композицию системы, называется моделью структуры системы. Модели данного типа наряду с характеристикой состава системы отражают взаимосвязи между объектами системы: элементами, частями, компонентами и подсистемами. Таким образом, модель структуры системы является дальнейшим развитием модели состава. Для того чтобы отразить композицию системы, недостаточно перечислить ее состав; необходимо установить между элементами определенные связи, отношения.

При рассмотрении модели структуры системы приходится сталкиваться с аналогичными особенностями, о которых уже частично шла речь ранее, а именно, анализируя реальные системы, можно констатировать, что между объектами, входящими в состав системы, имеется большое количество отношений. В любой структуре реализуется бесконечность природы. Отношения между элементами могут быть самыми разнообразными. Однако можно попытаться их классифицировать и по возможности перечислить. Трудность состоит в том, что заранее не известно, какие отношения реально существуют, и является ли их число конечным. Задача аналитика заключается в следующем: из множества реально существующих отношений между объектами, вовлеченными в систему, отобрать наиболее существенные. Критерием существенности отношений должна выступать опять же цель, для достижения которой строится модель. Таким образом, модель структуры является очередным шагом в развитии модели систем, описывающей существенные связи между элементами.

Развивая модели описания системы от модели черного ящика до модели структуры, приходим к описанию системы в виде структурной схемы. Структурная схема отражает, как правило, статическое состояние системы. В ней указываются все существенные с точки зрения выполнения поставленной цели элементы системы, все связи между элементами внутри системы и связи с окружающей средой – то, что названо входами и выходами. Для изображения структурной схемы абстрагируются от содержательной стороны схемы, оставив в рассматриваемой модели только общее для каждой схемы. В результате получается модель, в которой отмечено только наличие элементов и связей между ними. Как было отмечено в п. 2.2., для такого представления используют изображение в виде графа. На графе элементы отображаются в виде вершин, связи между элементами – в виде дуг. Если связи в схеме направленные, они изображаются стрелками, и тогда граф будет направленным или ориентированным. Если направление связей не обозначается, граф называется неориентированным. Для изображения и преобразования структур разработана специальная математическая дисциплина – теория графов, задачи которой связаны с различными преобразованиями графов, с рассмотрением различных отношений на графах. Отношения могут быть отражены в виде весовых характеристик, рангов, вероятностных характеристик и т.п.

Таким образом, еще раз отметим, что структурная схема системы является наиболее подробной моделью, отражающей статическое состояние системы. Однако для решения задач системного анализа статические структуры имеют важное, но, как правило, вспомогательное значение. Большинство задач системного анализа связано с изучением либо характеристик системы, либо с прогнозированием развития системы во времени, либо с анализом возможных траекторий развития и т.п. Короче говоря, цели большинства задач системного анализа связаны с изучением динамики системы, ее динамического поведения. В этом случае появляется необходимость построения новых моделей – динамических.

Динамические модели систем

Динамические модели отражают поведение систем, описывают происходящие с течением времени изменения, последовательность операций, действий, причинно-следственные связи. Системы, в которых происходят какие бы то ни было изменения со временем, называются *динамическими*, а модели, отображающие эти изменения, – *динамическими моделями* систем.

Говоря о динамике систем, следует остановиться на двух типах динамических процессов – это функционирование и развитие. Под функционированием понимают процессы, которые происходят в системе, стабильно реализующей фиксированную цель. Развитием называют изменения, происходящие с системой при смене ее целей. Характерной чертой развития является то обстоятельство, что изменение цели, как правило, с неизбежностью приводит к изменению всей системы. Это касается либо изменения структуры, либо изменения состава системы, иногда приходится проводить коренную перестройку системы. Таким образом, при построении динамических моделей на первом шаге анализируют тип отображаемого изменения системы, который хотят описать. Далее приступают к анализу происходящих изменений с целью более конкретного отображения динамики анализируемых процессов. На этом этапе вычленяют части, этапы происходящего процесса, рассматривают их взаимосвязь.

Заключительный этап построения динамической модели системы состоит в более глубокой формализации процессов, иными словами, в построении математического описания анализируемых процессов. При построении модели черного ящика был записан функционал, отображающий зависимость выхода системы от ее входов, в виде

$$\{Y\} = \Phi\left[\{X\},\{Z\},\{V\}\right].$$

Было отмечено, что в модели черного ящика характер зависимости или вид функционала не исследуется. Решение этого вопроса является задачей настоящего этапа.

Для построения математической модели динамического поведения системы вводится понятие состояния системы. *Состояние системы* есть некоторая внутренняя характеристика системы, значение которой

в настоящий момент времени определяет значение выходной величины. Состояние можно рассматривать как некий информационный объект, необходимый для предсказания влияния настоящего на будущее. Состояние есть множество **Z**. Конкретизируя множества **X**, **Y** и **Z**, а также отображения множества входов и состояний на множество выходов, можно перейти к моделям различных систем. Если ввести время как зависимую переменную, то получим два разных типа систем: дискретные и непрерывные. Примерами дискретных систем являются цифровые устройства: измерительные, вычислительные, управляющие. Примерами непрерывных систем являются производственные системы, аналоговые вычислительные машины и др., т.е. объекты, в которых не проводится дискретизация времени.

В зависимости от вида оператора отображения **Ф** различают линейные и нелинейные системы. Выделяют также класс стационарных систем, т.е. систем, свойства которых со временем не изменяются.

И наконец, говоря о динамических моделях, следует остановиться на подчиненности реальных систем принципу причинности. Согласно этому принципу, отклик системы на некоторое воздействие не может начаться раньше самого воздействия. Строя математическую модель системы, необходимо следовать сформулированному принципу. Дело в том, что в практике построения моделей встречаются ситуации, когда данный принцип игнорируется. В этом случае возникает ситуация, когда теоретические модели не могут быть реализованы на практике. Задача, стоящая перед исследователями при построении динамических моделей, – выяснение условий физической реализуемости теоретических моделей. Для обеспечения физической реализуемости требуется проводить тщательный анализ конкретных ограничений, которые приходится накладывать на модель.

3.3. Анализ и синтез – методы исследования систем

В предыдущем параграфе рассмотрены модели систем, которые строят с целью проведения системного анализа. Построение модели системы – это процесс формализации ее описания, которая достигается за счет изучения системы и в некоторой степени за счет упрощения ее реальных структур, связей и отношений. Остановимся на методах, позволяющих проводить исследования систем для дальнейшего построения их моделей. В качестве таких методов первостепеиное значение имеют анализ и синтез. Указанные методы противоположны друг другу по смыслу, так как анализ есть совокупность операций разделения целого на части, в то время как синтез – объединение частей в целое. Однако в приложении к решению задач исследования сложных систем применение этих методов автономно невозможно. Исследовать сложную систему можно только используя два указанных метода в совокупности. В применении анализа и синтеза к исследованию сложных систем проявляется диалектический принцип единства и борьбы противоположностей. Изложим суть каждого из подходов.

Аналитический метод состоит в расчленении сложного целого на все менее сложные части. Кроме того, он также предполагает, что части снова образуют единое целое в случае их соединения надлежащим образом. Этот момент соединения частей в целое является конечным этапом анализа, так как только после этого появляется возможность объяснить целое через его части и представить результат анализа в виде структурной схемы целого. С другой стороны, роль синтеза не сводится только к сборке частей, полученных при анализе. Необходимо подчеркнуть целостность системы, которая нарушается при анализе; при разбиении системы утрачиваются не только существенные свойства самой системы, но исчезают и существенные свойства ее частей, оказавшихся отделенными от нее. Результатом анализа является лишь вскрытие структуры; знание о том, как система работает, ответ на вопрос, почему она это делает так, дает синтез. Приведем цитату из работы Р. Акоффа [25] «Синтетическое мышление требует объяснить поведение системы. Оно существенно отличается от анализа. На первом шаге анализа вещь, подлежащая объяснению, разделяется на части; в синтетическом мышлении она должна рассматриваться как часть большего целого. На втором шаге анализа объясняются содержимые части; в синтетическом мышлении объясняется содержащее нашу вещь целое. На последнем шаге анализа знание о частях агрегируется в знание о целом; в синтетическом мышлении понимание содержащего целого дезагрегируется для объяснения частей. Это достигается путем вскрытия их ролей или функций в целом. Синтетическое мышление открывает не структуру, а функцию; оно открывает, почему система работает так, а не то, как она делает это». Подводя итог изложенному в данной цитате материалу, можно сказать, что аналитический метод не возможен без синтеза, так как синтез позволяет провести агрегирование частей в структуру, но и синтетический метод невозможен без анализа, так как необходима дезинтеграция целого для объяснения функций частей. Анализ и синтез дополняют, но не заменяют друг друга. Проведение системного анализа требует совмещения обоих указанных методов.

Автономное применение аналитического метода возможно лишь в тех случаях, когда систему удается разделить на не зависимые друг от друга части. Действительно, в этом случае отдельное рассмотрение частей позволяет составить правильное представление об их вкладе в общий эффект функционирования системы. Однако в реальной жизни найти систему, которая бы являлась результатом арифметического объединения своих компонентов, практически невозможно. Если удастся привести пример такой системы, то это будет удачное исключение, а не правило. В реальных ситуациях все части системы функционируют взаимосвязано; вклад одной части в общесистемный эффект зависит от вкладов других частей, т.е. можно констатировать, что реальные системы неаддитивны. В теории проектирования сложных систем проявление свойства неаддитивности хорошо известно. Оно проявляется в следующем. При проектировании систем формулируют условие оптимальности проекта. Но поскольку проект создается для сложной системы, постановка математической задачи проектирования имеет очень большую размерность и не решается путем проведения прямой оптимизации. В этом случае проводят разбиение системы на подсистемы и пытаются решить задачу оптимального проектирования подсистем с последующим объединением проектных решений. Оказывается, что в силу свойства неаддитивности систем созданный таким образом проект системы не будет обладать свойствами оптимальности. В ряде случаев спроектированная таким способом система будет просто неработоспособна. Попытка создать оптимальную систему из оптимальных частей, таким образом, не достигает успеха. Следовательно, при исследовании системы, ее динамики, свойств и, в конечном итоге, при построении модели системы требуется наряду с разбиением системы на части исследовать взаимодействие частей. Методом решения второй задачи является синтез.

Аргументом в пользу совместного применения аналитических и синтетических методов исследования систем является необходимость установления всесторонних отношений между рассматриваемыми явлениями. Описывать с помощью моделей можно только познанные явления, а таковыми они будут лишь тогда, когда известна совокупность условий, необходимых и достаточных для их реализации. Аналитические методы устанавливают только причинно-следственные отношения. В этом случае из рассмотрения исключаются все другие виды взаимодействия, в том числе и взаимодействия с окружающей средой. Установление причины для некоторого следствия означает определение необходимых условий для реализации этого события. Но кроме необходимых, коренных событий есть большое количество необходимых и достаточных условий, которые не выявляются аналитическими методами, и, следовательно, эти методы не гарантируют полную картину описываемых взаимодействий. В качестве примера, иллюстрирующего данные высказывания, можно привести систему «семя–растение». С точки зрения аналитических методов, устанавливающих причинно-следственные отношения, для того чтобы получить растение, необходимо иметь семя. Но аналитические методы не учитывают, что для произрастания растения необходима еще почва, вода, воздух, тепло, свет и т.д. Подобную полноту исследований гарантируют синтетические методы. Использование синтетических методов основано на признании того, что отношение «причина–следствие» является не единственно возможным и приемлемым описанием взаимодействия. Имеется еще масса необходимых и достаточных условий, требующих учета при рассмотрении наступления некоторых явлений или процессов, без учета которых модель будет не только неадекватной, но и неработоспособной.

3.4. Декомпозиция – метод математического описания систем

Основной операцией анализа является представление целого в виде частей. При решении задач системных исследований объектами анализа являются системы и цели, для достижения которых они проводятся. В результате анализа решаемые системой задачи разбиваются на подзадачи, системы на подсистемы, цели на подцели. Этот процесс разбиения продолжается до тех пор, пока не удастся представить соответствующий объект анализа в виде совокупности элементарных компонентов. Операция разложения целого на части называется *декомпозицией*. Обычно объект системного анализа сложен, слабо структурирован, плохо формализован, поэтому операция декомпозиции представляет собой также плохо формализованный процесс, сложный для выполнения. Обычно декомпозицию проводят высококвалифицированные эксперты, имеющие богатый опыт работы в данной области.

Итак, задача системного аналитика при построении модели системы заключается в разделении сложной системы на подсистемы. Аналогично целевая функция объекта должна быть представлена в виде последовательности подцелей, задач, функций, операций, выполнение которых ведет к достижению глобальной цели системного исследования. Далее желательно каждой подсистеме поставить в соответствие некоторую подцель (задачу, функцию, операцию) и наоборот. В этом и заключается смысл декомпозиции. Необходимость таких действий

обусловлена тем, что для отдельных подсистем объекта существенно проще предложить математическое описание, чем для всего объекта. В дальнейшем математическое описание объекта строится как совокупность математических описаний подсистем. Таким образом, декомпозиция – один из основных подходов к разработке математических моделей сложных систем. Однако проведение декомпозиции существенно зависит от вида объекта, для которого разрабатывается математическая модель. Рассмотрим некоторые классы таких объектов.

Объект – *техническая система*. Технической системой называется система, в которой поставленные цели могут быть полностью достигнуты в результате протекания внутренних явлений: физических, физико-химических, тепловых и т.п. Задача исследователя состоит в определении наиболее благоприятных для протекания требуемых процессов условий и обеспечении поддержания необходимых условий на заданном уровне. В технических системах роль человека минимальна, как правило, достаточно детально описывается инструкциями и другими регламентирующими документами. Декомпозицию в технических системах проводят таким образом, чтобы функционирование каждого элементарного объекта, полученного в результате декомпозиции, определялось одной физической, физико-химической или какой-либо другой закономерностью, и, следовательно, описывалось одним уравнением.

Объект – *социотехническая*, *организационная* или *человеко-машинная система*. В системах такого типа предполагается, что цели достигаются в результате совместной работы механизмов, агрегатов, станков и людей, производственного персонала, осуществляющих производственную деятельность и определяющих направления функционирования технических средств. Наличие человека – основная черта организационных систем. Ввиду этого организационные системы имеют следующие особенности.

• Целенаправленность: человек всегда стремится так определить функционирование системы, чтобы доля его участия была минимальной, т.е. развитие производственного процесса в организационных системах направлено на сокращение живого труда.

• Наличие неопределенности: разные исполнители выполняющие одни и те же виды работ будут иметь различные результаты. Здесь сказывается наличие опыта, квалификации, психологическое состояние конкретного человека и прочее. Это, в свою очередь, может существенно повлиять на общие показатели функционирования системы в целом. Такие факторы как опыт, настроение, дисциплинированность и т.п. субъективные факторы трудно предусмотреть в модели заранее и соответственно трудно формализовать.

• Активность: человек как активный элемент системы в процессе своей деятельности старается изменить условия и характер труда в сторону улучшения, повышения производительности труда, качества продукции и пр. Это осуществляется за счет рационализации, изобретательства, введения в производственный процесс новых форм и методов работы, которые до рассматриваемого момента времени не применялись, т.е. человеку присуща творческая составляющая, которую учесть при составлении моделей практически невозможно.

Таким образом, функционирование организационных систем имеет вероятностный характер, что требует применения соответствующего математического аппарата при формализации процесса функционирования такого рода систем. В качестве рекомендаций при математическом моделировании организационных систем можно предложить разделить функции технической части системы и человека как участника производственного процесса и отдельно как лица, принимающего решения относительно направления функционирования системы.

Следующий тип объектов – *социальные системы*. Такие системы представляют собой коллектив людей, участвующий в некотором едином процессе. Особенностью социальных систем является то, что отдельные личности помимо общей для всей системы цели могут иметь еще свои подцели, которые не всегда совпадают с целями системы, а зачастую могут даже входить с ней в противоречие. Декомпозиция такого рода систем представляет особые трудности.

При проведении декомпозиции требуется соблюдать правило, которое гласит, что необходимо сопоставление модели объекта с моделью цели и наоборот, т.е., например, при рассмотрении целей системного анализа проводится сопоставление объекта анализа, – цели развития системы – с соответствующей моделью системы. Тогда операция декомпозиции представляется как выделение в структуре целей элементарных функций, которые соответствуют элементам модели системы. Иными словами, строится дерево целей, в котором цель разбивается на подцели, подцели на функции, функции на операции и т.д. При этом отмечается, что цель соответствует модели системы. Например, цель состоит в проведении экономического анализа деятельности предприятия. Ей соответствует экономическая модель предприятия. Если цель заключается в определении показателей надежности или безопасности функционирования объекта, то соответствующая модель будет моделью надежности или безопасности. Далее выделяются подцели, которые способствуют достижению глобальной цели. Подцелям ставятся в соответствие подсистемы или группы подсистем, реализующих данные подцели. У каждой подцели выделяют функции, решение которых при-

водит к выполнению подцелей. Функциям ставят в соответствие блоки. Далее функции делят на операции, операциям соответствуют элементы, их реализующие. Аналогично выполняются действия по декомпозиции системы на множество подсистем, частей, элементов, комплектующих систему. В результате декомпозиции должно получиться столько частей, сколько элементов содержит модель, взятая в качестве основания. Вопрос о полноте декомпозиции – это вопрос завершенности модели.

Таким образом, объект декомпозиции должен сопоставляться с каждым элементом модели-основания. Однако и сама модель-основание может с разной степенью детализации отображать исследуемый объект. Скажем, при проведении исследований приходится использовать модель «жизненного цикла», которая позволяет проводить декомпозицию процессов на последовательные этапы от его возникновения до завершения. Разбиение процесса на этапы дает представление о последовательности действий, начиная с обнаружения проблемы и заканчивая ее ликвидацией. Степень такого разбиения может быть различной. Например, когда говорят о функционировании объекта с точки зрения надежного выполнения им своих функций, выделяют этапы приработки, нормального функционирования и старения. Ясно, что такое разбиение довольно условно. В период нормального функционирования можно выделить этапы работы под нагрузкой и простои оборудования, далее можно выделить этап исправного функционирования и восстановления работоспособности, профилактики и ремонта, причем эти этапы могут иметь разную длительность и чередоваться друг с другом. Этап старения можно разделить на начало старения, когда объект начал терять свою работоспособность, но еще удовлетворительно выполняет функции, и этап глубокого старения, когда его требуется заменить. Следовательно, при выборе вида и степени детализации модели-основания также необходимо исходить из постановки задачи системного анализа и существа решаемой проблемы.

При проведении декомпозиции имеется еще один вопрос, который требует проведения дополнительных исследований, – это взаимоотношение между полнотой и простотой модели. Иными словами вопрос состоит в следующем – до какой степени детализации следует проводить процесс декомпозиции. Ответ на этот вопрос весьма существенен. С одной стороны, чем более подробно проведена декомпозиция, чем на более мелкие объекты разбивается система, тем детальнее получается ее модель, тем более тонкие эффекты и особенности системы она может отразить и учесть. Но, с другой стороны, чем больше элементов представлено в структуре системы, тем больше взаимосвязей требуется учесть при объединении моделей объектов в модель системы, поскольку модель системы не является простой суммой моделей элементарных составляющих. Далее при слишком подробном представлении системы математическая модель системы содержит слишком большое количество математических операторов, отражающих модели элементов, а следовательно, такая модель труднореализуема. Еще один фактор, ограничивающий детализацию представления системы, – наличие информации о параметрах и коэффициентах модели для ее идентификации, т.е. при слишком детальном разбиении системы может оказаться, что для описания представленных компонентов не имеется информации о параметрах, необходимых для включения в модель. Таким образом, для каждой конкретной системы, каждой задачи и цели исследования существует некоторая разумная степень декомпозиции, переступать которую нецелесообразно.

Наконец следует отметить еще одно обстоятельство. Поскольку декомпозиция объекта проводится путем сопоставления модели объекта с моделью-основанием, а сама модель-основание, в свою очередь, тоже подвержена изменениям, следовательно, процесс декомпозиции целесообразно проводить путем постепенной детализации используемых моделей. Естественно, что такой процесс будет иметь итеративный характер.

3.5. Агрегирование – метод обобщения моделей

Операцией, противоположной декомпозиции, является *агрегирование* – объединение частей в целое. Операция декомпозиции применяется на этапе анализа системы. Цель декомпозиции – представить систему в виде иерархической структуры, т.е. разбить ее на подсистемы, их, в свою очередь, на части, далее выделить блоки, блоки представить в виде элементов и т.д. Аналогичные действия производят с целями, выделяют подцели, далее задачи, функции, операции. Затем для выделенных элементарных компонентов строят математические модели. Далее начинается операция сбора моделей компонентов системы в единую модель. Эта операция и есть агрегирование. Цель агрегирования – составление модели систем из моделей составляющих компонентов. Если декомпозиция системы осуществляется сверху вниз, то агрегирование идет снизу вверх.

Будучи объединенными, взаимодействующие элементы образуют систему, которая обладает не только внешней целостностью, обособленностью от окружающей среды, но и внутренним единством. Проявлением внутренней целостности системы является наличие у системы новых свойств, которые отсутствовали у отдельных элементов. Система не является только лишь объединением элементов, она представляет собой нечто большее. Система в результате ее создания приобретает такие свойства, которых нет ни у одного из ее элементов или частей. Естественно, что эти свойства появляются у системы ни вдруг, ниоткуда. Система обязана появлением качественно новых свойств благодаря наличию конкретных связей между конкретными элементами. Задача агрегирования заключается в том, чтобы сформировать модель системы из моделей элементов и не упустить при этом тех свойств, которые получаются при объединении элементов. Поскольку модель есть лишь слепок системы, ее отражение, то в ней должны быть реализованы хотя бы основные свойства, выражающие целевую направленность данной модели.

Приведем пример. Допустим, решается вопрос о расчете характеристик надежности некоторой системы. В результате выполнения декомпозиции построены модели надежности элементов. На этапе агрегирования последовательно от элементов к блокам, от блоков к каналам, от каналов к подсистемам и т.д. собирается модель системы. Какие новые свойства могут появиться у системы с точки зрения надежности функционирования? Одно из таких свойств – это повышение надежности за счет дублирования элементов или каналов, как это показано на рис. 2.3, в результате чего вся система более надежно выполняет свои функции. Далее может иметь место функциональное дублирование подсистем. В случае выхода из строя одной подсистемы частично может взять на себя выполнение ее функций другая подсистема. В системе управления и защиты энергоблоков атомных станций есть подсистема выработки сигнала на срабатывание аварийной защиты при превышении уровня мощности выше заданного предела и подсистема выработки сигнала при превышении скорости нарастания мощности. Наличие данных подсистем приводит к тому, что система в целом выполнит задачу остановки реактора в случае наступления аварийной ситуации даже при неисправности одной из них. Налицо функциональное дублирование. Задача агрегирования – реализовать данное свойство системы при составлении конкретной модели, в данном случае модели надежности системы.

Как и в случае декомпозиции, техника агрегирования основана на использовании определенных моделей исследуемой системы. Именно избранные модели жестко определяют, какие части должны войти в состав модели и как они должны быть связаны между собой. Разные постановки задач приводят к разным целям агрегирования и, следовательно, к необходимости использования разных моделей. Так при построении модели надежности не используется информация о стоимости того или иного блока, не принимаются во внимание стоимостные модели. Если ставится задача оптимизации структуры с использованием стоимостных критериев, то используются модели надежности и стоимости, но игнорируются, скажем, модели физических процессов, протекающих в блоках. Таким образом, тип окончательного агрегата определяется постановкой задачи и общей целью проводимого исследования. Отметим, что *агрегатом* называется результат выполнения операции агрегирования, т.е. модель, получаемая в ходе агрегирования. Точно также техника построения агрегата определяется условиями и целями агрегирования. В общем виде *агрегирование определяют как установление отношений на заданном множестве элементов*.

Объектом системных исследований являются большие или сложные системы широкой прикладной направленности. Системный анализ применяется для решения задач исследования технических, социотехнических, социальных, природных систем, т.е. объектом анализа может быть и технологический процесс, и экологическая ситуация обширной территории, и технико-экономическое развитие промышленного объекта, и социально-психологические исследования внутри коллектива. Естественно, что приходится наблюдать и описывать разнообразные процессы и структуры в ходе проведения исследований. Количество таких процессов очень многообразно и требует для своего описания применения разнообразных моделей. Здесь следует отметить одно важное обстоятельство. Конечная модель системы должна давать полное представление о системе с точки зрения поставленной цели исследования. Только совместное описание в терминах нескольких качественно различающихся языков позволяет охарактеризовать явление с достаточной полнотой. Например, при проектировании автоматизированной системы управления предприятием систему необходимо описывать в виде структурной схемы ее элементов, в виде функциональной схемы решаемых задач, в виде организационной схемы, в которой отражается связь данной системы с верхним и нижним уровнями управления, роль системы в принятии управленческих решений, далее необходимо в виде схемы отразить информационные потоки, циркулирующие в системе и прочие особенности. Если не будет представлена хотя бы одна из схем, система утратит свою целостность. Здесь приходится опять сталкиваться с проблемой полноты описания и возможной минимизацией описания явления. Причем, говоря о процессе агрегирования, необходимо заметить, что неполнота описания становится почти недопустимой. При неполноте описания речь может идти вообще не о том предмете, который имеется в виду. С другой стороны, переопределение связано с большими затратами. Таким образом, для создания агрегата необходимо привлечение качественно различных языков описания системы, причем число языков должно быть минимально, но в необходимом количестве для реализации заданной цели. Перечислив языки, на основании которых строится модель системы, тем самым определяется тип системы, фиксируется понимание природы системы.

Итак, приходится констатировать, что для разработки моделей систем используются разнообразные языки описания. Количество языков возрастает, когда приходится говорить о динамическом описании поведения систем. Однако, несмотря на многообразие описываемых процессов и структур, они могут быть классифицированы и представлены в виде ограниченного набора классов-агрегатов. Рассмотрим некоторые из наиболее употребимых видов агрегатов.

Агрегаты-структуры

Важной формой агрегирования является образование структур. Как и любой вид агрегата, структура является моделью системы и определяется совокупностью: объект, цель и средства моделирования. В результате получается многообразие типов структур: сетевые, древовидные, матричные. При синтезе создается структура будущей системы. Если это реальная система, то в ней установятся не только те связи, которые заложены в ходе проектирования, но и те, которые возникают из самой природы сводимых в систему элементов. Вспомним пример с подсистемами системы управления и защиты энергоблока АС. Функциональное дублирование возникает ввиду наличия соответствующих физических процессов, происходящих в установке, существует объективно, получается само собой.

Далее, говоря об агрегатах-структурах, следует отметить, что при проектировании системы важно задать ее структуру во всех существенных отношениях. Рассмотрим пример, иллюстрирующий данный тезис. При проектировании радиотехнических приборов требуется разработка нескольких видов структур, а именно, блок-схема, принципиальная и монтажная схемы. Блок-схема определяется выпускаемыми промышленностью радиоэлементами, и прибор делится на такие элементы. Принципиальная схема предполагает совершенно иное деление, так как она должна объяснять функционирование этого прибора. На ней выделены функциональные единицы – конденсаторы, диоды, транзисторы, которые могут не иметь пространственно локализованных аналогов, т.е. в реальности они выполнены в виде интегральных схем. Монтажная схема является результатом представления пространственной геометрии прибора, в пределах которого производится его монтаж. Таким образом, проект любой системы должен содержать разработку стольких структур, на скольких языках эта система описывается. Например, в организационных системах можно выделить иерархическую структуру подчиненности, структуру циркуляции информации, структуру производственного процесса и т.д. Эти структуры могут существенно отличаться топологически, но все они описывают с разных сторон одну и ту же систему и поэтому не могут быть не связаны между собой.

Агрегаты-операторы

Тип агрегата-оператора имеет место тогда, когда агрегируемые признаки фиксируются в числовых шкалах. В этом случае задается отношение на множестве признаков в виде числовой функции многих переменных, которая и является агрегатом. Основное применение агрегаты-операторы находят при описании динамических свойств системы. Представление зависимости выходных показателей системы в виде функционала от входных переменных есть пример агрегата-оператора.

Рассмотрим формализованное определение агрегата-оператора. Пусть T – множество моментов времени; X – множество входных сигналов; U – множество сигналов управления; Y – множество выходных сигналов; Z – множество состояний системы. Элементы указанных множеств назовем $t \subset T$ – моментом времени; $x \subset X$ – входным сигналом; $u \subset U$ – управляющим сигналом; $y \subset Y$ – выходным сигналом; $z \subset Z$ – состоянием системы. Все перечисленные сигналы будем рассматривать как функции времени $x(t)$, $u(t)$, $y(t)$, $z(t)$. Под агрегатом-оператором будем понимать объект, определяемый множествами T, X, U, Y, Z и операторами H и G, которые являются оператором переходов H и оператором выходов G. Данные операторы реализуют соответственно функции $z(t)$ и $y(t)$.

Рассмотрим оператор переходов H. Пусть даны состояния системы в моменты времени t и $t + \Delta t$, т.е. предполагается, что система за время Δt переходит из состояния $z(t)$ в состояние $z(t + \Delta t)$. Если известно, что в момент времени t в систему поступают входные сигналы $x(t)$ и управление $u(t)$, то оператор переходов однозначно определяет состояние системы в следующий момент времени $z(t + \Delta t)$:

$$z(t+\Delta t) = H\{t,\ x(t),\ u(t),\ z(t)\}.$$

Аналогично оператор выходов однозначно определяет значения выходных характеристик системы и выражается следующим образом:

$$y(t) = G\{t,\ x(t),\ u(t),\ z(t)\}.$$

Если для системы удается представить зависимость ее выходных и входных параметров, управляющие воздействия и состояния в виде агрегата-оператора, то получается довольно хорошо формализованная математическая модель. Ограничивающим фактором для решения такого рода моделей, как правило, является только лишь большая размерность входящих в нее параметров.

Агрегаты-статистики

Процессы функционирования реальных сложных систем во многих случаях носят случайный характер. Выходные характеристики таких систем принимают случайные значения из множества величин, описываемых некоторой функцией распределения $F(\theta, t)$, где θ – вектор параметров закона распределения; t – некоторый момент времени. Если элементы вектора параметров функции распределения выражаются через достаточные статистики, тогда нет необходимости хранить всю информацию о реализованных характеристиках системы. Эту информацию можно заменить оценками параметров, полученными по реализовавшимся результатам наблюдений. Достаточные статистики – это агрегаты, которые извлекают всю полезную информацию об интересующем параметре из совокупности наблюдений. Примерами достаточных статистик являются параметры нормального закона распределения – математическое ожидание и дисперсия, параметр экспоненциального закона распределения – λ-характеристика. Использовать достаточные статистики необходимо с большой осторожностью. Их применение оправдано только в том случае, когда обоснован вид закона распределения, описывающий совокупность выходных величин. Дело в том, что агрегирование в данном случае является необратимым преобразованием, которое может привести к потере информации. Например, по сумме нельзя восстановить совокупность случайных величин слагаемых суммы.

Стохастические модели, в основе которых лежат предположения о законе распределения исследуемой случайной величины, так же, как и агрегаты-операторы хорошо изучены. Имеется соответствующий математический аппарат, в современных операционных системах представлено обширное прикладное программное обеспечение, позволяющее успешно работать с подобного рода моделями.

Агрегат как случайный процесс

Если процесс функционирования реальной сложной системы по своему существу носит характер случайного процесса, для агрегата как математической модели системы используются основные понятия теории случайных процессов. Случайный процесс, протекающий в любой физической системе, представляет собой случайные переходы системы из состояния в состояние. Состояние системы может быть охарактеризовано с помощью численных переменных: в простейшем случае – одной, в более сложных – несколькими. Понятие случайного процесса представляет собой обобщение понятия случайной величины. *Случайным процессом* $X(t)$ называется процесс, значение которого при любом фиксированном $t = t_0$ является случайной величиной $X(t_0)$. Случайная величина $X(t_0)$, в которую обращается случайный процесс при $t = t_0$, называется сечением случайного процесса, соответствующим данному значению аргумента t. Теория случайных процессов бурно развивается в настоящее время. Ее аппарат изложен в обширной литературе, например, [27]. Имеются частные случаи случайных процессов: марковские, полумарковские, винеровские, кусочно-непрерывные и т.п.

Таким образом, можно подвести итог. Существует большое количество форм агрегирования, т.е. объединения частей в целое. Их общность состоит в том, что агрегирование диктуется выбранной моделью описываемой системы. Агрегирование есть установление отношений между агрегируемыми элементами. Наиболее важными видами агрегатов являются агрегаты-структуры, агрегаты-операторы, агрегаты-статистики и случайные процессы.

Глава 4

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ – МЕТОД ПРОВЕДЕНИЯ СИСТЕМНЫХ ИССЛЕДОВАНИЙ

4.1. Сущность имитационного моделирования

Особым видом моделей являются имитационные модели. Имитационное моделирование проводится в тех случаях, когда исследователь имеет дело с такими математическими моделями, которые не позволяют заранее вычислить или предсказать результат. В этом случае для предсказания поведения реальной сложной системы необходим эксперимент, имитация на модели при заданных исходных параметрах. Имитация представляет собой численный метод проведения на ЭВМ экспериментов с математическими моделями, описывающими поведение сложной системы в течение заданного или формируемого периода времени. Поведение компонентов сложной системы и их взаимодействие в имитационной модели чаще всего описывается набором алгоритмов, реализуемых на некотором языке моделирования. Термин «имитационная модель» используют в том случае, когда речь идет о проведении численных расчетов и в частности о получении статистической выборки на математической модели, например, для оценки вероятностных характеристик некоторых выходных параметров. Моделирование на системном уровне применяется в системном анализе для проведения расчетов характеристик будущей системы. При построении имитационной модели исследователя, прежде всего, интересует возможность вычисления некоторого функционала, заданного на множестве реализаций процесса функционирования изучаемой системы. Наиболее важным для исследователя функционалом является показатель эффективности системы. Имитируя различные реальные ситуации на модели, исследователь получает возможность решения таких задач как оценка эффективности тех или иных принципов управления системой, сравнение вариантов структурных схем, определение степени влияния изменений параметров системы и начальных условий на показатель эффективности системы. Примерами расчетов на имитационных моделях также могут служить вычисления характеристик производительности, надежности, качества функционирования и т.п., которые необходимо определить как функции внутренних и внешних параметров системы.

Ответственный этап создания имитационной модели представляет собой этап составления формального описания объекта моделирования сложной системы. Цель этапа – получение исследователем формального представления алгоритмов поведения компонентов сложной системы и отражение вопросов взаимодействия между собой этих компонентов. При составлении формального описания модели исследователь использует тот или иной язык формализации. В зависимости от сложности объекта моделирования и внешней среды могут использоваться три вида формализации: аппроксимация явлений функциональными зависимостями, алгоритмическое описание происходящих в системе процессов, комбинированное представление в виде формул и алгоритмических записей.

Сложность системы и вероятностный характер процессов, происходящих в объекте исследования, свидетельствуют о том, что для определения выходных характеристик системы необходимо использовать стохастические модели. Вероятностный характер процессов, происходящих в сложных системах, приводит к невозможности аппроксимации явлений функциональными зависимостями. Доминирующим методом при моделировании сложных систем является способ алгоритмического описания происходящих в системе процессов.

Отметим еще одну особенность, которую необходимо учитывать при моделировании процесса функционирования сложной системы. В социотехнических системах люди решают часть задач из общей последовательности задач, решаемых системой, например, задачи управления, принятия решения и т.п. Следовательно, они принципиально не устранимы из системы и должны быть представлены в модели системы как ее элементы. Однако учет так называемого «человеческого фактора» имеет принципиальные сложности. При выполнении человеком производственных операций требуется учитывать квалификацию конкретного исполнителя, его опыт и стаж работы. Необходимо также иметь в виду, что на качество выполняемых процедур могут оказывать влияние состояние его здоровья, эмоционально-психологический настрой и прочие факторы, которые практически не удается формализовать при составлении модели. Поэтому в моделях принимают определенного рода допущения, приводящие к упрощению модели, к решению задачи «в среднем», т.е. задают некоторые средние характеристики выполнения человеком своих функций и при данных значениях проводят расчеты

модели. Для того, чтобы учесть возможные отклонения в процессе выполнения операций различными исполнителями, необходимо проводить анализ чувствительности модели.

4.2. Композиция дискретных систем

При моделировании системы в большинстве задач системного анализа интересуются не статической структурой, а ее динамическим поведением, т.е. тем, как система выполняет свои функции. Совокупность выполняемых сложной системой функций может быть представлена в виде последовательности задач: производственных, управления, принятия решений и т.д. Часть из этих задач в процессе функционирования системы может быть формализована и решена с помощью ЭВМ или других технических средств, часть – неформализуема и может быть решена только человеком. В процессе составления имитационной модели любую задачу можно разбить на операции в соответствии с логикой ее решения. В таком случае процесс функционирования сложной системы представляется в виде последовательности операций как технологическая схема производственного процесса, процесса управления, обработки информации и пр. Причем этот процесс можно представить с разной степенью детализации.

При имитационном моделировании реализующий модель алгоритм воспроизводит процесс функционирования системы во времени. При этом имитируются элементарные явления, составляющие процесс, с сохранением их логической структуры и последовательности протекания во времени. В результате по исходным данным получают сведения о состояниях процесса в определенные моменты времени, дающие возможность оценить характеристики системы.

При моделировании работы системы важно знать, какими устройствами или людьми выполняются отдельные операции. Поэтому разбиение процесса на операции следует производить, руководствуясь и логикой выполнения производственного процесса, процесса управления, преобразования информации, т.е. логикой организации и проведения технологического процесса в самом общем смысле, а также логикой работы устройств и людей. Разумно выделять в качестве операций такие части процесса, которые реализуются без прерывания, без обращения к другим элементам одним устройством или одним человеком. В то же время при диалоговом решении какой-либо задачи не всегда необходимо отделять операции, выполняемые человеком и ЭВМ. При моделировании процесса в целом за операцию может быть принят весь цикл взаимодействия человека и ЭВМ, который осуществляется в ходе решения задачи. Расчленение процесса на операции всегда неформально. В зависимости от цели исследования, наличия исходных данных, требуемой разрешающей способности модели могут быть построены разные структурные схемы модели для отображения одного и того же процесса. Однако, несмотря на различия итогового представления общей схемы имитационной модели, у любого представления есть нечто общее, – это отображение динамического процесса функционирования системы в виде последовательности выполнения элементарных операций. Как правило, элементарные операции представляются в модели в виде единичных дискретных актов. Рассмотрим, каким образом осуществляется представление функциональных модулей при представлении их в имитационной модели.

Определим динамическую компоновку системы в терминах понятий, называемых активностями, процессами и событиями.

Динамическая компоновка системы

Активности. Характеристики производительности прямо или косвенно связаны со скоростью, с которой система выполняет свою работу, поэтому они содержат время в качестве независимой переменной. Работа совершается путем выполнения активностей. Активность является наименьшей единицей работы. Активность рассматривается как единый дискретный шаг. С каждой активностью связано время выполнения. Активность может соответствовать определенному этапу выполнения команды в моделировании на уровне регистровых передач или выполнения целого задания в макроскопическом моделировании вычислительной системы. Независимо от содержания представляемой деятельности активность является единым динамическим объектом, указывающим на совершение некоторой единицы работы.

Процессы. Логически связанный набор активностей образует процесс, который можно рассматривать как объект, вмещающий или инициирующий эти активности. Некоторый процесс может выступать в роли активности или субпроцесса в процессе более высокого уровня. Подобно активностям, процессы представляют собой единые динамические объекты. Выполнение в вычислительной системе определенной операции с дисками можно рассматривать как процесс, включающий в себя активности установки головки записи-чтения, задержки на вращение носителя и передачи данных.

Различие между активностями и процессами условно. Операция, определенная на одном уровне как активность, на другом уровне может

рассматриваться как процесс. Каждый процесс инициируется другим процессом, называемым инициатором. Инициатор может находиться как вне системы, так и внутри нее.

События. Активности инициируются в результате совершения событий. Событие представляет собой мгновенное изменение состояний некоторого объекта системы, который может быть как пассивным, так и активным. Окончание активности является событием, совершение которого может возбудить последующие активности. Такие события управляют следованием активностей внутри процесса и являются локальными или внутренними событиями данного процесса. Инициализация активности не означает немедленного ее выполнения. Начало выполнения находится в зависимости от наличия определенных условий, в число которых входит и инициализация данной активности. Инициализация активности является результатом окончания заранее указанных одной или нескольких активностей. Можно считать, что активность, выполнение которой не может быть начато из-за отсутствия необходимых условий, находится в ожидании совершения событий, приводящих к появлению этих условий. События можно разделить на две категории: события следования, которые управляют инициализацией активностей внутри данного процесса, и события изменения состояний, которые управляют выполнением активностей, относящихся в общем случае к независимым процессам. С точки зрения динамики система рассматривается как совокупность связанных друг с другом процессов, причем взаимодействие между ними управляется и координируется совершающимися событиями.

Описание процессов

В то время как динамическое поведение системы формируется в результате выполнения большого числа взаимодействующих процессов, сами эти процессы образуют относительно небольшое число классов. Чтобы описать поведение системы, достаточно указать некоторые классы процессов и задать значения атрибутов для конкретных процессов, являющихся элементами этих классов. Правила, описывающие поведение класса процессов, состоят из указаний активностей, входящих в процессы в определенных соотношениях следования, условий, управляющих их выполнением, и воздействий, оказываемых процессами на атрибуты и состояния активных и пассивных объектов системы. Эти правила в совокупности называются описанием процесса; конкретный процесс является случаем выполнения этого описания при заданных значениях атрибутов. Построение модели состоит из решения двух основных задач. Первая задача сводится к тому, чтобы описать правила, задающие различные виды процессов, происходящих в системе. Вторая, наиболее трудная задача, заключается в том, чтобы указать значения атрибутов процессов или задать правила генерации этих значений. Рассмотрим примеры классов и атрибутов на данных классах. Самым простым способом моделирования процесса является определение длительности его операций в виде детерминированных величин, т.е. класс, которым будет описываться порядок следования активностей, – это детерминированные величины. Задание атрибута на данном классе – это задание длительности выполнения конкретной операции. Следующим примером может служить моделирование длительности операции с помощью случайных величин. Тогда необходимо задать вид закона распределения, т.е. сказать, что случайные величины будут моделироваться, скажем, как нормально распределенные величины, экспоненциально распределенные величины, гамма-распределенные величины и т.п. Задание вида закона распределения означает задание класса (класс экспоненциально распределенных величин). Атрибутами на данном классе будут параметры закона распределения. Для нормального закона – это математическое ожидание и дисперсия, для экспоненциального – λ-характеристика, для Г-распределения – параметры масштаба и формы и т.д. Далее организовать выполнение процесса моделирования можно с помощью реализации случайных процессов. Тогда задание класса будет заданием вида случайного процесса, например, класс диффузионных процессов. Атрибуты класса – параметры процесса. Для диффузионных процессов – это параметры сноса и диффузии.

Рассмотренные объекты – активности, процессы и события – являются конструктивными элементами, с помощью которых описывается динамическое поведение дискретных систем, и на основе которых строятся языки моделирования этих систем. Система описывается на определенном уровне в терминах множества описаний процессов, каждое из которых включает в себя множество правил и условий возбуждения активностей. Такое описание системы может быть детализировано на более подробном уровне представления с помощью декомпозиции активностей в процессы; эти описания вместе с описаниями процессов предыдущего уровня образуют расширенное описание системы. Последовательное применение этой операции порождает множество описаний системы на различных уровнях детализации. Это обеспечивает многоуровневое исследование системы. Такое исследование играет важную роль в проектировании, описании и моделировании сложных систем.

Организация процесса моделирования

Система выполнения имитационного процесса во времени включает в себя механизм динамического управления активными объектами модели. Ее сложность варьируется в значительных пределах в зависимости от класса реализуемого алгоритма.

Алгоритмы моделирования дискретных систем в зависимости от подхода к описанию рассматриваемых объектов делят на классы алгоритмов, ориентированных на активности, события или процессы. Независимо от класса, к которому относятся все эти алгоритмы, они направлены, прежде всего, на представление активностей. Активность является объектом, потребляющим время. Для моделирования системы необходимо, чтобы время выполнения каждой активности в системе было известно или могло быть вычислено.

В реальной системе совместно выполняются несколько активностей, принадлежащих как связанным, так и не связанным между собой процессам. После выполнения этих активностей, вызывающих изменение состояний других системных объектов, совокупность происшедших событий следования и изменения состояний инициирует начало выполнения других активностей. Поведение системы во времени отображается порядком следования событий, установленным на уровне глобальных переменных. Моделирование системы по существу является задачей генерации этой временной последовательности событий. Все изменения состояний в моделируемой системе происходят в моменты времени, соответствующие окончанию активностей. Инициирование активности совершается как результат окончания ранее выполнившейся активности или совокупности активностей, которые могут существовать вне рассматриваемой системы. Будучи инициированным, выполнение активности может начаться немедленно или задержаться до появления определенных условий или состояний системных объектов.

Состояния изменяются в результате действий, совершаемых другими активностями как в начале, так и в конце их выполнения; в первом случае может начаться немедленное выполнение новых активностей, вызванное изменением состояний системных объектов. Инициирование первой активности системы происходит в результате окончания некоторой предшествующей активности, находящейся вне системы. Это рассуждение показывает, что наиболее важную роль в моделировании играют моменты, когда заканчиваются активности. Все события в системе происходят именно в эти моменты времени.

В алгоритмах, ориентированных на события, моделирующая программа организована в виде совокупности секций событий или процедур событий. Процедура события состоит из набора операций, которые в общем случае выполняются после завершения какой-либо активности. Выполнение процедуры синхронизируется во времени списковым механизмом планирования. Каждый элемент списка определяет время события, представляющего завершение выполняемой активности, вместе с именем или номером процедуры события, которая должна выполняться после совершения этого события. Подобный список часто называют списком следующих событий или просто списком событий. После ввода в процедуру события выполняются требуемые действия, относящиеся к завершению активности. В результате этих действий могут измениться состояния различных объектов системы, что позволяет возобновить выполнение ранее инициированных активностей, которые были задержаны до появления определенных условий, возможно оказавшихся теперь выполненными из-за происшедших изменений состояний. Процедура события проверяет, возникла ли такая ситуация. Если это так, она осуществляет действия, переводящие задержанные активности в состояние выполнения, и планирует времена их завершения. В общем случае завершение одной активности инициирует выполнение другой активности, связанной с первой порядком выполнения. В этой ситуации процедура события определяет, существуют ли условия запуска данной активности. В случае их наличия процедура события выполняет необходимые для запуска действия и планирует время ее завершения. После выполнения всех действий, соответствующих текущему моменту времени моделирования, активизируется механизм планирования, который выбирает новую процедуру события.

4.3. Содержательное описание сложной системы

Наиболее важным этапом построения модели сложной системы является этап содержательного описания объекта моделирования. Работа на данном этапе построения модели сложной системы начинается с анализа постановки задачи. В качестве исходной информации при построении модели сложной системы используется сформулированная цель системного исследования. Совокупность сведений об объекте моделирования представляется в виде схем, текстов, таблиц экспериментальных данных, характеризующих анализируемую структуру и характер функционирования системы. Кроме того, при составлении модели должна учитываться информация о внешних воздействиях и параметрах окружающей среды.

На начальном этапе построения модели системы необходимо четко определить цель будущего исследования на модели, а затем в соответствии с этой целью переработать весь объем исходной информации и постараться восполнить недостающую информацию. Этот процесс называется составлением содержательного описания сложной системы. Рекомендуется следующая последовательность действий при составлении содержательного описания сложной системы: выбор показателей качества, отражающих цели моделирования; определение управляющих переменных, выбор состава контролируемых характеристик объекта моделирования; детализация описания режимов функционирования системы; представление информации о воздействии внешней среды.

Выбор показателей качества моделируемой системы

Выбор показателей качества определяется теми задачами, для решения которых строится модель. Часто наблюдается тенденция имитировать все, что касается поведения объекта исследования. Однако такой подход неверен. При построении модели следует ориентироваться на решение лишь тех вопросов, которые сформулированы в постановке задачи, а не имитировать реальную систему во всех подробностях. На первом этапе решения задачи важно отделить главное, то, что действительно ведет к достижению сформулированной цели, от второстепенного. Выбор цели моделирования определяет характеристики, которые отражают поведение сложной системы. В дальнейшем вся работа сводится к выявлению и детализации тех аспектов функционирования системы, которые имеют отношение к выбранным показателям. Приведем конкретный пример. Пусть перед исследователем стоит задача анализа характеристик надежности системы. В этом случае в качестве показателей необходимо выбирать один или несколько из числа следующих: коэффициент готовности, вероятность выполнеиия задачи, возложенной на систему, вероятность безотказной работы, наработка на отказ и т.п. При этом не анализируется время выполнения задачи, средства, необходимые для ее решения, выделяемые ресурсы и пр.

Определение управляющих переменных системы

На этом этапе изучается техническая документация, по которой прослеживается информация, относящаяся к управлению системой. Согласно цели проведения системного анализа устанавливается состав управляемых и контролируемых характеристик объекта моделирования. Прежде всего выделяются те характеристики управления системой и контроля за ее работой, которые имеют отношение к цели моделирования. Все составляющие функциональной зависимости, определяющие значение показателя качества системы, включаются в состав управляющих переменных и контролируемых характеристик объекта моделирования. Продолжая рассматривать предыдущий пример, можно отметить, что с точки зрения анализа надежности системы на данном этапе важно заложить в модель следующую информацию: отметить наличие или отсутствие контроля за исправностью функционирования каждого из элементов, комплектующих систему; характер контроля (встроенный, периодический), если контроль периодический, то необходимо отметить время, через которое его проводят; полноту контроля; наличие профилактических мероприятий (плановых и аварийных); частоту проведения плановых профилактик и т.д.

Выбор состава контролируемых характеристик объекта моделирования

Выбрать состав контролируемых характеристик объекта моделирования, значит, указать те выходные параметры системы, которые имеют отношение к показателям качества, сформированным на первом этапе содержательного описания системы. Иными словами, необходимо указать те характеристики, через которые реализуются показатели качества. Поясним данную мысль примером. Пусть необходимо произвести расчет коэффициента готовности системы. Для определения этого показателя необходимо знать сколько времени в каждом модельном эксперименте система находилась в исправном состоянии и сколько в состоянии отказа, простоя и восстановления. Таким образом, в данном конкретном примере в качестве контролируемых характеристик объекта моделирования будет выступать время нахождения системы в каждом из перечисленных состояний. Следующий пример. Если требуется определить вероятность безотказной работы системы, то необходимо фиксировать состояние, в котором система находилась к концу каждого модельного эксперимента. Исправному состоянию приписывается значение 1, неисправному 0. Вероятность безотказной работы определяется как отношение количества успешных модельных реализаций к общему количеству испытаний. Следовательно, в данном примере контролируемой характеристикой модели системы будет ее состояние в каждом модельном эксперименте.

Детализация описания режимов функционирования системы

На данном этапе перерабатывается и дополняется имеющаяся информация для возможного выделения алгоритмов функционирования в каждом из режимов работы системы. Составляются временные диаграммы функционирования системы. Определяются наиболее неясные или сложные моменты функционирования компонентов системы, устанавливается последовательность их действий, выделяются вероятные места возникновения конфликтных ситуаций и описывается принятый порядок их разрешения в системе. Продолжим рассмотрение примера по анализу надежности системы. Для задач анализа надежности важно указать, в какие периоды система работает под нагрузкой, когда она находится в нерабочем состоянии. Важно также знать характер и величину нагрузки, потому что от этого зависят процессы старения, протекающие в элементах. Могут быть ситуации, когда разные элементы функционируют по своей собственной программе. Например, в системах управления и защиты энергоблоков атомных станций ряд элементов выполняют функции слежения за параметрами объекта управления. В случае, когда наблюдаемые параметры превышают допустимые уставки, элементы дают команду на срабатывание органов управления. Органы управления переводят установку в неработоспособное состояние. При этом они испытывают ударные нагрузки, так как стержни управления входят в активную зону под действием силы тяжести, кроме того на них может оказываться принудительное воздействие. Естественно, что для анализа надежности моменты непосредственного выполнения функций объектами куда важнее, чем моменты нахождения в состоянии простоя или ожидания.

Составление описания внешней среды

На этом этапе необходимо провести исследование факторов, оказывающих воздействие на моделируемую систему. В состав модели включаются только значимые факторы, влияние которых необходимо учитывать опять же с точки зрения сформулированной постановки задачи. В случае моделирования отдельных аспектов функционирования системы проводится исследовательская работа, цель которой состоит в определении алгоритмов взаимодействия системы с внешней средой. Иногда возможны модификация или пополнение состава управляющих переменных системы из-за детализации алгоритмов взаимодействия между системой и внешней средой.

С точки зрения анализа надежности важно указать факторы, которые способствуют деградации материалов. Например, повышенная влажность, высокие температуры, наличие радиационной активности – факторы, снижающие надежность элементов и системы в целом.

Таким образом, на каждом шаге данного этапа перерабатывается и дополняется имеющаяся информация о поведении системы в соответствии с поставленными целями моделирования. Результатом является содержательное описание сложной системы, выполненное в терминах соответствующего языка. Информация, не относящаяся к задаче моделирования, отбрасывается. После выполнения описанных этапов поставлена лишь одна цель моделирования. Далее необходимо перейти к математической постановке задачи моделирования и собственно построению модели. Отметим, что общего рецепта построения содержательной модели не существует. Однако можно утверждать, что при решении достаточно широкого круга задач модель системы представляется в виде сетевой структуры. Узлы сети являются моделями элементов системы. Дуги выражают связи между элементами. Сеть изображается в виде графа передачи, который строится на основе матрицы, отражающей действительные связи между элементами. Для построения имитационной модели необходимо задать поведение динамических элементов. Как было отмечено в предыдущем параграфе, для этого выделяются активности, процессы и события, имеющие место при функционировании системы. Эти динамические объекты описываются с помощью соответствующих классов, отражающих их поведение, и заданием на этих классах атрибутов.

4.4. Пример построения имитационной модели анализа надежности сложной системы

Построение имитационной модели системы с целью проведения расчетов характеристик надежности начинается с изучения структурной схемы системы и стратегии ее функционирования. На основании структурной схемы строится надежностная схема системы, которая характеризует статическую составляющую системы. В качестве аппарата для представления схем системы используется аппарат теории графов. Элементы системы изображаются в виде вершин графа, связи между элементами – в виде дуг. После построения надежностной схемы системы в виде графовой модели ее необходимо представить в виде функциональной зависимости (формализованное представление структуры системы). При построении имитационных моделей для формализованного представления надежностной схемы системы рекомендуют использовать аппарат алгебры логики. Используя этот аппарат, вероятностные характеристики надежности системы, такие как вероятность отказа или вероятность безотказной работы, вычисляют через логические функции работоспособности. Правила построения логических функций работоспособности описаны в [28].

Следующим этапом построения имитационной модели является отображение стратегии ее функционирования. На этом этапе осуществляется построение динамической составляющей модели системы. В качестве примера рассмотрим достаточно общую стратегию функционирования системы. Пусть в моменты времени T_k, $2T_k$, ..., nT_k производятся контрольные мероприятия по проверке неисправности элементов системы. Если в момент проведения контроля исправности элементов обнаруживается отказ, то начинаются восстановительные мероприятия. Могут быть ситуации, когда при проведении контрольных проверок отказ не обнаруживается, и элемент простаивает в состоянии отказа до следующего момента контроля. Функционирование системы продолжается до момента времени T_p, если система не отказала, или до момента отказа. В момент времени T_p начинается плановая профилактика, в момент отказа системы начинается аварийная профилактика. После проведения профилактического обслуживания система полностью обновляется, и процесс функционирования начинается заново.

Будем считать заданными периоды между проведением контрольных проверок T_k и период времени T_p, при достижении которого система подвергается восстановлению. Для организации процесса моделирования необходимо также задать вероятность обнаружения отказа P_o и исходные данные для моделирования отказов и восстановлений элементов, а именно, плотность распределения наработки до отказа для каждого элемента, входящего в состав системы, – $f_{oi}(\theta_i, t)$, где i – порядковый номер элемента; θ_i – вектор параметров закона распределения, плотность распределения времени восстановления для каждого элемента – $f_{вi}(w_i, t)$, w_i – вектор параметров закона распределения времени восстановления.

После задания всех исходных параметров переходим к организации процесса моделирования. Процесс функционирования элементов системы приведен на рис.4.1. На рисунке ступеньками обозначены периоды исправного функционирования элементов системы, линиями – периоды простоя элементов в неисправном состоянии до момента начала контроля и обнаружения неисправности, заштрихованной ступенькой обозначено время восстановления элемента после обнаружения отказа.

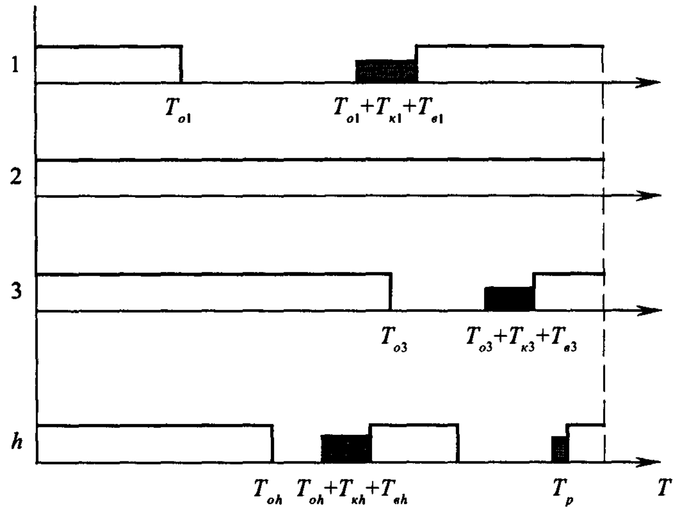

Рис. 4.1. Иллюстрация процесса функционирования элементов системы

Статистическое оценивание вероятности безотказной работы системы производится по следующей схеме. Для каждого элемента системы моделируется случайное время наработки до отказа T_{oi}. Моделирование осуществляется на основании заданной плотности распределения наработки до отказа $f_{oi}(\theta, t)$. Далее, на основании заданной вероятности обнаружения отказа моделируется событие, состоящее в обнаружении или необнаружении отказа. Если отказ обнаружен, то после ближайшего к наработке до отказа данного элемента момента контроля начинается восстановление элемента. Если выпало событие, состоящее в том, что в ближайший момент контроля отказ не обнаружен, то элемент находится в состоянии отказа до следующего момента контроля. В следующий момент контроля заново моделируется событие, состоящее в обнаружении или необнаружении отказа. Если отказ обнаружен, начинается восстановление элемента. Случайное время восстановления элемента моделируется на основании заданной плотности распределения времени восстановления.

После того как смоделированы наработки до отказа и времена восстановления каждого элемента из всего набора наработок T_{oi}, выбирают такие, для которых выполняется соотношение $T_{oi} < T_p$. Здесь необходимо отметить, что изменение состояния системы может произойти только в моменты изменения состояния элементов. Следовательно, для обнаружения отказа системы необходимо просматривать только изменения состояний элементов. Поэтому для каждого T_{oi}, для которого выполняется соотношение $T_{oi} < T_p$, проверяем условие

$$\left.\begin{aligned} T_{oi} < T_{ol} < T_{oi} + T_{\kappa i} + T_{вi} \\ T_{ol} < T_{oi} < T_{ol} + T_{\kappa l} + T_{вl} \end{aligned}\right\}$$

по всем $l = \overline{1, h}$ где h – количество элементов в системе. Проверка этого условия состоит в обнаружении элементов, находящихся в состоянии отказа в тот период, когда в состоянии отказа был i-й элемент. Введем идентификатор состояния элемента p_l. Определим его следующим образом: $p_l = 0$, если в данный период $[T_{oi}, T_{ol} + T_{\kappa l} + T_{вl}]$ элемент находился в состоянии отказа и $p_l = 1$, если элемент был работоспособен. Естественно, что в проверяемый момент $[T_{ol}, T_{oi} + T_{\kappa l} + T_{вl}]$ i-й элемент находится в состоянии отказа и для него $p_l = 0$. Сформировав массив $\{p_l\}$, на основании логической функции работоспособности определяем, был ли в данном интервале времени отказ системы. Если был, то $p_c = 0$, если отказа не было, $p_c = 1$. Если в рассматриваемый промежуток времени отказа системы не было, переходим к следующему интервалу времени. Процесс продолжается до тех пор, пока не будет равна нулю величина p_c. Если на одном из проверяемых периодов величина p_c приняла значение 0, это значение запоминается и начинается следующая итерация моделирования. Если ни на одном из рассматриваемых интервалов до момента T_p величина p_c не приняла значение 0, то отказа системы не было, и значение p_c в данном испытании равно 1. Проводя данную процедуру N раз, получаем N значений величины p_c. Статистическую оценку вероятности безотказной работы системы находим по формуле

$$P_c(T) = 1/N \sum_{j=1}^{N} p_{cj},$$

где p_{cj} – значение величины p_c в j-м испытании.

Описанная модель является концептуальной. После ее составления переходят к программной реализации и исследованию модели на ЭВМ.

Глава 5

ТЕОРИЯ ПОДОБИЯ – МЕТОДОЛОГИЯ ОБОСНОВАНИЯ ПРИМЕНЕНИЯ МОДЕЛЕЙ

5.1. Модели и виды подобия

Исследование систем с помощью моделей может быть обосновано только в том случае, когда модель адекватно описывает процессы и явления, происходящие в системах. При использовании моделей необходимо теоретически обосновать аналогию между моделью и реальным физическим явлением. Только в этом случае результаты, полученные на модели, могут быть перенесены на исследуемый объект. Без этого обоснования моделирование теряет познавательное значение, так как перестает быть достоверным источником информации о реальных явлениях и процессах. В основе доказательства степени соответствия модели и объекта лежит теория подобия. Необходимо отдавать отчет в том, что абсолютное подобие может иметь место только лишь при замене одного объекта другим, точно таким же. При моделировании абсолютного подобия добиться невозможно. Системные аналитики стремятся к тому, чтобы модель достаточно хорошо отображала исследуемую сторону функционирования объекта.

В предыдущих главах были рассмотрены основные понятия моделей, методы построения моделей систем, организация имитационного моделирования как метода проведения системных исследований. Остановимся еще раз на понятии моделирования с точки зрения теории подобия. Моделирование представляет собой процесс проведения исследований объекта, базирующийся на подобии модели и объекта, и включает в себя построение модели, ее изучение и, наконец, перенос полученных результатов на объект исследования. Под моделью понимают объект, например, явление, процесс, систему, экспериментальную установку, знаковое образование, математические выражения, находящиеся в отношении подобия к исследуемому объекту. Математическое моделирование использует подобие между величинами, входящими в математические выражения, описывающие поведение изучаемого объекта. Иначе говоря, в математическом моделировании в отличие от других форм моделирования предполагается замена явления его математическим описанием, воспроизводимым вычислительными средствами.

Физическое моделирование использует подобие между объектом и моделью, имеющей физическую природу. Основой физического или математического экспериментального исследования являются методы теории подобия, которые применяются при постановке эксперимента, обработке данных о результатах экспериментальных исследований и испытаний.

В процессе проведения системных исследований необходимо решать вопросы измерения характеристик и обработки статистических данных, распространения полученных данных на другие явления, выбора аналогов при проектировании систем, повышении точности и достоверности оценок рассчитываемых характеристик исследуемых объектов. Перечисленные вопросы решаются с помощью методов теории подобия, которая дает ответы на ряд вопросов о том, как организовать исследования и испытания объекта анализа, как обрабатывать опытные данные, обобщать и распространять полученные результаты на другие объекты.

Модель и отображаемый ею объект находятся в отношении сходства, а не тождества. Это означает, что модель по определенным признакам подобна изучаемой системе, а по каким-то может быть от нее отлична. Важное условие при проведении исследований – реализовать подобие по наиболее важным признакам с точки зрения проведения конкретного, данного исследования. Понятие модели взаимно связано с понятием подобия. Модель обеспечивает подобие тех процессов, которые удовлетворяют критериям, полученным с помощью теории подобия. Характеристики любого явления в группе подобных явлений могут быть реализованы с помощью критериев подобия путем некоторого преобразования характеристик другого подобного явления. Примером такого преобразования может служить масштабирование.

Подобие определяют как взаимно однозначное соответствие между двумя объектами, при котором функции перехода параметров, характеризующих один из объектов, к другим параметрам известны, а математические описания этих объектов могут быть преобразованы в тождественные.

Итак, сформулируем очевидный тезис: чтобы некоторая конструкция могла быть отражением, т.е. замещала в некотором смысле оригинал, между оригиналом и моделью должно быть установлено отношение подобия. Существуют разные виды подобия. Первый тип подобия –

это подобие, устанавливаемое в результате физического взаимодействия в процессе создания моделей. Приведем примеры такого подобия. Прежде всего, это масштабированные модели гидротехнических сооружений, самолетов, кораблей, автомобилей, макеты зданий и т.п. Такое подобие называется прямым. Только при прямом подобии возможна взаимозаменяемость модели и оригинала. Но даже в случае проведения исследований на модели, которая является макетом, созданным путем реализации отношения прямого подобия, возможны сложности с переносом результатов моделирования на оригинал. Например, при исследовании влияния водной среды на гидротехнические сооружения можно промасштабировать не только само сооружение, но и часть условий, в которых проводятся исследования (скорость течения воды, высоту волн), однако часть факторов масштабированию не поддается, например, вязкость воды, сила тяготения. В результате задача пересчета данных, полученных при проведении модельного эксперимента, на реальные условия становится нетривиальной.

Второй тип подобия – косвенное подобие. Косвенное подобие объективно существует в природе, обнаруживается в виде совпадений или достаточной близости оригинала и модели. Если установлена близость абстрактных моделей рассматриваемых объектов (оригинала и модели), то можно переходить к использованию их в практике реального моделирования. Наиболее известным примером косвенного подобия является подобие некоторых электрических и механических процессов, описываемых одинаковыми уравнениями. Различие в уравнениях состоит лишь в различной физической интерпретации переменных, входящих в эти уравнения. Рассмотрим примеры сходства математических описаний процессов различной физической природы: электрических, механических, гидравлических, процессов динамики жидкости и газа и пр. (табл. 5.1).

Законы Кулона и Ньютона описывают силу, действующую на единичные электрические заряды q_1 и q_2 или тела, массой m_1 и m_2, находящиеся друг от друга на расстояниях r, достаточно больших по сравнению с геометрическими размерами несущих их тел.

Уравнения электрической и гидравлической цепей описывают поведение потока зарядов (электрического тока J_e) и потока жидкости (гидравлического тока J_g) сквозь электрические и гидравлические сопротивления r_e, r_g, индуктивности L_e, L_g и емкости C_e, C_g. В этом случае аналогом напряжения электрической цепи является давление P в расчете на единицу объемной плотности ρ_m текущей среды, заполняющей гидравлическую цепь.

Таблица 5.1

Закон Кулона $F = \frac{q_1 q_2}{4\pi\varepsilon r^2}$	Закон Ньютона $F = \frac{\gamma m_1 m_2}{r^2}$
Уравнение электрической цепи $u = J_e r_e + \frac{L_e dJ_e}{dt} + \frac{1}{C_e}\int J_e dt$	Уравнение гидравлической цепи $\frac{P}{\rho_m} = J_g r_g + \frac{L_g dJ_g}{dt} + \frac{1}{C_g}\int J_g dt$
Принцип непрерывности электрического тока $\oint_s j_e dS = -\frac{dg}{dt}$	Принцип непрерывности потока жидкости $\oint_s j_g dS = -\frac{dm}{dt}$
Закон сохранения энергии $W_e = \rho_e\left(Eh + k_e v^2\right) + \Delta W_T = \mathrm{const}$	Уравнение Бернулли $P = \rho_m(gh + v^2/2) + \Delta P_T = \mathrm{const}$

Принципы непрерывности электрического тока и неразрывности жидкости означают, что суммарный поток зарядов J_e и суммарный ток жидкости J_g, проходящие в единицу времени сквозь любую замкнутую поверхность S, в точности равны скорости изменения соответственно заряда или массы внутри этой поверхности. Эти принципы представляют собой законы сохранения вещества (заряда или массы) при любых преобразованиях, кроме аннигиляции.

Из закона сохранения энергии в установившемся режиме течения следует, что плотность полной электрической W_e или механической P энергии постоянна вдоль всей цепи, хотя и может перераспределяться между потенциальной, кинетической и тепловой формами. В уравнениях используются обозначения: ρ_e и ρ_m – объемные плотности заряда и массы; g – ускорение в поле земного тяготения; h – расстояние от источника электрической энергии до данного места цепи или высота уровня жидкости над данным местом гидравлической цепи соответственно; k_e – коэффициент, характеризующий геометрию цепи; v – скорость течения зарядов или жидкости.

Убедившись в схожести описаний процессов различной физической природы, можно заменить исследования одних из них изучением других. Так, вместо громоздкого и сложного экспериментирования с механическими объектами можно проводить опыты с электрической схемой, исследовать при этом различные варианты, не переделывая механическую конструкцию. Примером косвенного подобия является проведение исследований на аналоговых вычислительных машинах. В свое время масштабы использования аналоговых вычислительных машин были очень широки.

Третий класс моделей – это модели, подобие которых оригиналу не является ни прямым, ни косвенным, а устанавливается в результате соглашения. Такое подобие называется условным. Примерами условного подобия являются чертежи (модели будущих объектов), карты (модели местности), сигналы (модели сообщений). Условное подобие не требует фактического сходства, но несмотря на это, оно должно строиться с учетом особенностей человека – создателя и потребителя моделей условного подобия.

5.2. Основные понятия физического подобия

Виды физического подобия

Теория подобия включает в себя такое обширное понятие как физическое подобие, которое объединяет геометрическое, динамическое, кинематическое, тепловое и другие виды подобия. При геометрическом подобии отношение любых сходных отрезков равно одному и тому же постоянному числу. Иными словами изучаемый объект подобен первоначальному, когда он получается путем изображения его в другом геометрическом масштабе. Кинематическое подобие означает, что в любых сходных точках систем скорости движущихся объектов параллельны и пропорциональны друг другу, т.е. отношение между их скоростями одинаково во всех точках системы. Если система рассматривается как состоящая из отдельных элементов, то у подобных систем отношение масс элементов между собой представляет постоянное число. Динамическое подобие заключается в параллельности и пропорциональности сил в сходных точках. Тепловое подобие означает пропорциональность друг другу всех характеризующих тепловое явление величин: температур, теплоемкостей, тепловых потоков, коэффициентов теплопроводности и т.д.

Приведем математическую формулировку подобия. Введем обозначения. Пусть l_1 и l_2 – сравниваемые отрезки длин первого и второго объектов, v_1 и v_2 – скорости объектов, m_1 и m_2 – массы, f_1 и f_2 – сравниваемые силы. Тогда можно записать

$$\frac{l_1}{l_2} = c_l, \quad \frac{v_1}{v_2} = c_v, \quad \frac{m_1}{m_2} = c_m, \quad \frac{f_1}{f_2} = c_f.$$

Коэффициенты, определяющие отношения длин c_l, скоростей c_v, масс c_m и сил c_f, называются константами подобия. Для каждого вида величин константы имеют свою особую численную величину.

В общем случае подобие явлений, процессов или систем определяется как пропорциональность друг другу всех величин, характеризующих данные объекты. Коэффициенты пропорциональности при этом сохраняют постоянное значение во всех точках системы для величин определенного наименования, но могут принимать отличные значения для величин разного наименования.

Подобных объектов может быть не два, а значительное количество, т.е. они могут составлять группу подобных объектов. Сравнивая все члены группы с одним объектом, который является образцовым или базовым, можно выявить закономерность: при переходе от одного, подобного базовому, объекта к другому константы подобия могут принимать разные значения. Но при этом сохраняется свойство постоянности констант во всех точках каждой системы, подобной базовому образцу. Объединяя переход от явлений образца ко всем подобным ему, можно рассматривать его выражение $x_2^i = c_{xi} x_1^i$ как групповое преобразование явления.

Инварианты подобия

Подобие явлений можно выражать не только константами подобия, но и так называемыми инвариантами подобия. Для пояснения понятия инварианта подобия необходимо перейти от абсолютной системы единиц измерений к относительной. С этой целью требуется отнормировать все величины каждого из подобных объектов. При этом за базовое значение принимаются характеристики объектов, измеренные в сходных точках, например, объект характеризуется линейными размерами длиной l_1, шириной d_1 и высотой h_1. Возьмем один из параметров за базовый, остальные отнормируем относительно него, получим

$$L_1 = l_1/h_1, \ D_1 = d_1/h_1, \ H_1 = h_1/h_1 = 1.$$

Аналогичные действия проведем для объекта, находящегося в отношении подобия к первому объекту, при этом в качестве базового возьмем аналогичный параметр, что и для первого объекта, в нашем случае это высота. В результате получим

$$L_2 = l_2/h_2, \ D_2 = d_2/h_2, \ H_2 = h_2/h_2 = 1.$$

Поскольку первый и второй объекты находятся в отношении подобия, то для них выполняется условие $l_1/l_2 = d_1/d_2 = h_1/h_2 = c_l$, откуда получаем

$$l_1/h_1 = l_2/h_2, \ d_1/h_1 = d_2/h_2,$$

10 — 4355

следовательно $L_1=L_2$, $D_1=D_2$. Точно такие же соотношения можно получить для любых других характеристик объектов, находящихся в отношении подобия друг к другу. Так для скоростей процессов будем иметь $v_1^i/v_1^0 = V_1^i$, $v_2^i/v_2^0 = V_2^i$, где нижний индекс означает номер объекта, а верхний индекс – номер точки, в которой производятся измерения: при этом одинаковыми индексами обозначены результаты измерений в сходных точках. Проводя такую же процедуру, как это было сделано в случае линейных размеров, получим $V_1^i = V_2^i$. Для масс объектов будет справедливо выражение $M_1^i = M_2^i$. Таким образом, получен результат, заключающийся в том, что значения соответствующих характеристик подобных объектов, выраженные в относительных единицах измерения и рассчитанные в сходных точках, равны друг другу. Эти величины и называются инвариантами подобия.

Необходимо различать понятия «константа подобия» и «инвариант подобия». Константа сохраняет постоянное значение во всех точках системы, но она изменяется, когда одна пара подобных явлений заменяется другой. Инвариант подобия, наоборот, различен для разных точек системы, так как он является отображением одной из величин этой системы, имеющей разное численное значение в разных точках системы. Инвариант подобия не меняется при переходе от одного явления к любому другому, подобному ему, т.е. сохраняет одно и то же значение в сходных точках всей группы подобных явлений.

Константы подобия не являются произвольными величинами. Характер взаимосвязи величин, входящих в константы подобия, определяется закономерностью физического явления и выражается в виде уравнений. Наличие взаимосвязи между величинами налагает определенные ограничения и на константы подобия.

Различают полное, неполное, приближенное и другие виды подобия, используемые в соответствующих способах моделирования. Полное подобие и соответственно способ моделирования в формализованном виде характеризуют тот случай, когда между всеми параметрами образцового объекта и модели выполняется соотношение $y_i = m_i x_i$, где y_i – i-й параметр системы-оригинала; x_i – i-й параметр модели; m_i – масштабный коэффициент, который обычно является постоянной величиной.

Неполное подобие и соответственно способ моделирования характеризуются частичным подобием протекания основных процессов в системе и модели или только в пространстве, или только во времени. Приближенное подобие имеет место в том случае, когда для части параметров системы и модели инварианты подобия приблизительно равны друг другу, т.е. значения инвариантов системы и модели укладываются в некоторые границы критической области.

Совокупность масштабных коэффициентов перехода от модели к натурному образцу представляет собой масштабный фактор системы в целом. Изменение масштабного коэффициента перехода от модели к образцу для одного из параметров приводит, как правило, к изменению масштабного фактора физической системы пропорционально значению этого параметра для исследуемого процесса. В сложной неоднородной системе при расчете масштабного фактора должны учитываться взаимовлияния подсистем. При этом необходимо обращать внимание на проверку критериев подобия в широком диапазоне изменения параметров в процессе проведения исследований. Для однородных моделей, полученных на основе теории подобия, как правило, выполняются требования автомодельности. Автомодельность – это свойство явления автоматически сохранять подобие явлению-оригиналу независимо от абсолютных величин параметров элементов той или иной системы, в которой данное явление протекает. Согласно теории подобия, если модель и натурный объект подобны, то они описываются одинаковыми критериями и эти критерии тождественны. Если отношения подобия между моделью и оригиналом установлены, то результаты, полученные при исследовании модели, можно переносить на объект-оригинал.

5.3. Формирование критериев физического подобия

До сих пор речь шла об установлении подобия объектов на основании сравнения определяющих параметров, характеризующих свойства сравниваемых объектов. Однако в теории подобия известны результаты, которые говорят о том, что количество сравниваемых величин можно уменьшить, сформировав так называемые критерии подобия. Критерием подобия π назовем безразмерный (т.е. нулевой размерности) функционал, зависящий от нескольких определяющих параметров объекта (двух и более):

$$\pi = \Phi(p_1, p_2, \dots, p_n).$$

Рассмотрим способ формирования критериев подобия для некоторого явления, процесса или системы. Пусть рассматриваемый объект характеризуется n параметрами $p_1, p_2, \dots, p_n$. Каждый параметр представляет собой некоторую измеримую физическую величину, для которой определена шкала измерения и установлена размерность. Размерность величины находится при помощи определительного уравнения, которое описывает эту величину в математической форме. Например, определительное уравнение для скорости имеет вид $v = dL / dT$, где L – расстояние; T – время; для силы $F = Mg$, где M, g – соответственно масса и ускорение. Размерность величины будем указывать при помощи символа, взятого в квадратные скобки, так, например, для размерности скорости движения объекта запишем $[v] = [L][T]^{-1}$, где $[L]$, $[T]$ – соответственно размерности длины и времени; для размерности силы получим $[F] = [M][L][T]^{-2}$, где $[M]$ – размерность массы. Короче говоря, размерность любой физической величины представляет собой произведение возведенных в степень размерностей первичных величин.

Рассмотрим механические системы, для которых первичными единицами измерения являются длина $[L]$, время $[T]$ и масса $[M]$. Размерность любого определяющего параметра можно выразить через эти единицы:

$$[p_i] = [L]^{\alpha_i}[T]^{\beta_i}[M]^{\gamma_i}.$$

Критерий подобия определим как комбинацию величин p_i, $i = \overline{1, n}$, т.е. $\pi = p_1^{z_1} p_2^{z_2} \ldots p_n^{z_n}$, или выразив через размерности соответствующих величин, будем иметь:

$$\pi = c[p_1]^{z_1}[p_2]^{z_2}\ldots[p_n]^{z_n},$$

где c – безразмерная величина, численно равная произведению значений параметров. Выразив далее размерность определяющих параметров системы через первичные единицы измерения и подставив эти выражения в критерий подобия, получим:

$$\pi = c[[L]^{\alpha_1}[T]^{\beta_1}[M]^{\gamma_1}]^{z_1}[[L]^{\alpha_2}[T]^{\beta_2}[M]^{\gamma_2}]^{z_2}\ldots[[L]^{\alpha_n}[T]^{\beta_n}[M]^{\gamma_n}]^{z_n}.$$

Группируя выражения одной размерности, получим

$$\pi = c[[L]^{\alpha_1 z_1 + \alpha_2 z_2 + \ldots + \alpha_n z_n}[T]^{\beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_n z_n}[M]^{\gamma_1 z_1 + \gamma_2 z_2 + \ldots + \gamma_n z_n}].$$

Поскольку критерий подобия был определен как безразмерный функционал, можно записать соотношение, связывающее степени у размерностей:

$$\begin{aligned}
\alpha_1 z_1 + \alpha_2 z_2 + \ldots + \alpha_n z_n &= 0;\\
\beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_n z_n &= 0;\\
\gamma_1 z_1 + \gamma_2 z_2 + \ldots + \gamma_n z_n &= 0.
\end{aligned}$$

Получена система из трех уравнений. Ранг системы равен трем; таким образом, число независимых переменных равно $n - r$, где r – количество первичных величин, определяющих систему единиц измерения, в рассматриваемом примере $r = 3$. Это число независимых переменных определяет количество критериев подобия, необходимых и

достаточных для решения задачи о подобии оригинала и модели.

Таким образом, для определения критериев подобия необходимо выявить число определяющих параметров, которые характеризуют исследуемый процесс, составить матрицы размерностей каждого параметра, установить число независимых между собой параметров, представить описание исследуемого явления в критериальной форме, записать выражения критериев подобия. Проиллюстрируем формирование критериев подобия на примере системы вынужденных механических колебаний с демпфированием. Пусть груз массой M колеблется на пружине жесткостью C, причем при перемещении его на расстояние l в вязкой среде появляется сила сопротивления, пропорциональная скорости v: $F_c = -Kv$. На груз действует возмущающая сила $F_b = F\sin(\omega t)$. Участвующих величин семь: $p_1 = M$, $p_2 = \omega$, $p_3 = F$, $p_4 = l$, $p_5 = K$, $p_6 = C$, $p_7 = t$. Функциональная зависимость, подлежащая исследованию, имеет вид $\Phi(M, \omega, F, l, K, C, t)$. Выберем три независимые единицы применительно к системе измерений LMT. Пусть $p_1 = M$, $p_2 = \omega_0$, $p_3 = F_0$, тогда получаем систему

$$\begin{aligned}
&[M]=[M]^{-1}[L]^{0}[T]^{0};\\
&[\omega]=[M]^{0}[L]^{0}[T]^{-1};\\
&[F]=[M]^{1}[L]^{1}[T]^{-2}.
\end{aligned} \tag{5.1}$$

Для остальных параметров будем иметь

$$\begin{aligned}
&[l]=[M]^{0}[L]^{1}[T]^{0};\\
&[K]=[M]^{1}[L]^{0}[T]^{-1};\\
&[C]=[M]^{1}[L]^{0}[T]^{-2};\\
&[t]=[M]^{0}[L]^{0}[T]^{1}.
\end{aligned}$$

Правильность выбора числа независимых параметров проверяется путем вычисления определителя, составленного на основании системы (5.1):

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & -2 \end{bmatrix} = 1.$$

Поскольку определитель не равен нулю, то выбранное значение числа независимых параметров, равное трем, оказывается верным, и величины M, ω, F_0 действительно независимы. Дальнейшие действия заключаются в определении формы записи критериев подобия согласно системе уравнений (5.1) и отыскании числовых значений показателей степеней в выражении для размерностей соответствующих параметров. Применительно к данному примеру будем иметь

$$\pi_1 = \frac{[l]}{[M]^{\alpha_1}[\omega]^{\beta_1}[F]^{\gamma_1}};$$

$$\pi_2 = \frac{[K]}{[M]^{\alpha_2}[\omega]^{\beta_2}[F]^{\gamma_2}};$$

$$\pi_3 = \frac{[C]}{[M]^{\alpha_3}[\omega]^{\beta_3}[F]^{\gamma_3}};$$

$$\pi_4 = \frac{[t]}{[M]^{\alpha_4}[\omega]^{\beta_4}[F]^{\gamma_4}}.$$

Определим значения показателей степеней α_i, β_i, γ_i. Выразив размерность каждой из составляющих, входящих в формулы приведенной системы, через первичные единицы измерения, получим

$$\pi_1 = \frac{[L]}{[M]^{\alpha_1}[T]^{-\beta_1}[M]^{\gamma_1}[L]^{\gamma_1}[T]^{-2\gamma_1}}.$$

Приравнивая показатели одноименных первичных единиц измерения, стоящих в числителе и знаменателе, получим $\gamma_1 = 1$, $\alpha_1 = -1$, $\beta_1 = -2$. Далее для второго критерия

$$\pi_2 = \frac{[M][T]^{-1}}{[M]^{\alpha_2}[T]^{-\beta_2}[M]^{\gamma_2}[L]^{\gamma_2}[T]^{-2\gamma_2}},$$

откуда получаем $\gamma_2 = 0$, $\alpha_2 = 1$, $\beta_2 = 1$. Для третьего критерия

$$\pi_3 = \frac{[M][T]^{-2}}{[M]^{\alpha_3}[T]^{-\beta_3}[M]^{\gamma_3}[L]^{\gamma_3}[T]^{-2\gamma_3}},$$

и можно определить показатели степеней $\gamma_3=0$, $\alpha_3=1$, $\beta_3=2$. И, наконец, для последнего критерия

$$\pi_4 = \frac{[T]}{[M]^{\alpha_4}[T]^{-\beta_4}[M]^{\gamma_4}[L]^{\gamma_4}[T]^{-2\gamma_4}},$$

показатели степеней равны $\gamma_4 = 0$, $\alpha_4 = 0$, $\beta_4 = -1$. Критерии подобия, таким образом, будут иметь следующий вид:

$$\pi_1 = \frac{Ml\omega^2}{F},\ \pi_2 = \frac{K}{M\omega},\ \pi_3 = \frac{C}{M\omega^2},\ \pi_4 = \omega t.$$

Поскольку в рассматриваемой системе имеется семь определяющих параметров, а выделено четыре независимых критерия, то можно сказать, что на основании данных критериев подобия имеется возможность сформировать новые группы независимых между собой критериев. Имеется в виду независимость критериев внутри группы. Критерии, взятые из разных групп, будут зависимыми. Новую группу независимых критериев можно построить, например, перемножив некоторые критерии из исходной группы.

Таким образом, показано, что для решения задачи определения подобия объектов (оригинала и модели) сравнивается не множество определяющих параметров, а множество критериев, размерность которого меньше, чем размерность множества определяющих параметров.

5.4. Элементы статистической теории подобия

В общем случае параметры исследуемых систем, процессов или явлений могут представлять собой случайные величины. Поэтому необходимо применять основные положения теории подобия с учетом стохастического характера процессов и явлений, происходящих в объектах. Принципы подобия в стохастическом смысле основаны на том, что сравниваемые параметры являются случайными величинами, а критерии подобия – функциями этих случайных величин.

Рассмотрим постановку задачи определения подобия системы-оригинала и модели. Пусть имеется соотношение [29] вида

$$\rho = \rho_1 \wedge \rho_2 \wedge \ldots \wedge \rho_m, \qquad (5.2)$$

где $\rho_i = \frac{x_{1i}}{x_{2i}}$, $i = \overline{1, m}$ – статистики критериев, сформированные для каждой группы одноименных параметров; $\{x_{1i}\}$ – параметры системы-оригинала; $\{x_{2i}\}$ – параметры модели.

Считается, что системы подобны, если отношение (5.2) равно единице. Параметры, входящие в выражение (5.2), в общем случае являются случайными величинами. В такой постановке можно говорить о равенстве критерия ρ единице только с некоторой долей вероятности. Если критерий ρ является непрерывной случайной величиной, то вероятность того, что $\rho = 1$, в точности равна 0. В стохастической постановке принято считать, что две системы подобны, когда функции распределения параметров, характеризующих эти системы, равны, а статистика критерия подобия находится в пределах верхней и нижней границ доверительного интервала. Вектор параметров системы в общем случае представляет собой набор функциональных и конструктивно-технологических параметров системы. Примерами таких параметров могут служить физические характеристики: коэффициент теплопередачи, рабочее давление, число оборотов вала турбины, параметры, характеризующие габаритные размеры объекта и т.п. В качестве параметров могут использоваться обобщенные параметры, например, характеристики надежности типа среднего времени до отказа (наработка на отказ), коэффициент готовности, вероятность выполнения задачи и т.д.

Примеры параметров промышленных объектов. Проведем анализ параметров на примере конкретных промышленных объектов. Рассмотрим агрегаты, входящие в структуру атомных электростанций (АЭС). Так при анализе модульных парогенераторов АЭС определяющими параметрами являются количество трубок в модуле, количество модулей, длина трубки, толщина стенок трубки, коэффициент теплопередачи, параметры рабочего тела. При исследовании подобия турбин определяющими параметрами являются мощность турбины, габаритные размеры, параметры пара, такие как влажность, температура и пр. Для электронных устройств помимо конструктивного подобия необходимо анализировать электрические параметры: напряжение и силу тока на входе и выходе, коэффициент усиления и т.п.

Изложим способ получения численного значения статистик критерия подобия. Пусть случайные параметры $x_{11}, x_{12}, \ldots, x_{1m}$ имеют плотность распределения $f_{1i}(x_{1i})$, соответственно параметры $x_{21}, x_{22}, \ldots, x_{2m}$ имеют плотность $f_{2i}(x_{2i})$. Тогда в терминах задачи проверки статистических гипотез нулевая гипотеза H_0 будет состоять в том, что $f_{1i}(x_{1i}) = f_{2i}(x_{2i})$ для всех одноименных параметров.

Альтернативная гипотеза заключается в том, что $f_{1i}(x_{1i}) \neq f_{2i}(x_{2i})$. Для критерия подобия можно записать, что сравниваемые системы подобны, если $\hat{\rho} \in [\rho_n, \rho_b]$. Здесь ρ_n, ρ_b – соответственно нижняя и верхняя границы доверительного интервала для критерия подобия, определяемые с некоторым уровнем значимости.

Для определения численного значения статистики критерия ρ при справедливости гипотезы H_0 необходимо по известным плотностям распределения случайных величин x_{1i} и x_{2i} определить плотности распределения $f_{1i}(x_{1i})$ и $f_{2i}(x_{2i})$, от этих плотностей перейти к плотности распределения величины ρ и, наконец, вычислить выборочное значение статистики $\hat{\rho}$ и границы критической области $[\rho_n, \rho_b]$.

Итак, пусть каждый объект характеризуется некоторым числом параметров x_i, $i = \overline{1, m}$. Перепишем формулу (5.2) в следующем виде:

$$\rho = \rho_1 \wedge \rho_2 \wedge \ldots \wedge \rho_m = \frac{x_{11}}{x_{21}} \wedge \frac{x_{12}}{x_{22}} \wedge \ldots \wedge \frac{x_{1m}}{x_{2m}}.$$

Иными словами сравнению подлежат параметры одного типа. Например, сравнивают длину трубок парогенераторов оригинала и модели, толщину стенок этих парогенераторов, их теплопроводность и т.д.

Множество величин ρ_i представляет собой набор независимых случайных величин. Тогда объекты – оригинал и модель – находятся в отношении подобия, если они подобны по каждому определяющему параметру.

Теперь задача сводится к определению доверительного интервала по каждому критерию подобия ρ_i с одним и тем же уровнем значимости. Если такие интервалы найдены и для каждого критерия выполняются условия $\hat{\rho} \in [\rho_n, \rho_b]$, то объекты можно считать подобными. Если же хотя бы по одному критерию условие подобия не выполняется, то оснований считать объекты подобными нет.

Следует заметить, что часть анализируемых параметров может иметь детерминированный характер. Так, в примере с парогенератором можно считать, что длина трубок, толщина стенок трубок – величины неслучайные, поэтому при анализе подобия объектов по этим параметрам решение тривиальное, не требующее привлечения аппарата случайных функций. Подобие по этим параметрам определяется путем деления параметров объекта-оригинала на соответствующие параметры модели. Таким образом, размерность задачи (5.2), требующей анализа с использованием аппарата случайных функций, существенно понижается.

Методика определения плотности распределения величины ρ состоит в следующем. Итак $\hat{\rho} = \frac{x_{1i}}{x_{2i}}$ – выборочное значение статистики критерия, определенное по результатам испытаний системы оригинала и модели, $F_{\rho_i}(\kappa)$ – функция распределения величины ρ_i. Не нарушая общности, опустим индекс в записи для функции распределения, так как аналогичные выкладки будут иметь место для любых параметров системы, имеющих характер случайных величин. Запишем выражение для функции распределения:

$$F_\rho(\kappa) = P(\rho \le \kappa) = P\left(\frac{x_1}{x_2} \le \kappa\right).$$

Величины x_1 и x_2, независимы, следовательно, их совместная плотность распределения есть произведение плотностей этих величин. Согласно [31], вероятность соотношения $\frac{x_1}{x_2} \le \kappa$ выражается интегралом от совместной плотности по области, определенной неравенствами $x_2 > 0$, $x_1 \le x_2\kappa$:

$$F_\rho(\kappa) = \iint\limits_{\substack{x_2>0 \\ x_1 \le x_2\kappa}} f_1(y) f_2(z) dy dz. \quad (5.3)$$

Будем считать, что функция $F_\rho(\kappa)$ дифференцируема, т.е. существует плотность $f_\rho(\kappa)$.

Точные доверительные границы определяются из соотношений

$$\int\limits_0^{\rho_n} f_\rho(\kappa / H_0) d\kappa = \alpha; \qquad \int\limits_{\rho_b}^{\infty} f_\rho(\kappa / H_0) d\kappa = \alpha, \quad (5.4)$$

где α – уровень значимости.

В данных соотношениях неизвестными величинами являются значения границ доверительного интервала ρ_n, ρ_b, относительно которых необходимо решить интегральные уравнения (5.4). Определив функцию распределения статистики критерия (5.3) можно ограничиться вычислением приближенных границ доверительного интервала. Для этого необходимо получить плотность распределения статистики критерия ρ, вычислить математическое ожидание

$$m_\rho = \int\limits_{\Omega_\rho} \kappa f_\rho(\kappa) d\kappa$$

и среднее квадратическое отклонение статистики критерия

$$\sigma_\rho^2 = \int\limits_{\Omega_\rho} \kappa^2 f_\rho(\kappa) d\kappa - m_\rho^2,$$

где Ω_ρ – область определения критерия ρ. В общем случае область определения критерия $\Omega_\rho = [0, \infty]$. Определив математическое ожидание и среднее квадратическое отклонение, можно вычислить приближенный доверительный интервал по формулам $\rho_n = m_\rho - t_\alpha \sigma_\rho$, $\rho_b = m_\rho + t_\alpha \sigma_\rho$, где t_α – квантиль распределения Стьюдента, определенный для уровня значимости α.

Рассмотрим методику проверки гипотезы о подобии объектов оригинала и модели. Решение о подобии объектов будем принимать на основании сравнения оценок, полученных на различных этапах исследования, а именно, при функционировании объекта-оригинала и при исследовании модели. Особо подчеркнем, что в качестве сравниваемых параметров фигурируют оценки. Поскольку оценки определяются путем обработки выборки случайных чисел ограниченного объема, то они сами являются случайными величинами, имеющими плотность распределения $f_x(t)$. В ряде случаев плотность распределения оценок характеристик довольно просто получить на основании плотности распределения исходной случайной величины. Известно [30], что если оценка характеристики выражается через достаточную статистику $M(T)$ (через $M(T)$ в математической статистике принято обозначать сумму случайных величин), то для определения ее плотности распределения можно воспользоваться аппаратом характеристических функций. Такая ситуация имеет место при определении средней наработки, интенсивности отказа, вероятности отказа, коэффициента готовности и ряда других параметров системы. Рассмотрим два случая распределения наблюдаемых случайных величин.

1. Случайная величина (например, наработка до отказа) имеет гамма-распределение $t \sim \Gamma(t, \lambda, \alpha)$ с параметрами: λ – параметр масштаба и α – параметр формы. Определим плотность распределения средней величины. Рассмотрим функционирование объекта-оригинала. Пусть за время его работы наблюдаемая случайная величина реализовалась k раз. Положим, что все случайные величины независимы и распределены по гамма-закону

$$\Gamma(t, \lambda_1, \alpha_1) = \frac{\lambda_1^{\alpha_1} t^{\alpha_1 - 1} \exp(-\lambda_1 t)}{\Gamma(\alpha_1)}.$$

Характеристическая функция для данной плотности имеет вид

$$\varphi_1(y) = \left(\frac{\lambda_1}{\lambda_1 - iy} \right)^{\alpha_1}.$$

Характеристическая функция величины $X = \sum\limits_{j=1}^{k} T_j$, где k – объем статистических данных, определяется из соотношения

$$\varphi_X(y) = \left(\frac{\lambda_1}{\lambda_1 - iy} \right)^{k\alpha_1}.$$

Применяя к данному выражению обратное преобразование Фурье, получаем плотность распределения случайной величины X

$$f(x,\lambda_1,\alpha_1)=\frac{\lambda_1^{k\alpha_1}x^{k\alpha_1-1}\exp(-\lambda_1 x)}{\Gamma(k\alpha_1)}. \qquad (5.5)$$

Перейдем от плотности распределения случайной величины X, представляющей собой сумму случайных величин, к величине $T_{ср}$, являющейся средней величиной

$$T_{ср}=\frac{\sum_{j=1}^{k}T_j}{k}.$$

Плотность распределения средней величины будет иметь вид

$$f_T(t,\lambda_1,\alpha_1)=\frac{(k\lambda_1)^{k\alpha_1}t^{k\alpha_1-1}\exp(-k\lambda_1 t)}{\Gamma(k\alpha_1)}. \qquad (5.6)$$

Аналогичное выражение получим для распределения оценки, рассчитанной по результатам испытаний модели:

$$f_{Tm}(t,\lambda_2,\alpha_2)=\frac{(n\lambda_2)^{n\alpha_2}t^{n\alpha_2-1}\exp(-n\lambda_2 t)}{\Gamma(n\alpha_2)}, \qquad (5.7)$$

здесь n – объем статистических данных, полученных при проведении исследования модели. Определим теперь функцию распределения статистики критерия ρ, для чего подставим выражения (5.6) и (5.7) в (5.3) и получим

$$F_\rho^{\Gamma}(\kappa)=\iint\limits_{\substack{t_2>0\\ t_1\le t_2\kappa}} f_1(t_1)f_2(t_2)dt_1dt_2=$$

$$=\iint\limits_{\substack{t_2>0\\ t_1\le t_2\kappa}}\frac{(k\lambda_1)^{k\alpha_1}t_1^{k\alpha_1-1}\exp(-k\lambda_1 t_1)}{\Gamma(k\alpha_1)}\frac{(n\lambda_2)^{n\alpha_2}t_2^{n\alpha_2-1}\exp(-n\lambda_2 t_2)}{\Gamma(n\alpha_2)}dt_1dt_2.$$

Перейдем в данном выражении к повторному интегралу и произведем замену переменных $t_2=v$, $t_1=uv$, получим

$$F_\rho^{\Gamma}(\kappa)=\int_0^{\kappa}\frac{(n\lambda_2)^{n\alpha_2}(k\lambda_1)^{k\alpha_1}u^{k\alpha_1-1}}{\Gamma(n\alpha_2)\Gamma(k\alpha_1)}\int_0^{\infty}v^{n\alpha_2+k\alpha_2-1}\exp\{-v(k\lambda_1 u+n\lambda_2)\}dvdu.$$

Внутренний интеграл в данном выражении представляется через гамма-функцию следующим образом:

$$\int_0^{\infty}v^{n\alpha_2+k\alpha_1-1}\exp\{-v(k\lambda_1 u+n\lambda_2)\}dv=\frac{\Gamma(n\alpha_2+k\alpha_1)}{(n\lambda_2+k\lambda_1 u)^{n\alpha_2+k\alpha_1}}.$$

Таким образом, функция распределения критерия ρ будет иметь вид

$$F_\rho^{\Gamma}(\kappa)=\int_0^{\kappa}\frac{(n\lambda_2)^{n\alpha_2}(k\lambda_1)^{k\alpha_1}\Gamma(n\alpha_2+k\alpha_1)u^{k\alpha_1-1}}{\Gamma(n\alpha_2)\Gamma(k\alpha_1)(n\lambda_2+k\lambda_1 u)^{n\alpha_2+k\alpha_1}}du.$$

Соответствующая данной функции плотность распределения получается дифференцированием данного выражения по κ и равна

$$f_\rho^{\Gamma}(\kappa)=\frac{(n\lambda_2)^{n\alpha_2}(k\lambda_1)^{k\alpha_1}\Gamma(n\alpha_2+k\alpha_1)\kappa^{k\alpha_1-1}}{\Gamma(n\alpha_2)\Gamma(k\alpha_1)(n\lambda_2+k\lambda_1\kappa)^{n\alpha_2+k\alpha_1}}. \qquad (5.8)$$

При справедливости гипотезы H_0 должны выполняться соотношения $\lambda_1=\lambda_2=\lambda$, $\alpha_1=\alpha_2=\alpha$, тогда условная плотность распределения статистики критерия будет иметь вид

$$f_\rho^{\Gamma}(\kappa/H_0)=\frac{(k/n)^{n\alpha}\Gamma((n+k)\alpha)\kappa^{k\alpha-1}}{\Gamma(n\alpha)\Gamma(k\alpha)(\kappa k/n+1)^{(n+k)\alpha}}. \qquad (5.9)$$

Точные верхнюю и нижнюю границы доверительного интервала ρ_n, ρ_b определяем, решая уравнения (5.4), подставляя в них выражение (5.9). Данные уравнения решаются численными методами, например, методом последовательных приближений. Гамма-распределение является довольно общим распределением. К семейству гамма-распределений относятся распределение Рэлея, экспоненциальное распределение, χ^2-распределение. Таким образом, полученные результаты могут быть обобщены на ряд других законов распределения случайной величины, для которой формируется критерий подобия.

2. Второй случай характеризует ситуацию, когда наблюдаемые случайные величины подчиняются нормальному распределению. Методику разработки критерия будем рассматривать на примере сравнения средних арифметических выборочных данных, полученных при функционировании оригинала и модели. Пусть наблюдаемые случайные величины имеют плотность распределения соответственно

$$f_1^N(t)=\frac{1}{\sqrt{2\pi}s_1}\exp\left(-\frac{(t-m_1)^2}{2s_1^2}\right);$$

$$f_2^N(t)=\frac{1}{\sqrt{2\pi}s_2}\exp\left(-\frac{(t-m_2)^2}{2s_2^2}\right),$$

где m_1, s_1 – параметры плотности распределения случайной величины, полученной при наблюдении за объектом-оригиналом; m_2, s_2 – параметры плотности распределения случайной величины, полученной при исследовании модели. Тогда выборочные средние значения, определяемые из выражения $X = \sum_{j=1}^{k} T_j$, будут иметь соответственно следующие выражения для плотностей распределения:

$$f_1^N(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-m_1)^2}{2\sigma_1^2} \right);$$

$$f_2^N(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x-m_2)^2}{2\sigma_2^2} \right),$$

где $\sigma_1^2 = \frac{s_1^2}{k}, \sigma_2^2 = \frac{s_2^2}{n}$.

Запишем функцию распределения статистики критерия ρ:

$$F_\rho^N(\kappa) = \iint\limits_{x_2>0} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x_1-m_1)^2}{2\sigma_1^2} \right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x_2-m_2)^2}{2\sigma_2^2} \right) dx_1 dx_2 .$$

Переходим в данном выражении к повторному интегралу, выполняя при этом замену переменных $x_2 = v$, $x_1 = uv$:

$$F_\rho^N(\kappa) = \int_0^\kappa du \int_0^\infty v \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(uv-m_1)^2}{2\sigma_1^2} \right) \exp\left(-\frac{(v-m_2)^2}{2\sigma_2^2} \right) dv. \quad (5.10)$$

Произведем в данном выражении следующую замену переменных: $\mu = \frac{m_1}{u}$, $s = \frac{\sigma_1}{u}$, далее индексы у параметров для простоты опустим. Запишем показатель у экспоненты, предварительно суммируя слагаемые:

$$\nu = \frac{(v-\mu)^2\sigma^2 + (v-m)^2 s^2}{2s^2\sigma^2} = \frac{v^2\sigma^2 - 2v\mu\sigma^2 + \mu^2\sigma^2 + v^2 s^2 - 2vms^2 + m^2 s^2}{2s^2\sigma^2}.$$

Приведем подобные члены в данном выражении:

$$\nu = \frac{v^2(\sigma^2 + s^2) - 2v(\mu\sigma^2 + ms^2) + \mu^2\sigma^2 + m^2 s^2}{2s^2\sigma^2}.$$

Разделим числитель и знаменатель на (σ^2+s^2) , получим

$$\nu = \frac{v^2 - 2v(\mu\sigma^2 + ms^2)/(\sigma^2 + s^2) + (\mu^2\sigma^2 + m^2 s^2)/(\sigma^2 + s^2)}{2s^2\sigma^2/(\sigma^2 + s^2)}.$$

В числителе данного выражения добавим и отнимем величину $\left(\frac{(\mu\sigma^2 + ms^2)}{(\sigma^2 + s^2)} \right)^2$, тогда выражение примет вид

$$\nu = \frac{1}{2s^2\sigma^2/(\sigma^2 + s^2)} \times \left[v^2 - 2v\frac{(\mu\sigma^2 + ms^2)}{(\sigma^2 + s^2)} + \left(\frac{(\mu\sigma^2 + ms^2)}{(\sigma^2 + s^2)} \right)^2 - \right.$$

$$\left. - \left(\frac{(\mu\sigma^2 + ms^2)}{(\sigma^2 + s^2)} \right)^2 + \frac{(\mu^2\sigma^2 + m^2 s^2)}{(\sigma^2 + s^2)} \right] =$$

$$= \frac{1}{2s^2\sigma^2/(\sigma^2 + s^2)} \times \left[\left(v - \frac{(\mu\sigma^2 + ms^2)}{(\sigma^2 + s^2)} \right)^2 - \left(\frac{(\mu\sigma^2 + ms^2)}{(\sigma^2 + s^2)} \right)^2 + \frac{(\mu^2\sigma^2 + m^2 s^2)}{(\sigma^2 + s^2)} \right].$$

В первом слагаемом получили квадрат разности двух выражений, причем выделили переменную v. Второе слагаемое получилось не зависящим от интегрируемой переменной. Упростим второй член данной суммы, для чего вынесем за скобку $(\sigma^2+s^2)^{-2}$, тогда получим

$$\nu_2 = \frac{(\mu^2\sigma^2 + m^2 s^2)(\sigma^2 + s^2) - (\mu\sigma^2 + ms^2)^2}{2s^2\sigma^2(\sigma^2 + s^2)} =$$

$$= \frac{m^2 s^4 + m^2 s^2\sigma^2 + \mu^2\sigma^2 s^2 + \mu^2\sigma^4 - \mu^2\sigma^4 - 2\mu\sigma^2 ms^2 - m^2 s^4}{2s^2\sigma^2(\sigma^2 + s^2)}.$$

Приведем подобные члены и получим простое выражение

$$\nu_2 = \frac{(\mu - m)^2}{2(\sigma^2 + s^2)}.$$

Преобразуем выражение (5.10) с учетом полученных результатов:

$$F_\rho^N(\kappa) = \int_0^\kappa du \int_0^\infty v \frac{1}{\sqrt{2\pi}\sigma_1} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{(\mu - m)^2}{2(\sigma^2 + s^2)} \right] \exp\left[-\frac{\left(v - \frac{\mu\sigma^2 + ms^2}{\sigma^2 + s^2} \right)^2}{2\sigma^2 s^2/(\sigma^2 + s^2)} \right] dv .$$

Первая экспонента в данном выражении не зависит от v, поэтому ее можно вынести за знак внутреннего интеграла. Тогда получим

$$F_\rho^N(\kappa)=\int_0^\kappa \frac{1}{\sqrt{2\pi}\sigma_1}\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(\mu-m)^2}{2(\sigma^2+s^2)}\right]du\int_0^\infty v\exp\left[-\frac{\left(v-\dfrac{\mu\sigma^2+ms^2}{\sigma^2+s^2}\right)^2}{2\sigma^2s^2/(\sigma^2+s^2)}\right]dv.$$

Внутренний интеграл в данной формуле выражается через функцию Лапласа. Покажем это. Для этого введем обозначение:

$$z=\frac{v-\dfrac{\mu\sigma^2+ms^2}{\sigma^2+s^2}}{\sqrt{2}\dfrac{\sigma s}{\sqrt{\sigma^2+s^2}}}.$$

Тогда интеграл можно записать в виде

$$I_2=\sqrt{2}\frac{\sigma s}{\sqrt{\sigma^2+s^2}}\int_a^\infty v\exp(-z^2)dz,$$

где нижний предел интегрирования равен $a=-\dfrac{\mu\sigma^2+ms^2}{\sqrt{2(\sigma^2+s^2)}s\sigma}$.

По определению функция Лапласа имеет вид

$$\Phi(y)=\frac{2}{\sqrt{\pi}}\int_0^y \exp(-z^2)dz.$$

Следовательно, можно записать

$$I_2=\frac{\sqrt{\pi}s\sigma}{(\sigma^2+s^2)}\frac{(\mu\sigma^2+ms^2)}{\sqrt{2(\sigma^2+s^2)}}\left[1+\Phi\left[\frac{(\mu\sigma^2+ms^2)}{\sqrt{2(\sigma^2+s^2)}s\sigma}\right]\right]+$$
$$+\frac{s^2\sigma^2}{\sigma^2+s^2}\exp\left[-\left(\frac{(\mu\sigma^2+ms^2)}{\sqrt{2(\sigma^2+s^2)}s\sigma}\right)^2\right].$$

Подставляя полученное значение в (5.10) и переходя к прежним обозначениям, будем иметь

$$F_\rho^N(\kappa)=\int_0^\kappa \frac{1}{\sqrt{2\pi}\sigma_1}\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(m_1u-m_2)^2}{2(\sigma_1^2u^2+\sigma_2^2)}\right)\times$$
$$\times\left\{\frac{\sqrt{\pi}\sigma_1\sigma_2}{\sqrt{2(\sigma_1^2u^2+\sigma_2^2)}}\frac{(m_2\sigma_1^2u+m_1\sigma_2^2)}{\sigma_1^2u^2+\sigma_2^2}\left[1+\Phi\left[\frac{(m_2\sigma_1^2u+m_1\sigma_2^2)}{\sqrt{2(\sigma_1^2u^2+\sigma_2^2)}\sigma_1\sigma_2}\right]\right]+\right.$$
$$\left.+\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2u^2+\sigma_2^2}\exp\left[-\left(\frac{(m_2\sigma_1^2u+m_1\sigma_2^2)}{\sqrt{2(\sigma_1^2u^2+\sigma_2^2)}\sigma_1\sigma_2}\right)^2\right]\right\}du.$$

Соответствующая плотность распределения статнстнки критерия выглядит следующим образом:

$$f_\rho^N(u)=\frac{1}{\sqrt{2\pi}\sigma_1}\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(m_1u-m_2)^2}{2(\sigma_1^2u^2+\sigma_2^2)}\right)\times$$
$$\times\left\{\frac{\sqrt{\pi}\sigma_1\sigma_2}{\sqrt{2(\sigma_1^2u^2+\sigma_2^2)}}\frac{(m_2\sigma_1^2u+m_1\sigma_2^2)}{\sigma_1^2u^2+\sigma_2^2}\left[1+\Phi\left[\frac{(m_2\sigma_1^2u+m_1\sigma_2^2)}{\sqrt{2(\sigma_1^2u^2+\sigma_2^2)}\sigma_1\sigma_2}\right]\right]+\right.$$
$$\left.+\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2u^2+\sigma_2^2}\exp\left[-\left(\frac{(m_2\sigma_1^2u+m_1\sigma_2^2)}{\sqrt{2(\sigma_1^2u^2+\sigma_2^2)}\sigma_1\sigma_2}\right)^2\right]\right\}.$$

При условии справедливости нулевой гипотезы должны выполняться соотношения $m_1=m_2$, $s_1=s_2$. Следовательно, условная плотность распределения статистики критерия при условии справедливости нулевой гипотезы будет иметь вид

$$f_\rho^N(u/H_0)=\frac{1}{2\pi}\exp\left(-\frac{r^2(u-1)^2}{2(u^2/k+1/n)}\right)\times$$
$$\times\left\{\frac{\sqrt{\pi}r}{\sqrt{2(u^2/k+1/n)}}\frac{(u/k+1/n)}{u^2/k+1/n}\left[1+\Phi\left[\frac{r\sqrt{nk}(u/k+1/n)}{\sqrt{2(u^2/k+1/n)}}\right]\right]+\right. \tag{5.11}$$
$$\left.+\frac{1}{\sqrt{nk}(u^2/k+1/n)}\exp\left[-\left(\frac{r\sqrt{nk}(u/k+1/n)}{\sqrt{2(u^2/k+1/n)}}\right)^2\right]\right\},$$

здесь $r=m/s$.

Зная плотность распределения статистики критерия подобия, можно вычислить по аналогии с предыдущим случаем границы доверительного интервала для заданного уровня значимости. Для этого необходимо подставить выражение (5.11) в (5.4) и решить уравнения относительно верхней и нижней границ доверительного интервала. Нормальный закон распределения также является общим случаем, частными случаями которого будут усеченное нормальное распределение, логарифмическое нормальное распределение.

Такнм образом, проверяя по каждой группе одноименных параметров гипотезу о их подобии, решаем задачу подобия системы оригинала и модели.

Глава 6

ЭКСПЕРИМЕНТ – СРЕДСТВО ПОСТРОЕНИЯ МОДЕЛИ

6.1. Характеристика эксперимента

В основе любого исследования лежит эксперимент. Для построения модели объекта необходима информация о нем. Средством получения информации являются наблюдения за объектом исследования. Наблюдения могут проводиться как с помощью пассивных способов их организации (то, что в философии называется простое созерцание), так и в процессе специально организованных исследований.

Отношение между экспериментом и моделью такое же, как между курицей и яйцом, – нельзя определить, что было «в самом начале». Эксперимент с некоторым объектом проводится с целью уточнения его модели. С другой стороны, постановка эксперимента определяется имеющейся до опыта моделью. Противоречия в данном высказывании нет, так как в ходе проведения эксперимента исследователь получает новую информацию, позволяющую усовершенствовать модель, развивать и усложнять ее, т.е. в модели появляются новые составляющие, отражающие более полно процессы, явления и эффекты взаимодействия, которые ранее не были учтены в модели.

Термин «эксперимент» обычно используется при:

• целенаправленном наблюдении исследуемого явления в точно учитываемых условиях, позволяющих следить за ходом явления и воссоздавать его каждый раз при повторении этих условий;

• преднамеренных действиях или операциях, предпринятых с целью установления неизвестных причин, их проверки или иллюстрации;

• воспроизведении объекта познания, организации особых условий его существования;

• наблюдении развития явления в естественных для него условиях.

Общей чертой в характеристике эксперимента является то, что он определяется как осмысленная деятельность человека, связанная с различными целями, средствами и объектами познания. Наиболее общей целью проведения экспериментов является получение новой информации об изучаемом явлении (процессе, объекте). Обобщенно эксперимент определяется как форма познания объективной действительности. Он является одним из основных способов научного исследования наряду с теоретическим мышлением, наблюдением, математическими расчетами и др. Отличительной особенностью эксперимента является использование при его проведении специальных средств исследования, позволяющих исследователю осуществлять вмешательство в явления и процессы внешнего мира, воспроизводить ход процесса, планомерно изменять различные условия в целях получения искомого результата. Эксперимент характеризуется определенной направленностью и организованностью, что сводит к минимуму элемент случайности, неожиданности, хотя полностью его не исключает. Таким образом, *эксперимент – это совокупность действий исследователя, осуществляемая посредством материальных средств исследования с целью получения новой информации об изучаемом объекте (процессе, явлении) путем построения информационных (описательных) моделей, характеризующих различные его стороны и проявления.*

Основные элементы эксперимента: 1) экспериментатор и его деятельность как познающего субъекта; 2) объект экспериментального исследования; 3) средства экспериментального исследования.

Различают пассивный и активный эксперименты. Пассивный эксперимент подразумевает сбор исходного статистического материала в режиме нормальной эксплуатации объекта наблюдения. Активный эксперимент ставится по заранее составленному плану с использованием методов планирования эксперимента. При этом предусматривается одновременное изменение всех параметров, влияющих на процесс, что приводит к сокращению общего числа опытов. Для проведения активного эксперимента требуются специальные установки. Примерами таких установок могут служить термобарокамеры, вибростенды, аэродинамические трубы и пр. Эксперимент может быть управляемым и неуправляемым. Одним из основных принципов организации научных экспериментальных иселедований является стремление к изоляции изучаемого объекта от влияния окружающей среды, т.е. проведение контролируемого активного эксперимента. В таком эксперименте независимые переменные могут варьироваться по желанию исследователя, а влияние внешних переменных исключается. Управляемый эксперимент предполагает управляемость объекта исследований, которая определяется возможностью перевода объекта с наперед заданной точностью в любое из различимых состояний, в котором он находится в течение требуемого промежутка времени. При этом под состоянием объекта понимается все множество значений его характеристик и соотношений

между ними, присущих ему в данный момент времени. В неуправляемом эксперименте наблюдатель пассивно фиксирует спонтанно протекающие процессы.

Различают также контролируемый и неконтролируемый эксперименты. В контролируемом эксперименте независимые переменные, воздействующие на объект исследования и называемые факторами, могут быть измерены с достаточно высокой точностью. Неконтролируемый эксперимент характеризуется тем, что исследователь предполагает воздействие некоторых факторов внешней среды, но у него нет технической возможности произвести количественные измерения уровней воздействующих факторов.

В реальных условиях любой, даже самым тщательным образом организованный, эксперимент обладает свойствами неуправляемого и неконтролируемого эксперимента. Никакая исследовательская установка не может быть полностью изолирована от воздействия факторов внешней среды. Часть этих факторов может быть измерена, но сами факторы будут неуправляемы, другая же часть факторов не подлежит измерению и, таким образом, является неконтролируемой. К неконтролируемым факторам, как правило, относятся такие, как уровень солнечной радиации, естественный радиоактивный фон и магнитное поле Земли и т.п. Данные факторы при проведении активных экспериментов, как правило, не учитываются. Наличие неуправляемых и неконтролируемых факторов может привести к нарушению воспроизводимости результатов эксперимента в сериях – повторениях. Воспроизводимость опыта – одно из главных требований, предъявляемых при организации экспериментальных исследований. Воспроизводимость означает, что в ходе проведения эксперимента различия в выходных результатах опытов, проводимых в условиях воздействия одних и тех же факторов на одних и тех же уровнях, обусловлены случайными факторами: погрешностью приборов, ошибками измерения, дрейфом параметров в экспериментальной установке. Если различия в выходных результатах велики и не объяснимы случайными воздействиями, это означает, что имеет место влияние неучитываемого фактора, которое ведет к нарушению воспроизводимости эксперимента. Если неуправляемые факторы оказывают влияние на результат, получаемый в ходе эксперимента, то эксперимент будет активно-пассивным.

Рассмотрим примеры экспериментальных исследований. Для отработки радиоэлектронной аппаратуры, эксплуатируемой на движущихся объектах в специальных условиях (например, в условиях Крайнего Севера, пустынях, на высокогорье, на летательных аппаратах, морских судах при повышенной влажности и пр.), проводят эксперименты данной аппаратуры в термобарокамерах с установленными в них вибростендами. При обычных условиях внешней среды установки такого типа можно достаточно хорошо изолировать от влияния внешних факторов. Эксперимент в данном случае будет активным. Но, если в ходе проведения эксперимента случится землетрясение, то естественная вибрация Земли наложится на вибрацию, моделируемую вибростендом, и это приведет к нарушению воспроизводимости результатов эксперимента.

Другой пример с физическими экспериментами, проводимыми на ускорителях или исследовательских реакторах. К нарушению воспроизводимости таких экспериментов могут привести резкие всплески солнечной активности. В данном случае происходит наложение активности внешних источников на активность объектов исследования.

Несмотря на большое разнообразие научно-технических экспериментов, отличающихся по физической природе, используемым техническим средствам и объему задач исследований, можно выделить их некоторые общие свойства.

1. Процесс экспериментальных исследований обязательно связан с непосредственным участием человека-исследователя и диктуется его интересами. Исследователь формулирует постановку задачи системного анализа, определяет план его проведения, разрабатывает алгоритм обработки результатов исследования и принимает решение о дальнейших действиях, т.е. человек определяет ход системных исследований на каждом из этапов его проведения.

2. Специфической особенностью экспериментов любой разновидности является наличие неопределенности, обусловленной уровнем априорной информации об исследуемом объекте и степенью ее достоверности. Основной целью проведения экспериментов является получение новых знаний, новой информации и соответственно понижение степени неопределенности.

3. Результаты каждого эксперимента всегда имеют некоторый элемент неопределенности, который вносится ограниченностью экспериментального материала. Его оценка проводится путем статистического анализа результатов наблюдений. Если целью эксперимента является построение модели исследуемого процесса, то выполнение этой цели достигается с определенной точностью. Таким образом, всегда имеется элемент неопределенности в формировании результата эксперимента, обусловленный случайностью исследуемых процессов, объектов или явлений и ограниченностью числа опытов.

4. Любой научно-технический эксперимент ведет к определенным действиям исследователя – принятию решения по продолжению или прекращению исследований – и заканчивается представлением результатов, формулировкой выводов, выдачей рекомендаций. Процесс принятия решений в экспериментальных исследованиях не удается полностью формализовать даже в самых простых ситуациях.

5. Сложность объекта исследования определяется числом различных состояний, в которых он может находиться. Сложность объекта характеризуется уровнем его организации, степенью детерминированности. Какими бы сложными ни были те или иные эксперименты, по форме организации они мало различаются и включают в себя этапы планирования эксперимента, его проведение и анализ результатов.

6. Общий принцип организации экспериментальных исследований – системный подход. Элементами такой организации должны стать планирование эксперимента, исключение или учет случайных воздействий окружающей среды, анализ получаемых результатов с оценкой ошибок и их совокупного влияния, проверка приемлемости результатов и их интерпретация, представление полученных данных в упорядоченном и наглядном виде. Повышению эффективности именно этих сторон процесса экспериментальных исследований должна способствовать его автоматизация.

Отметим, что современная теория системного анализа трактует понятие эксперимента несколько шире классического, предусматривающего лишь количественные, однозначные измерения. Выделяют следующие черты эксперимента.

1. Имеются наблюдаемые явления, в принципе не допускающие числовой меры, но которые можно фиксировать в «качественных», «слабых» шкалах. Результаты таких экспериментов, однако, можно учитывать в моделях, получая качественные, но вполне научные выводы.

2. Неотъемлемым природным свойством некоторых наблюдений признана их расплывчатость. Тем не менее, таким наблюдениям придана строгая математическая форма и разработан формальный аппарат работы с ними.

3. Погрешности измерений являются неотъемлемым естественным свойством самого процесса измерения, обусловленным наличием неопределенностей, шумами аппаратуры, квантованием измеряемых сигналов.

4. Широкое распространение получили статистические измерения, т.е. оценивание функционалов распределений вероятностей по реализации случайного процесса.

Путем обработки результатов наблюдений, фиксируемых в ходе проведения эксперимента, генерируется информация для включения в модель с целью ее усовершенствования. Таким образом производится перевод модели на более высокий качественный уровень.

6.2. Классификация экспериментальных исследований

Эксперименты, описываемые совокупностью однотипных свойств, целесообразно объединить в некоторые классы. Наиболее полный вариант классификации научно-технических экспериментов приведен в [33]. В этой работе приняты следующие обобщенные классификационные признаки: структура эксперимента; стадия научных исследований, к которым относится эксперимент; организация эксперимента; постановка задачи; способ проведения эксперимента. Данный набор признаков, по признанию самого автора, не является единственно возможным и охватывающим все многообразие свойств экспериментов. Изложим схему классификации научно-технических экспериментов по [33], дополнив ее. В качестве первого уровня классификации рассмотрим качественный и количественный эксперименты. **Качественный** эксперимент – более простой вид экспериментов. Его цель – установление только факта существования явления. Качественный эксперимент реже обставляется сложными измерительными системами и системами обработки данных. Но кажущаяся простота качественного эксперимента пропадает, если изучаемое явление или процесс является стохастическим (случайным). Стохастичность может быть вызвана, во-первых, тем, что уровень шумов, на фоне которых измеряется полезный сигнал, одного порядка или даже выше значения самого сигнала. Во-вторых, стохастичность может лежать в основе самого процесса.

Количественный эксперимент встречается чаще, чем качественный. Требует для своего проведения более сложного оборудования. Задачей измерительного или количественного эксперимента является установление количественных связей между параметрами, описывающими состояние системы.

Следующий уровень – разделение экспериментов по их структуре на натурные, модельные и модельно-кибернетические (машинные). В **натурном** эксперименте средства экспериментального исследования взаимодействуют непосредственно с объектом исследования, в **модельном** эксперименте – не с самим объектом, а с его моделью. При этом модель играет двоякую роль. Во-первых, она является непосредственно объектом экспериментального исследования. Во-вторых, по отношению к подлинному изучаемому объекту или процессу модель выступает в качестве средства экспериментального исследования. **Модельно-кибернетический** эксперимент является разновидностью модельного, при котором соответствующие характеристики изучаемого объекта исследуются с помощью модели на ЭВМ.

Эксперименты на моделях можно, в свою очередь, подразделить на масштабное, аналоговое, полунатурное и математическое моделирование.

Масштабное моделирование. Этот вид экспериментальных исследований один из самых старых. Чтобы качественно или количественно изучить явление, делали его модель (уменьшенную копию). Примером масштабного моделирования может служить изучение поведения гидротехнических сооружений, потоков жидкости в трубопроводах, устойчивости судов при воздействии на них течений различной направленности. Достоинством данного вида моделирования является изучение явлений и процессов в натуре. Недостаток масштабного моделирования состоит в том, что геометрическое подобие не обеспечивает подобия явления.

Аналоговое моделирование. Следующий тип моделирования – исследования, проводимые на аналоговых моделях. Если различные явления описываются одними и теми же уравнениями, то можно одно из явлений выбрать за основу модели, а остальные выражать через него. Модельным выбирается то явление или процесс, в котором можно легче и точнее произвести измерения. Так как лучше всего разработаны измерения электрических величин, то и модели стараются выполнить на электросхемах (моделирование на аналоговых вычислительных машинах).

Полунатурное моделирование. Полунатурное моделирование чаще всего применяется при исследовании систем автоматического или полуавтоматического регулирования или управления. Примером может служить исследование характеристик самолетов на специальных стендах по обработке навыков в управлении объектом, скажем, автопилот. На основе полунатурного моделирования создаются различные тренажеры.

Математическое моделирование. Если удается выразить весь моделируемый процесс в форме математических уравнений и отношений, то проблема может облегчиться тем, что эти математические уравнения и отношения исследуются на ЭВМ. В этом случае экспериментатор уже сам распоряжается планом проведения эксперимента: какие параметры и как надо варьировать, а какие стабилизировать. Эксперимент ведется в строгих рамках принятых допущений и введенных в рассмотрение параметров. Составляя математическую модель нужно стремиться оставлять для рассмотрения лишь наиболее существенные параметры, делать математическое описание процесса как можно проще.

Следующий уровень предполагает деление экспериментов согласно стадиям проведения научных исследований. Здесь можно выделить лабораторные, стендовые и промышленные эксперименты.

К **лабораторным** относятся эксперименты по изучению общих закономерностей различных явлений и процессов, по проверке научных гипотез и теорий. Лабораторный эксперимент характеризуется небольшим числом измерительных и управляющих каналов, малыми энергетическими затратами экспериментальной установки, немногочисленным штатом обслуживающего персонала.

При лабораторном эксперименте велика роль самого экспериментатора. Установка для экспериментального исследования, как правило, создается им самим и находится в его подчинении на все время исследования. Этот фактор определяет и сравнительно низкий коэффициент ее загрузки, так как часть времени она простаивает (в период анализа полученных результатов или перемонтажа и наладки оборудования).

Стендовые исследования проводят при необходимости изучить вполне конкретный процесс, протекающий в исследуемом объекте с определенными физическими, химическими и другими свойствами. При стендовых исследованиях на основе сведений, полученных на стадии лабораторных экспериментов, уточняются характеристики объекта, его поведение при варьировании факторов, воздействующих на объект, определяются оптимальные условия функционирования объекта исследования. По результатам стендовых испытаний судят о различных наработках при расчетах или проектировании объекта, изделия или технического процесса. Также в ходе стендовых исследований вырабатываются рекомендации относительно серийного выпуска изделия и условий его эксплуатации.

Разновидностью стендовых исследований является **сложный исследовательский** эксперимент. Ускорители, реакторы, химические колонны – примеры сложных экспериментальных установок для исследовательского эксперимента.

Промышленный эксперимент проводят при создании нового изделия или организации технологического процесса по данным лабораторных или стендовых исследований, при оптимизации технологического процесса, при проведении контрольно-выборочных испытаний качества выпускаемой продукции. Этот вид эксперимента по своему принципу является как бы зеркальным отображением математического моделирования. В математическом моделировании экспериментальным инструментом является ЭВМ. На ней по составленным уравнениям и значениям параметров, выбранных в качестве определяющих и полученных из измерительного эксперимента, воспроизводится исследуемый процесс. В промышленном эксперименте экспериментальная установка (например, аэродинамическая труба, прочностной стенд и т.п.) применяется для сложного измерительного эксперимента. В нем тип исследуемого процесса и уравнения, его описывающие, известны. Но сам процесс настолько сложен, что произвести его математическое моделирование при современном уровне средств вычислительной техники оказывается невозможным. С появлением более мощных ЭВМ часть наиболее простых промышленных экспериментов заменяется математическим моделированием.

Информативность промышленного эксперимента, как правило, велика, что в сочетании со сложностью обработки данных делает его чрезвычайно трудоемким. Промышленный эксперимент может быть модельным, полунатурным и натурным.

Следующий признак классификации учитывает организацию экспериментов. По данному классификационному признаку выделяют обычные (рутинные), специальные (технические), уникальные и смешанные эксперименты, проводимые в стационарных условиях или на подвижных объектах.

Наиболее часто встречаются **обычные** эксперименты. Такие эксперименты выполняются по стандартным методикам с использованием сравнительно простого локального экспериментального оборудования.

Технические эксперименты связаны с созданием и исследованием различных приборов и аппаратов.

Уникальные эксперименты проводятся на сложном дорогостоящем экспериментальном оборудовании (типа ядерного реактора, синхрофазотрона, аэродинамической трубы). Такие эксперименты отличаются большими объемами экспериментальных данных, высокой скоростью протекания исследуемых процессов, широким диапазоном изменения характеристик объектов исследования.

Смешанный эксперимент обладает особенностями разных типов экспериментов. Названные разновидности экспериментов организуются как в стационарных условиях, так и на подвижных объектах (морских, авиационных, космических, наземных).

Следующий уровень классификации экспериментальных исследований по признаку, определяемому их частными целями, или по типу моделей определенного вида, восстанавливаемых по результатам исследований. Говорят также, что данный уровень классификации осуществляется по постановке задачи определения вида модели.

В отдельных случаях экспериментатора могут интересовать установление наличия связей между некоторыми переменными объекта исследования, вида взаимосвязей между ними, конкретных аналитических зависимостей, количественно описывающих объект исследования, уточнение вида и параметров этих аналитических зависимостей. Проведение эксперимента приводит к понижению степени неопределенности в априорно известной модели исследуемого объекта. Постановка задачи конкретного экспериментального исследования определяется уровнем сложности исследуемого объекта, количеством и качеством априорной информации об объекте, т.е. степенью его изученности, особыми условиями существования объекта, например, подверженностью случайным неконтролируемым внешним воздействиям, наличием дрейфа характеристик и т.д. и требуемой степенью детализации его описания. Эти общие принципы постановки задачи рассматриваются как составные элементы признака классификации.

В постановке задачи можно выделить ее характеристики, достаточно общие и вместе с тем выражающие наиболее существенные отличия данной постановки от других. На основе таких характеристик формируются классы экспериментов. Критерием формирования определенного класса экспериментов на данном уровне схемы классификации будем считать существование некоторого набора методов, позволяющих решить задачу с учетом существенных характеристик ее постановки.

На данном уровне классификации можно выделить следующие классы.

1. Эксперименты по нахождению модели объекта исследования при наличии неоднородностей разного вида.

2. Эксперименты по нахождению модели объекта исследования при взаимосвязанных входных переменных.

3. Эксперименты по нахождению модели объекта исследования при наличии у него «памяти», т.е. свойства сохранять последействие.

4. Эксперименты по нахождению модели объекта исследования при выяснении механизма явлений.

5. Эксперименты по нахождению модели объекта исследования, описывающей локальную область пространства его параметров, соответствующую экстремуму некоторого критерия оптимальности при наличии временного дрейфа параметров.

6. Эксперименты по нахождению модели объекта исследования, описывающей локальную область пространства его параметров, соответствующую экстремуму некоторого критерия оптимальности при отсутствии временного дрейфа параметров.

7. Эксперименты по нахождению модели объекта исследования, описывающей степень влияния входных переменных на выходные переменные.

8. Эксперименты по нахождению математической модели объекта исследования, позволяющей преобразовать набор переменных объекта исследования.

9. Эксперименты по нахождению математической модели объекта исследования, прогнозирующей его поведение.

10. Эксперименты по нахождению моделей классификации объектов исследования и проверки степени соответствия экспериментальных данных определенным известным моделям.

Наконец, последний уровень схемы классификации делит эксперименты по способу их проведения, определяющему характер взаимодействия системы автоматизации с объектом исследований. С этой точки зрения различают пассивный, активный с программным управлением, активный с обратной связью, активно-пассивный эксперименты.

Характеристика активного и пассивного экспериментов дана ранее. Активный эксперимент с программным управлением проводится по заранее разработанному плану. В соответствии с этим планом исследователь воздействует на факторы, влияние которых исследуется, переводя их с одного уровня на другой согласно плану эксперимента. При этом изменение функции отклика, отражая реакцию исследуемого объекта на управляющие воздействия, позволяет выяснить природу происходящих в объекте исследования процессов. В случае активного эксперимента с обратной связью система автоматизации интерпретирует результаты на каждом шаге эксперимента и выбирает оптимальную стратегию управления им. Активно-пассивный эксперимент характеризуется тем, что при его проведении часть факторов просто контролируется, а по другой части осуществляется управление.

6.3. Обработка экспериментальных данных

Результаты любого эксперимента фиксируют в той или иной форме, затем их используют с целью обработки. Операции сбора и обработки в одних случаях могут быть совмещены во времени, в других случаях обработка экспериментальных данных является самостоятельным этапом. Практически совмещенными во времени сбор и обработка данных являются в автоматизированных системах управления научными исследованиями и комплексными испытаниями, проводимыми в реальном масштабе времени. Отдельным этапом работ обработка данных выступает при проведении учебных экспериментов, на этапе обобщения результатов научных исследований, при проведении системного анализа.

Методы обработки экспериментальной информации зависят от того, какова модель, для уточнения которой проводится эксперимент. Фактически обработка экспериментальных данных – это преобразование информации к виду, удобному для использования, перевод результатов наблюдений с языка измерений на язык уточняемой модели. Модель, в свою очередь, может принадлежать к одному из двух типов: классификационным или числовым моделям. Тип моделей зависит от знаний об объекте, для которого строится модель. Знания могут быть как первоначальными, приближенными, так и достаточно полными, хорошо структурированными, хотя и требующими уточнения. Классификационная модель носит качественный характер, хотя в ней могут участвовать и количественные переменные. Например, классифицируют состояние объекта «работоспособен – неработоспособен» по результатам численных измерений параметров. С другой стороны, в числовых моделях часть переменных может измеряться в слабых шкалах. Рассмотрим особенности экспериментальных данных и их обработки для обоих типов моделей.

Классификационные модели

Классификационные модели являются первичными, исходными формами знания. Человек в своей повседневной деятельности, сталкиваясь с новыми явлениями или предметами, очень часто их распознает, т.е. без особых затруднений относит к тому или иному классу. Увидев животного неизвестной породы, человек тем не менее относит его к определенному типу животных. Человек может читать рукописи, написанные разными людьми, хотя каждый из них имеет свой почерк. Люди узнают своих знакомых даже в случае, если те поменяли прическу, одежду, нанесли макияж. Таким образом, человек выделяя основные признаки, способен относить объекты к тому или иному классу, т.е. решать задачу классификации.

Необходимость решения задачи классификации проявляется во многих сферах человеческой деятельности. Ряд профессий связан исключительно с умением правильно классифицировать ситуации. От врачей требуется умение правильно поставить диагноз больному, криминалисты занимаются идентификацией почерка, археологи устанавливают принадлежность найденных предметов определенной эпохе, геологи по косвенным данным определяют наличие и характер полезных ископаемых. В каждом из перечисленных видов деятельности проявляется умение человека правильно отнести наблюдаемый объект к тому или иному классу. Также и в науке, познание начинается со сравнения изучаемого объекта с другими, выявления сходства и различия между ними. Наблюдаемые данные, полученные в ходе эксперимента, проводимого на классификационном уровне, содержат результаты из-

мерения ряда признаков X для подмножества A объектов, выбранных из множества Γ: каждый объект $a_i \in A \subseteq \Gamma$ обладает значениями признаков

$$x_i = (x_{i_0}, x_{i_1}, \ldots, x_{i_n}) \in \{X_0, X_1, \ldots, X_n\} = X, i = \overline{1, N},$$

где n – число признаков; N – число объектов в A. Каждый признак характеризует конкретное свойство объекта.

Способ обработки данных зависит от цели обработки. Часто приходится решать задачи определения по наблюдаемым значениям признаков $x = (x_1, \ldots, x_n)$ значений ненаблюдаемого признака x_0. Целевыми являются те параметры модели, которые требуется уточнить по экспериментальным данным.

Для построения классификационных моделей решают следующие типы задач: кластеризации, классификации или распознавания образов, упорядочивание объектов и уменьшение размерности модели.

Задача **кластерного анализа** характеризуется следующими условиями: считается, что и границы классов в пространстве признаков, и число классов являются неизвестными. Требуется определить классы исходя из близости, похожести или различия описаний объектов $x_i = (x_{i_0}, x_{i_1}, \ldots, x_{i_n})$. Компоненты вектора X_0 – признаки кластера, значения которых подлежат определению.

В задаче **классификации или распознавания образов** число классов считается заданным. Если границы между классами заданы, то имеем априорную классификацию, если границы требуется оценить по классификационным примерам, то задача будет называться распознаванием образов по обучающей выборке. Целевой признак X_0 имеет значения в номинальной шкале.

При решении задачи **упорядочивания объектов** требуется установить отношение порядка между признаками объектов $x_{10}, x_{20} \ldots, x_{n0}$ (или некоторой их частью) по определенному критерию предпочтения.

Задача **уменьшения размерности** заключается в следующем. Классификационные модели могут учитывать множество предположений, которые еще надо проверить. Список признаков X формируется эвристически и может содержать дублирующие признаки. Поэтому задача состоит в совершенствовании классификационных моделей, в уменьшении размерности модели с помощью отбора наиболее информативных признаков или путем формирования обобщенных признаков.

Числовые модели

Числовые модели отличаются от классификационных рядом особенностей:

1) целевые признаки x_0 измеряются в числовых шкалах;

2) числа x_0 представляют собой функционалы или функции признаков переменных, которые не обязательно имеют числовые выражения;

3) в числовых моделях переменные могут зависеть от времени.

Если в задаче классификации для получения экспериментальной информации необходимо организовывать наблюдения за группой однотипных объектов, то в задаче построения числовых моделей в качестве первичной информации могут присутствовать результаты длительных наблюдений за одним объектом или небольшой по объему группой однотипных объектов. Числовые модели могут задавать связь между переменными как в виде параметрических, так и в виде непараметрических зависимостей. Типичными задачами для числовых моделей являются задачи косвенных измерений и поиска экстремума.

В задаче **косвенных измерений** (или как ее еще называют задачей оценки параметров) требуется по результатам наблюдений $\{x_{ij}\}$ оценить параметр x_0. В отличие от задачи классификации x_0 измеряется не в номинальной шкале, а в числовой. Если статистические данные $\{x_{ij}\}$ представляют собой результаты наблюдения до некоторого момента времени t_0, а x_0 требуется оценить для момента $t > t_0$, то задача оценивания называется *прогнозированием*.

Задача **поиска экстремума** состоит в организации наблюдений за исследуемым процессом таким образом, чтобы по результатам наблюдений $\{x_{ij}(t_k)\}$, $t_k = t_0+k\Delta t$, $k = 0, 1, 2,\ldots$ можно было получить экстремальное значение целевого признака x_0. Задачи такого рода решаются с помощью методов планирования эксперимента.

6.4. Вероятностное описание событий и процессов

Экспериментальные исследования проводят с целью получения новых сведений об объекте анализа. Экспериментальные данные необходимы для того, чтобы устранить неопределенность в знаниях об объекте, для которого производится построение модели. Основной причиной неопределенности является случайность явлений и процессов, происходящих в объектах исследования. Совершенно очевидно, что в природе нет ни одного физического явления, в котором не присутствовали бы в той или иной степени элементы случайности. Как бы точно и тщательно ни были бы фиксированы условия проведения эксперимента, невозможно достигнуть того, чтобы при повторении опыта результаты полностью и в точности совпадали. Случайные отклонения неизбежно сопутствуют любому закономерному явлению. В ряде практических задач этими случайными элементами можно пренебречь, предполагая, что в данных условиях проведения наблюдений явление протекает вполне определенным образом. При этом из множества воздействующих на процесс факторов выделяются самые главные, влиянием остальных факторов пренебрегают. В других исследованиях исход опыта зависит от большого количества факторов, к тому же на исход эксперимента влияют не только сами факторы, но и их сочетание, их взаимодействие. В результате приходим к необходимости изучения случайных явлений, исследованию закономерностей и выяснению причин возникновения случайностей в наблюдаемом явлении. При рассмотрении результатов отдельных экспериментов бывает трудно обнаружить какие-либо устойчивые закономерности. Однако, если рассмотреть последовательность большого числа однородных экспериментов, можно обнаружить некоторые интересные свойства, а именно: если индивидуальные результаты опытов ведут себя непредсказуемо, то средние результаты обнаруживают устойчивость.

Случайность можно определить как вид неопределенности, подчиняющейся некоторой закономерности, которая выражается распределением вероятностей. Зная распределение вероятностей, можно ответить на следующие вопросы: в каком интервале находятся возможные значения случайной величины, каково наиболее вероятное значение случайной величины, каково рассеивание реализовавшихся случайных величин, какова связь между разными реализациями и т.д. Но для того, чтобы определить закон или плотность распределения случайной величины, необходима информация об исследуемом объекте. В основе проведения любых расчетов лежат исходные данные, результаты наблюдений случайной величины или случайного процесса. Измерения случайных величин и процессов по существу есть измерение выходного параметра, характеризующего определенные свойства объекта исследования. На основании таких измерений решаются вопросы восстановления вида и параметров законов распределения, вычисление коэффициентов регрессии и корреляции, восстановление спектральных плотностей и тому подобные расчеты.

Следует отметить, что результаты наблюдений за функционированием сложных систем, каковые являются, в первую очередь, объектом системного анализа, имеют ряд специфических особенностей, приводящих к необходимости применения и разработки неклассических методов анализа. Остановимся на рассмотрении данных особенностей.

Большая размерность массива данных. Для построения модели сложной системы требуется проводить наблюдения за большой группой выходных параметров, причем некоторые параметры могут характеризоваться рядом признаков. Существенным является также необходимость учета фактора времени, т.е. фиксация изменения свойств объекта в зависимости от времени жизни системы. Современные методы организации баз данных на ЭВМ способны решать задачи сбора и хранения данных, но тем не менее проблема размерности все-таки остается.

Разнотипность данных. Разные признаки могут измеряться в различных шкалах. Здесь возникает проблема согласования данных.

Зашумленность данных. Наблюдаемая величина отличается от истинного значения параметра на некоторую случайную величину. Примерами таких зашумляющих факторов могут служить дрейф нуля измерительного прибора, погрешности приборов, наличие помех в каналах передачи информации и т.п. Статистические свойства помех могут не зависеть от измеряемой величины, тогда помехи можно рассматривать как аддитивный шум. В противном случае имеет место неаддитивная или зависимая помеха. Различные варианты зашумленности должны по-разному учитываться при разработке алгоритмов обработки данных.

Отклонения от предположений, искажения результатов. Приступая к обработке данных, аналитик всегда исходит из определенных предположений о природе величин, подлежащих обработке. Любой способ обработки дает результаты надлежащего качества только в том случае, когда обрабатываемые данные отвечают заложенным в алгоритм обработки предположениям. Во-первых, большинство наблюдаемых параметров имеет характер непрерывных величин, но при обработке неизбежно округление данных, что может привести к искажениям результатов. Далее – измерительный прибор может обладать нелинейной характеристикой и если это не учитывается в алгоритме обработки, то итоговые данные будут также иметь искажения. Чтобы повысить качество выводов, получаемых при обработке данных, необходимо обеспечить соответствие свойств данных и требований к алгоритмам их обработки.

Наличие пропущенных значений. Данная ситуация имеет место в том случае, когда часть наблюдений не доводится до реализации наблюдаемого признака. Примерами таких ситуаций могут служить эксперименты по определению надежности группы однотипных изделий. Современные изделия обладают достаточно высоким уровнем надежности и даже длительные по времени наблюдения за их функционированием не приводят к отказам всей совокупности изделий. В результате выборка данных будет иметь характер цензурированной выборки, в которой для части изделий имеется информация о времени их отказа,

для другой же части такой информации нет. Другим примером могут служить социологические исследования, которые допускают либо отсутствие определенных сведений об опрашиваемых субъектах, предполагают возможность неконкретного ответа на вопросы (типа «не знаю»).

Отмеченные особенности поступающей для обработки статистической информации накладывают определенные ограничения на выбор методов и предъявляют требования к разработке специальных алгоритмов ее обработки.

Одним из подходов, позволяющих учитывать различного рода неопределенности при обработке статистической информации, явилась теория статистического интервального оценивания. Ключевым при построении вероятностных моделей является утверждение о том, что в строгом смысле точные средние и вероятности – это параметры статистически устойчивого явления и достигаются они усреднением при неограниченном повторении того же самого явления в независимых и устойчивых условиях. Так как организовать устойчивое повторение затруднительно, а неограниченное число раз просто невозможно, то часто подразумевают мыслимый повтор. Но чтобы проиграть явление в уме или на ЭВМ, нужно более или менее знать физическую модель явления. Реальные же явления таковы, что их внутренние механизмы до конца не поддаются исследованиям, опыты уникальны, их повторы неустойчивы. В результате точные характеристики остаются как идеальное понятие, достигаемое в пределе, применение которого сопровождается многими оговорками. Таким образом, не только неустойчивость явлений, но и любая неабсолютность статистических знаний, такая как недостаточность, неточность, ограниченность, свойственная почти всем реальным задачам, естественно вынуждает переходить к интервальным понятиям.

В отличие от теории вероятностей, освещающей поточечную структуру моделей, интервальный анализ оперирует только имеющейся информацией, всегда конечной, представленной в интервальной, размытой, доверительной форме.

6.5. Описание ситуаций с помощью нечетких моделей

Одна из основных целей построения математических моделей реальных систем состоит в поиске способа обработки имеющейся информации либо для выбора рационального варианта управления системой, либо для прогнозирования путей ее развития. При решении задач системных исследований достаточно часто, особенно при исследовании экономических, социальных, социотехнических систем, в функционировании которых принимает участие человек, значительное количество информации о системе получают от экспертов, имеющих опыт работы с данной или подобными системами, знающих ее особенности и имеющих представление о целях ее функционирования. Эта информация носит субъективный характер и ее представление в терминах естественного языка содержит большое число неопределенностей – «много», «мало», «высокий», «низкий», «очень эффективный» и т.п., которые не имеют аналогов в терминах языка классической математики. Язык традиционной математики, опирающийся на теорию множеств и двузначную логику, недостаточно гибок для представления встречающихся неопределенностей в характеристике объектов. В нем нет средств достаточно адекватного описания понятий, которые имеют неопределенный смысл. Представление подобной информации на языке традиционной математики обедняет математическую модель исследуемой реальной системы и делает ее слишком грубой. В классической математике множество понимается как совокупность элементов (объектов), обладающих некоторым общим свойством, например, множество чисел, не меньших заданного числа, множество векторов, сумма компонент каждого из которых не превосходит единицы и т.д. Для любого элемента при этом рассматривается лишь две возможности: либо элемент принадлежит множеству, т.е. обладает данным свойством, либо не принадлежит множеству и соответственно не обладает рассматриваемым свойством. Таким образом, в описании множества в обычном смысле должен содержаться четкий критерий, позволяющий судить о принадлежности или непринадлежности любого элемента данному множеству. Разработка математических методов отражения нечеткости исходной информации позволяет построить модель, более адекватную реальности.

Одним из начальных шагов на пути создания моделей, учитывающих нечеткую информацию, считается направление, связанное с именем математика Л. Заде [34] и получившее название теории нечетких множеств. Лежащее в основе этой теории понятие нечеткого множества предлагается в качестве средства математического моделирования неопределенных понятий, которыми оперирует человек при описании своих представлений о реальной системе, своих желаний, целей и т.д. Нечеткое множество – это математическая модель класса с нечеткими или размытыми границами. В этом понятии учитывается возможность постепенного перехода от принадлежности к непринадлежности элемента рассматриваемому множеству. Иными словами, элемент может иметь степень принадлежности множеству, промежуточную между полной принадлежностью и полной непринадлежностью. Понятие нечеткого

множества – это попытка математической формализации нечеткой информации с целью ее использования при построении математических моделей сложных систем. В основе этого понятия лежит представление о том, что составляющие данное множество элементы, обладающие общим свойством, могут обладать этим свойством по-разному, в большей или меньшей степени. При таком подходе высказывания типа «элемент принадлежит данному множеству» теряют смысл, поскольку необходимо указать «насколько сильно» или с какой степенью данный элемент принадлежит рассматриваемому множеству. Одним из важных направлений применения этого нового подхода является проблема принятия решений при нечеткой исходной информации.

Идеи теории нечетких множеств нашли развитие в теоретическом направлении, называемом статистикой объектов нечисловой природы. Особенностью этих объектов является то, что для них не определена совокупность арифметических операций. Объекты нечисловой природы лежат в пространствах, не имеющих векторной структуры.

Примерами объектов нечисловой природы являются:

- значения качественных признаков, т.е. результаты кодировки объектов с помощью заданного перечня категорий (градаций);
- упорядочения (ранжировки) экспертами образцов продукции (при оценке ее технического уровня);
- классификации, т.е. разбиения объектов на группы сходных между собой (кластеры);
- бинарные отношения, описывающие сходство объектов между собой, например, сходство тематики научных работ, оцениваемое экспертами с целью рационального формирования экспертных советов внутри определенной области науки;
- результаты парных сравнений или контроля качества продукции по альтернативному признаку («годен» – «брак»), т.е. последовательности из нулей и единиц;
- множества (обычные или нечеткие), например, зоны, пораженные коррозией, или перечни возможных причин аварии, составленные экспертами независимо друг от друга;
- слова, предложения, тексты;
- векторы, координаты которых представляют собой совокупность значений разнотипных признаков, например, результат составления статистического отчета о научно-технической деятельности или заполненная компьютеризированная история болезни, в которой часть признаков носит качественный характер, а часть – количественный;
- ответы на вопросы экспертной, маркетинговой или социологической анкеты, часть из которых носит количественный характер (возмож-

но, интервальный), часть сводится к выбору одной из нескольких подсказок, а часть представляет собой тексты и т.д.

Статистические методы анализа нечисловых данных нашли широкое применение в экономике, социологии, при проведении экспертного анализа. Дело в том, что в этих областях от 50 до 90% данных являются нечисловыми.

6.6. Характеристика и классификация статистической информации

Классическая схема обработки результатов наблюдений, состоит в предположении, что в каждом испытании реализуется наблюдаемый признак. Например, при испытании объектов на надежность каждый объект доводится до отказа. Такая схема является идеализацией реально проводимых исследований. В реальной жизни, в особенности при проведении обследования функционирующих объектов, информация, поступающая на обработку, крайне ограничена. Например, при эксплуатации объектов их стараются не доводить до отказа. Более того, на предприятии, как правило, существует система предупредительных профилактических мероприятий, суть которых заключается в том, чтобы не допустить возникновение отказов изделий в процессе их функционирования. Даже при организации специальных экспериментов с целью определения характеристик надежности партии испытываемой продукции не удается всю партию довести до отказа, так как для этого потребовалось бы большое время проведения эксперимента. Аналогичные данные поступают на обработку и в других областях проведения исследований. Например, в области социологии или психологии для части испытуемых рассматриваемый признак может наблюдаться, для части – нет (скажем, при определении среднего возраста вступления в брак часть анкетируемых может ответить, что до настоящего времени в браке не состоит). При проведении исследований в медицине у части больных за время наблюдения исследуемый признак может не реализоваться. Так, если анализируется воздействие некоторого препарата на состояние больного и фиксируется время, в течение которого наступает выздоровление, то у одних пациентов процесс выздоровления может пойти быстро, у других медленнее, а у некоторой части за время наблюдения он может не наступить. Возможно он реализуется в дальнейшем, но вывод о результатах исследования формируется в данный момент времени и часть наблюдений, таким образом, является не доведенной до конца. Но, несмотря на то, что для части объектов исследования являются не доведены до конца, в них содержится полезная информация, которую необходимо использовать при обработке результатов наблюдений. Данные, для которых имеется неопределенность в наблюдениях за реализацией исследуемого признака, называются цензурированными данными.

Цензурирование – это процесс возникновения неопределенности момента реализации признака объекта (в теории надежности момента отказа), причем интервал неопределенности считается известным. Интервалом неопределенности называется интервал времени, внутри которого произошла либо произойдет реализация наблюдаемого признака объекта, при этом точное значение времени реализации признака объекта неизвестно.

Понятие о цензурированной выборке

Рассмотрим основные понятия и определения, применительно к информации, поступающей на обработку на примере задачи оценивания показателей надежности.

В процессе анализа надежности приходится сталкиваться с ситуациями, когда определенная часть объектов или систем не отказывает за период наблюдения, а другая часть отказывает, но моменты отказов точно неизвестны. В таких ситуациях возникает необходимость проведения статистического анализа надежности на основе специфических выборок, основной особенностью которых является отсутствие сведений о моментах отказов контролируемой части изделий. Это явление носит название цензурированных данных, а получаемые в результате выборки – цензурированными выборками (ЦВ).

Под **данными**, применительно к задачам надежности, понимают фиксированные значения наработок изделий, полученные по результатам испытаний или эксплуатационных наблюдений. Данными цензурированной выборки являются наработки как отказавших объектов, так и неотказавших объектов, а также интервалы времени, в течение которых объект отказал, но момент отказа точно неизвестен.

Цензурированной выборкой называется выборка, элементами которой являются значения наработки до отказа и наработки до цензурирования, либо только значения наработки до цензурирования. Как было отмечено ранее, **цензурирование** – это процесс возникновения неопределенности момента отказа объекта, причем интервал неопределенности известен аналитику. **Интервал неопределенности** – интервал наработки, внутри которого произошел либо произойдет отказ объекта, причем точное значение наработки до отказа неизвестно. Этот

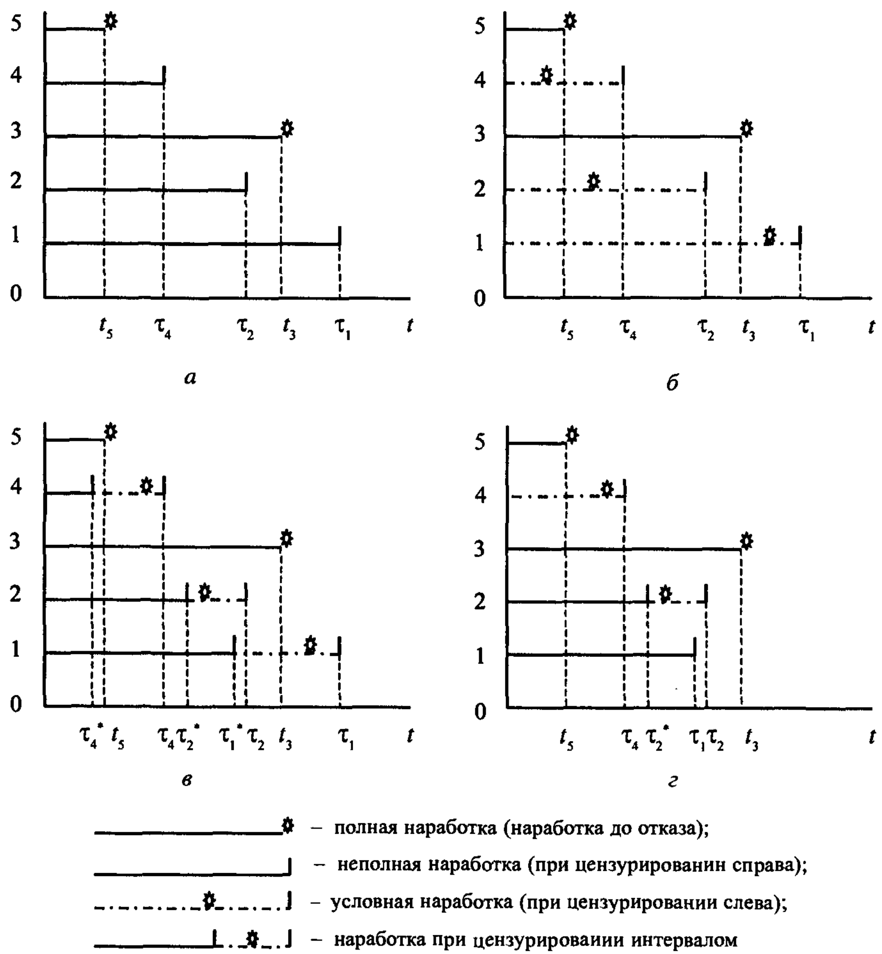

Рис. 6.1. Распределение значений наработок объектов для цензурированной выборки:

а – справа; *б* – слева; *в* – интервалом; *г* – комбинированного

интервал может быть неограниченным справа, тогда говорят о **цензурировании справа**, либо ограниченным справа, тогда говорят о **цензурировании слева**. Если интервал неопределенности момента отказа ограничен слева и справа, то говорят о **цензурировании интервалом**. Следует отметить, что в задачах надежности при цензурировании слева левая граница интервала неопределенности равна нулю, а при цензурировании интервалом – больше нуля. На рис. 6.1 приведены реализации случайных наработок изделий до отказа и до цензурирования.

Необходимо обратить внимание на то, что цензурирование интервалом является наиболее общим видом цензурирования, так как при устремлении правой границы интервала к бесконечности этот вид цензурирования превращается в цензурирование справа, а при устремлении левой границы интервала к нулю – в цензурирование слева. При устремлении границ интервала друг к другу цензурирование исчезает.

Рассмотрим понятие «наработка». **Полная наработка** – это наработка изделия до отказа. **Неполная наработка** – наработка объекта от начала испытаний или эксплуатации до прекращения испытаний или эксплуатационных наблюдений до отказа. **Условная наработка при цензурировании слева** – это значение интервала, измеряемого в единицах наработки, в пределах которого произошел отказ. Эта наработка названа условной потому, что объект может не работать в пределах всего интервала, так как отказ может наступить в некоторой части этого интервала. **Наработка при цензурировании интервалом** складывается из неполной наработки и условной наработки при цензурировании слева.

Причины появления цензурированных данных

Отметим причины появления цензурированных данных на примере обработки результатов наблюдений с целью определения характеристик надежности объектов. Причиной появления цензурированных данных в данном случае является специфика организации функционирования объектов, состоящая в том, что реально функционирующие объекты в процессе работы до отказа стараются не доводить. На предприятии регулярно проводятся планово-профилактические работы (ППР), цель которых состоит в восстановлении работоспособности объектов. В большинстве случаев схема функционирования элементов следующая: в период проведения ППР объекты выводят из работы и на их место ставят новые. Работоспособность снятых объектов восстанавливается до первоначального уровня; если есть необходимость, производят их ремонт, настройку, чистку и прочие мероприятия. При проведении следующих ППР эти устройства ставятся в систему, а объекты, которые находились в работе, выводят для проведения восстановительных мероприятий. Особенностью функционирования является также наличие контроля исправности работы элементов, их замена при достижении определенной наработки, независимо от того, отказал элемент к данному моменту или нет.

Рассмотрим более детально причины возникновения цензурированных данных.

1. Разное время установки в систему и снятие с эксплуатации однотипных объектов. Такая схема организации эксплуатации характерна для элементов, которые выводятся из работы в период проведения ППР независимо от того, отказали они или нет; на их место устанавливаются аналогичные объекты из состава запасных изделий. Указанный режим организации эксплуатации имеет место для большинства объектов системы управления и защиты (СУЗ) энергоблоков АЭС, для ряда контрольно-измерительных приборов и устройств автоматики. Другой причиной может служить наличие резервных каналов и технологических петель, которые временно выводятся из работы и включаются в случае обнаружения отказа на работающем оборудовании.

2. Снятие объектов с эксплуатации из-за отказов составных частей, надежность которых не исследуется. Например, при оценке надежности различных устройств СУЗ, таких как устройство задания мощности, устройство измерения и контроля и т.п., они могут быть сняты с эксплуатации из-за отказов блоков питания, находящихся в соответствующих схемах каналов СУЗ.

3. Переход объектов из одного режима применения в другой в процессе их эксплуатации. Часть объектов используется в течение определенного промежутка времени, и далее наблюдения за их функционированием прекращаются. Примерами могут служить системы локализации аварии, скажем, система аварийного расхолаживания реактора, система аварийного электроснабжения и т.п. Имеются объекты, у которых в течение короткого времени проверяется работоспособность и на этом функционирование прекращается. Так периодически производится опробование дизель-генераторов, после непродолжительной работы их отключают.

4. Необходимость оценки надежности различных систем до наступления отказов ее комплектующих элементов. В настоящее время ко всем системам энергоблоков АЭС, важных для безопасности, предъявляется требование периодической оценки их надежности.

5. Наличие периодического контроля за исправностью функционирования объектов приводит к поступлению информации в отдельные моменты времени на границах интервалов наблюдений. Таким образом, наблюдателю не известно, как ведут себя объекты внутри интервала наблюдения; известными становятся состояния объектов только в моменты контроля. Данная схема наблюдений называется цензурированием интервалом или группированные данные.

6. Ненадежность устройств контроля, которые должны фиксировать отказы отдельных приборов или каналов. Это приводит к тому, что отказы объектов или каналов, у которых, в свою очередь, отказало устройство контроля, выявляются в моменты проведения проверок. Например, отказы обнаруживаются во время проведения ППР.

Цензурирование для группы однотипных объектов может быть задано в одной точке; с другой стороны, могут наблюдаться реализации, когда цензурирование проводится в разных точках. Примером первого случая цензурирования может служить план испытаний $[N, U, T]$ или $[N, U, r]$ [38]. При плане $[N, U, T]$ наблюдения производят за N объектами, длительность наблюдений равна T единиц времени. По истечении этого времени испытания прекращаются не зависимо от того, сколько элементов отказало. Если отказало m из N объектов, то для оставшихся $(N-m)$ объектов время наблюдения будет цензурировано величиной T, т.е. известно, что на интервале $[0, T]$ $(N-m)$ объектов не отказало и, вероятно, их отказы произойдут на интервале $(T, \infty]$.

При плане $[N, U, r]$ наблюдения производят за N объектами. Испытания прекращают тогда, когда откажет r объектов. В этом случае цензурирующим моментом является T_r – момент отказа r-го объекта. Отказы остальных $N-r$ изделий произойдут в интервале $[T_r, \infty]$. Такое цензурирование называется однократным.

Когда цензурирование производится в разных точках, оно называется многократным. Цензурирование может быть случайным и неслучайным. Цензурирование будет неслучайным тогда, когда цензурирующие моменты y_i детерминированы. Например, заранее спланированы моменты времени, в которые начинается профилактика. Примером неслучайного цензурирования являются результаты наблюдения за функционированием объектов, которые проходят испытания по плану $[N, U, T]$, когда момент T остановки испытаний известен заранее. При случайном цензурировании цензурирующие моменты y_i являются реализациями случайной величины. Примером выборки со случайным цензурированием являются результаты наблюдений за испытанием объектов по плану $[N, U, r]$. Здесь момент приостановки наблюдений T_r является случайной величиной и определяется моментом отказа r-го объекта.

Глава 7

ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНОЙ ИНФОРМАЦИИ

7.1. Оценивание показателей систем и определение их точности

При решении вопросов построения моделей систем особую актуальность имеет задача формирования исходной информации о параметрах элементов, входящих в состав системы. От точности и достоверности исходной информации зависит точность оценок анализируемых характеристик систем, точность расчетов по оптимизации стратегий функционирования и правил их обслуживания, решение проблем, связанных с прогнозированием поведения системы в будущем, и другие вопросы. При формировании исходной информации о параметрах элементов, как правило, за основу берется информация, получаемая в ходе проведения обследования систем и изучения опыта ее эксплуатации. Иными словами за основу берется информация о поведении комплектующих элементов системы в процессе ее функционирования.

Анализ исходных показателей элементов, узлов, составных частей, который производят на этапах эксплуатации, испытаний, конструкторских разработок, выполняется в целях разрешения следующих вопросов:

1) определения фактических значений исследуемых характеристик комплектующих элементов в условиях их реальной эксплуатации;

2) выявления взаимосвязи изучаемых характеристик элементов и условий их эксплуатации, анализа влияния на исследуемые показатели внешних воздействий;

3) прогнозирования поведения вновь создаваемого оборудования.

Таким образом, для решения указанных задач, в первую очередь, необходимо организовать контроль за поведением оборудования в реальных условиях его эксплуатации. В дальнейшем информация, получаемая в процессе эксплуатации объектов, используется для построения моделей систем, в отношении которых проводится анализ.

При проведении экспериментальных исследований большую роль играет информация, полученная в результате наблюдений за объектами, поведение которых имеет вероятностную природу. Изучение таких систем осуществляется по результатам реализации выходных параметров, являющихся случайными величинами. Наиболее общей характеристикой, описывающей поведение одномерной случайной величины, является ее плотность распределения $f(t)$. Зная плотность распределения случайной величины, можно однозначно определить такие характеристики, как вероятность реализации некоторого события, интенсивность наступления события, среднее время между реализациями событий и пр. Приведем формулы, позволяющие оценить соответствующие показатели.

Вероятность реализации события за время t определяется по формуле

$$Q(t) = F(t) = \int_0^t f(t)dt.$$

На практике часто находит применение величина, определяемая через функцию распределения следующим образом:

$$P(t) = 1 - F(t).$$

Например, в теории надежности так определяется вероятность безотказной работы.

Среднее время между реализациями событий определяется из соотношения

$$T_{\text{ñö}} = \int_0^\infty tf(t)dt = \int_0^\infty P(t)dt.$$

Интенсивность наступления события можно определить по формуле

$$\lambda(t) = \frac{f(t)}{P(t)} = \frac{dF(t)}{dt}\frac{1}{P(t)} = -\frac{dP(t)}{dt}\frac{1}{P(t)}.$$

Таким образом, зная плотность или функцию распределения случайной величины, можно перейти к определению характеристик сложной системы. На практике функция распределения бывает неизвестна. Ее приходится восстанавливать по статистическим данным реализации случайной величины. Поскольку статистика о результатах наблюдений всегда присутствует в ограниченном виде, восстановление функции распределения возможно с некоторой долей достоверности. Следовательно, если функция распределения оценена с определенной ошибкой, то и вычисление характеристик системы будет также осуществляться с ошибкой.

Точность оценивания показателей сложных систем характеризуется величиной дисперсии. Пусть необходимо произвести оценивание некоторого показателя $R(t)$. Покажем, как определяется дисперсия в его оценке. Будем считать, что показатель $R(t)$ определяется через функцию распределения. Пусть функция распределения зависит от двух параметров α и β. Примерами двухпараметрических функций являются нормальное распределение, усеченное нормальное, логарифмически нормальное, гамма-распределение, распределение Вейбулла и ряд других. Итак, пусть $F(t) = F(t, \alpha, \beta)$. Соответственно оцениваемый показатель сложной системы можно представить как функционал от $F(t) = F(t, \alpha, \beta)$:

$$R(t) = R[F(t,\alpha,\beta)] = R(t,\alpha,\beta).$$

Разложим оценку $\hat{R}(t)$ в ряд Тейлора в точке α, β и ограничимся тремя членами:

$$\hat{R}(t) = R(t) + \frac{\partial R(t)}{\partial \alpha}(\hat{\alpha} - \alpha) + \frac{\partial R(t)}{\partial \beta}(\hat{\beta} - \beta).$$

К обеим частям данного выражения применим операцию вычисления дисперсии

$$D[\hat{R}(t)] \cong \left[\frac{\partial R(t)}{\partial \alpha}\right]^2 D[\hat{\alpha}] + \left[\frac{\partial R(t)}{\partial \beta}\right]^2 D[\hat{\beta}] + 2\frac{\partial R(t)}{\partial \alpha}\frac{\partial R(t)}{\partial \beta}\operatorname{cov}(\hat{\alpha},\hat{\beta}),$$

где $\operatorname{cov}(\hat{\alpha}, \hat{\beta})$ – ковариация между параметрами $\hat{\alpha}$ и $\hat{\beta}$. Таким образом, для оценки дисперсии некоторого показателя необходимо определить частные производные данного показателя по параметрам закона распределения и дисперсии в оценке параметров закона распределения.

Рассмотрим вопросы определения частных производных для показателей, введенных выше для конкретных законов распределения. Определение дисперсии оценок параметров законов распределения будет описано далее.

В качестве примера рассмотрим определение частных производных оцениваемого показателя по параметрам закона распределения для нормального закона.

Нормальное распределение

Плотность нормального закона распределения имеет вид

$$f_N(t,m,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

Соответственно функция распределения записывается следующим образом:

$$F_N(t,m,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{t}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx.$$

Определим один из показателей надежности – вероятность безотказной работы:

$$P_N(t,m,\sigma) = 1 - \frac{1}{\sqrt{2\pi}\sigma}\int_0^t\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx.$$

Вычислим частные производные:

$$\frac{\partial P_N(t,m,\sigma)}{\partial m} = \frac{1}{\sqrt{2\pi}\sigma}\left[\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) - \exp\left(-\frac{m^2}{2\sigma^2}\right)\right];$$

$$\frac{\partial P_N(t,m,\sigma)}{\partial \sigma^2} = \frac{1}{2\sqrt{2\pi}\sigma^3}\left[(t-m)\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) - m\exp\left(-\frac{m^2}{2\sigma^2}\right)\right].$$

Среднее время между реализациями событий определяется по формуле

$$T_N(m,\sigma) = \int_0^\infty t\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right)dt.$$

Соответственно частные производные определяются как

$$\frac{\partial T_N(m,\sigma)}{\partial m} = \frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right)dt = \Phi\left(\frac{m}{\sigma}\right);$$

$$\frac{\partial T_N(m,\sigma)}{\partial \sigma^2} = \frac{1}{2\sqrt{2\pi}\sigma}\exp\left(-\frac{m^2}{2\sigma^2}\right).$$

И, наконец, для интенсивности наступления события имеем

$$\lambda(t,m,\sigma)=\frac{\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left(-\dfrac{(t-m)^2}{2\sigma^2}\right)}{1-\dfrac{1}{\sqrt{2\pi}\sigma}\displaystyle\int_0^t\exp\left(-\dfrac{(x-m)^2}{2\sigma^2}\right)dx}.$$

Выражения для частных производных имеют вид

$$\frac{\partial\lambda_N(t,m,\sigma)}{\partial m}=\frac{f_N(t,m,\sigma)'_m(1-F_N(t,m,\sigma))-f_N(t,m,\sigma)[1-F_N(t,m,\sigma)]'_m}{[1-F_N(t,m,\sigma)]^2};$$

$$f_N(t,m,\sigma)'_m=\frac{t-m}{\sqrt{2\pi}\sigma^3}\exp\left\{-\frac{(t-m)^2}{2\sigma^2}\right\};$$

$$[1-F_N(t,m,\sigma)]'_m=\frac{1}{\sqrt{2\pi}\sigma}\left[\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right)-\exp\left(-\frac{m^2}{2\sigma^2}\right)\right];$$

$$\frac{\partial\lambda_N(t,m,\sigma)}{\partial\sigma^2}=\frac{f_N(t,m,\sigma)'_{\sigma^2}(1-F_N(t,m,\sigma))-f_N(t,m,\sigma)[1-F_N(t,m,\sigma)]'_{\sigma^2}}{[1-F_N(t,m,\sigma)]^2};$$

$$f_N(t,m,\sigma)'_{\sigma^2}=\frac{[(t-m)^2-\sigma^2]}{2\sqrt{2\pi}\sigma^3}\exp\left\{-\frac{(t-m)^2}{2\sigma^2}\right\};$$

$$[1-F_N(t,m,\sigma)]'_{\sigma^2}=\frac{1}{2\sqrt{2\pi}\sigma^3}\left[(t-m)\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right)-m\exp\left(-\frac{m^2}{2\sigma^2}\right)\right].$$

Таким образом, представлены формулы для определения соответствующих производных показателей по параметрам закона распределения для нормального закона. Обобщением нормального закона распределения является усеченное нормальное распределение. Рассмотрим применение одностороннего усеченного нормального распределения в задачах оценивания показателей сложных систем. В ряде задач системного анализа случайные параметры положительно определены. Примером могут служить задачи теории надежности, в которых случайные параметры имеют область определения от 0 до ∞, например, наработка до отказа – величина положительно определенная. В этом случае нормальный закон распределения применять для описания данных случайных величин неправомерно. В таких ситуациях применяют усеченное слева нормальное распределение. Рассмотрим данный случай применительно к оцениванию показателей надежности.

Одностороннее усеченное нормальное распределение

Плотность распределения усеченного нормального закона с односторонним усечением слева в точке 0 имеет вид

$$f_{у.н}(\mu,b,t)=\begin{cases}\dfrac{c}{\sqrt{2\pi b}}\exp\left(-\dfrac{(x-\mu)^2}{2b}\right), & x>0;\\ 0, & x\le 0,\end{cases}$$

где

$$c=\frac{1}{\displaystyle\int_0^\infty\frac{1}{\sqrt{2\pi b}}\exp\left(-\frac{(x-\mu)^2}{2b}\right)dx}.$$

Соответственно функция распределения запишется как

$$F_{у.н}(t,\mu,b)=\frac{c}{\sqrt{2\pi b}}\int_0^t\exp\left(-\frac{(x-\mu)^2}{2b}\right)dx.$$

Перейдем к определению вероятностных показателей. Вероятность безотказной работы вычисляется по формуле

$$P_{у.н}(t,\mu,b)=1-\frac{\displaystyle\int_0^t\exp\left(-\frac{(x-\mu)^2}{2b}\right)dx}{\displaystyle\int_0^\infty\exp\left(-\frac{(x-\mu)^2}{2b}\right)dx}.$$

Введем обозначения:

$$R=\int_0^t\exp\left(-\frac{(x-\mu)^2}{2b}\right)dx;\quad Q=\int_0^\infty\exp\left(-\frac{(x-\mu)^2}{2b}\right)dx.$$

Соответствующие производные имеют вид

$$\frac{\partial P_{у.н}(t,\mu,b)}{\partial\mu}=-\frac{R'_\mu Q-Q'_\mu R}{Q^2};$$

$$R'_\mu=\exp\left(-\frac{\mu^2}{2b}\right)-\exp\left(-\frac{(t-\mu)^2}{2b}\right);\quad Q'_\mu=\exp\left(-\frac{\mu^2}{2b}\right);$$

$$R'_b=\frac{\mu-t}{2b}\exp\left(-\frac{(t-\mu)^2}{2b}\right)-\frac{\mu}{2b}\exp\left(-\frac{\mu^2}{2b}\right)+\frac{\sqrt{\pi}}{\sqrt{2b}}\left[\Phi\left(\frac{t-\mu}{\sqrt{b}}\right)+\Phi\left(\frac{\mu}{\sqrt{b}}\right)\right];$$

$$Q'_b = -\frac{\mu}{2b}\exp\left(-\frac{\mu^2}{2b}\right) + \frac{\sqrt{\pi}}{\sqrt{2b}}\left[\Phi\left(\frac{\mu}{\sqrt{b}}\right) + 0,5\right].$$

Среднее время между реализациями событий определяется по формуле

$$T_{у.н}(\mu, b) = \frac{b\exp\left(-\frac{\mu^2}{2b}\right) + \mu\int_0^\infty \exp\left(-\frac{(x-\mu)^2}{2b}\right)dx}{Q}.$$

Обозначим числитель через L.

Соответствующие производные вычисляются по формулам

$$\frac{\partial T_{у.н}(\mu,b)}{\partial\mu} = -\frac{L'_\mu Q - Q'_\mu L}{Q^2};\quad L'_\mu = b\sqrt{\pi}\left[\Phi\left(\frac{\mu}{\sqrt{b}}\right) + 0,5\right];$$

$$\frac{\partial T_{у.н}(\mu,b)}{\partial b} = -\frac{L'_b Q - Q'_b L}{Q^2};\quad L'_b = \exp\left(-\frac{\mu^2}{2b}\right) - \mu\frac{\sqrt{\pi}}{\sqrt{2b}}\left[\Phi\left(\frac{\mu}{\sqrt{b}}\right) + 0,5\right].$$

Наконец, интенсивность наступления событий равна

$$\lambda_{у.н}(t,\mu,b) = \frac{\exp\left(-\frac{(t-\mu)^2}{2b}\right)}{\int_0^\infty \exp\left(-\frac{(x-\mu)^2}{2b}\right)dx - \int_0^t \exp\left(-\frac{(x-\mu)^2}{2b}\right)dx}.$$

Введем обозначение

$$M = \exp\left(-\frac{(t-\mu)^2}{2b}\right).$$

Определим производные интенсивности по параметрам

$$\frac{\partial\lambda_{у.н}(t,\mu,b)}{\partial\mu} = \frac{M'_\mu(Q-R) - (Q-R)'_\mu M}{(Q-R)^2};$$

$$M'_\mu = \frac{t-\mu}{b}\exp\left(-\frac{(t-\mu)^2}{2b}\right);\quad (Q-R)'_\mu = \exp\left(-\frac{(t-\mu)^2}{2b}\right)$$

и последнее выражение

$$\frac{\partial\lambda_{у.н}(t,\mu,b)}{\partial b} = \frac{M'_b(Q-R) - (Q-R)'_b M}{(Q-R)^2},$$

где соответствующие составляющие определяются по формулам

$$M'_b = \frac{(t-\mu)^2}{2b^2}\exp\left(-\frac{(t-\mu)^2}{2b}\right);$$

$$(Q-R)'_b = \frac{t-\mu}{2b}\exp\left(-\frac{(t-\mu)^2}{2b}\right) + \frac{\sqrt{\pi}}{2\sqrt{2b}} - \frac{\sqrt{\pi}}{2\sqrt{2b}}\Phi\left(\frac{t-\mu}{\sqrt{b}}\right).$$

Логарифмически-нормальное распределение

Логарифмически-нормальному закону распределения подчиняется случайная величина t, логарифм которой распределен по нормальному закону. Плотность распределения логарифмически-нормального закона имеет вид

$$f_{л.н}(t) = \frac{1}{tb\sqrt{2\pi}}\exp\left(-\frac{(\ln t-\mu)^2}{2b^2}\right).$$

Функция распределения имеет вид

$$F_{л.н}(t) = \int_{-\infty}^{\ln t}\frac{1}{\sqrt{2\pi B}}\exp\left(-\frac{(x-\mu)^2}{2B}\right)dx,$$

где $B = b^2$.

Запишем формулы для определения показателей надежности

$$P_{л.н}(t,\mu,B) = 1 - \int_{-\infty}^{\ln t}\exp\left(-\frac{(x-\mu)^2}{2B}\right)dx.$$

Соответствующие производные имеют вид

$$\frac{P_{л.н.}(t,\mu,B)}{\partial\mu} = \frac{1}{\sqrt{2\pi B}}\exp\left(-\frac{(\ln t-\mu)^2}{2B}\right);$$

$$\frac{P_{л.н}(t,\mu,B)}{\partial B} = \frac{\ln t-\mu}{2B\sqrt{2\pi B}}\exp\left(-\frac{(\ln t-\mu)^2}{2B}\right).$$

Для определения средней наработки до отказа используют формулу

$$T_{л.н}(t,\mu,B) = \exp\left(\mu + \frac{B}{2}\right).$$

Производные равны

$$\frac{\partial T_{\text{л.н}}(t,\mu,B)}{\partial\mu}=\exp\left(\mu+\frac{B}{2}\right);\quad \frac{\partial T_{\text{л.н}}(t,\mu,B)}{\partial B}=\frac{1}{2}\exp\left(\mu+\frac{B}{2}\right).$$

Выражение для определения интенсивности отказов имеет вид

$$\lambda_{\text{л.н}}(t,\mu,B)=\frac{\frac{1}{t\sqrt{2\pi B}}\exp\left(-\frac{(\ln t-\mu)}{2B}\right)}{1-\frac{1}{\sqrt{2\pi B}}\int\limits_{-\infty}^{\ln t}\exp\left(-\frac{(x-\mu)^2}{2B}\right)dx}=\frac{f_{\text{л.н}}(t,\mu,B)}{1-F_{\text{л.н}}(t,\mu,B)}.$$

Частные производные определяются из выражений

$$\frac{\partial\lambda_{\text{л.н}}(t,\mu,B)}{\partial\mu}=\frac{\left(f_{\text{л.н}}(t)\right)'_{\mu}(1-F_{\text{л.н}}(t))-f_{\text{л.н}}\left(1-F_{\text{л.н}}(t)\right)'_{\mu}}{[1-F_{\text{л.н}}(t)]^2},$$

где $\left(f_{\text{л.н}}(t)\right)'_{\mu}=\frac{1}{t\sqrt{2\pi B}}\frac{\ln t-\mu}{B}\exp\left(-\frac{(\ln t-\mu)^2}{2B}\right);$

$$\left(1-F_{\text{л.н}}\right)'_{\mu}=\frac{1}{\sqrt{2\pi B}}\exp\left(-\frac{(\ln t-\mu)^2}{2B}\right);$$

$$\frac{\partial\lambda_{\text{л.н}}(t,\mu,B)}{\partial B}=\frac{\left(f_{\text{л.н}}(t)\right)'_{B}(1-F_{\text{л.н}}(t))-f_{\text{л.н}}\left(1-F_{\text{л.н}}(t)\right)'_{B}}{[1-F_{\text{л.н}}(t)]^2};$$

$$\left(f_{\text{л.н}}(t)\right)'_{B}=\frac{1}{2B^2t\sqrt{2\pi B}}\left((\ln t-\mu)^2-B\right)\exp\left(-\frac{(\ln t-\mu)^2}{2B}\right);$$

$$\left(1-F_{\text{л.н}}\right)'_{B}=\frac{\ln t-\mu}{2B\sqrt{2\pi B}}\exp\left(-\frac{(\ln t-\mu)^2}{2B}\right).$$

Распределение Вейбулла

Плотность распределения Вейбулла имеет вид

$$f_{\text{В}}(t,a,b)=\frac{b}{a}\left(\frac{t}{a}\right)^{b-1}\exp\left(-\left(\frac{t}{a}\right)^{b}\right),$$

функция распределения

$$F_{\text{В}}(t,a,b)=1-\exp\left(-\left(\frac{t}{a}\right)^{b}\right),$$

где a – параметр масштаба; b – параметр формы распределения Вейбулла.

Запишем выражение для вероятности безотказной работы

$$P_{\text{В}}(t,a,b)=\exp\left(-\left(\frac{t}{a}\right)^{b}\right).$$

Вычислим производные данного выражения по параметрам распределения:

$$\frac{\partial P_{\text{В}}(t,a,b)}{\partial a}=\frac{b}{a}\left(\frac{t}{a}\right)^{b}\ln\left(\frac{t}{a}\right)\exp\left(-\left(\frac{t}{a}\right)^{b}\right);$$

$$\frac{\partial P_{\text{В}}(t,a,b)}{\partial b}=-\left(\frac{t}{a}\right)^{b}\ln\left(\frac{t}{a}\right)\exp\left(-\left(\frac{t}{a}\right)^{b}\right).$$

Средняя наработка до отказа определяется по формуле

$$T_{\text{В}}(a,b)=\int\limits_{0}^{\infty}\exp\left(-\left(\frac{t}{a}\right)^{b}\right)dt.$$

Соответствующие производные равны

$$\frac{\partial T_{\text{В}}(a,b)}{\partial a}=\int\limits_{0}^{\infty}\frac{b}{a}\left(\frac{t}{a}\right)^{b}\exp\left(-\left(\frac{t}{a}\right)^{b}\right)dt;\quad \frac{\partial T_{\text{В}}(a,b)}{\partial b}=-\int\limits_{0}^{\infty}\left(\frac{t}{a}\right)^{b}\ln\left(\frac{t}{a}\right)\exp\left(-\left(\frac{t}{a}\right)^{b}\right)dt.$$

Интенсивность отказа равна

$$\lambda_{\text{В}}(t,a,b)=\frac{b}{a}\left(\frac{t}{a}\right)^{b-1}.$$

Производные по параметрам имеют вид

$$\frac{\partial\lambda_{\text{В}}(a,b)}{\partial a}=-\frac{b^2}{a^2}\left(\frac{t}{a}\right)^{b-1},\quad \frac{\partial\lambda_{\text{В}}(a,b)}{\partial b}=\frac{t^{b-1}}{a^{b}}+\frac{b}{a}\left(\frac{t}{a}\right)^{b-1}\ln\left(\frac{t}{a}\right).$$

Гамма-распределение

Плотность гамма-распределения записывается следующим образом

$$f_{\Gamma}(t,\lambda,\alpha)=\frac{\lambda^{\alpha}t^{\alpha-1}\exp(-\lambda t)}{\Gamma(\alpha)},$$

где $\Gamma(\alpha)$ – гамма-функция.

Соответственно функция распределения имеет вид

$$F_\Gamma(t,\lambda,\alpha)=\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^t x^{\alpha-1}\exp(-\lambda x)dx.$$

Вероятность безотказной работы вычисляется по формуле

$$P_\Gamma(t,\lambda,\alpha)=1-\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^t x^{\alpha-1}\exp(-\lambda x)dx.$$

Производные по параметрам равны

$$\frac{\partial P_\Gamma(t,\lambda,\alpha)}{\partial\lambda}=-\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}\int_0^t x^{\alpha-1}\exp(-\lambda x)(\alpha-\lambda x)dx;$$

$$\frac{\partial P_\Gamma(t,\lambda,\alpha)}{\partial\alpha}=-\frac{\lambda^\alpha}{\Gamma^2(\alpha)}\int_0^t x^{\alpha-1}\exp(-\lambda x)[\Gamma(\alpha)(\ln\lambda-\ln t)-\Gamma'(\alpha)]dx,$$

где $\Gamma(\alpha)=\int_0^\infty \lambda^\alpha t^{\alpha-1}\exp(-\lambda t)dt=\int_0^\infty z^{\alpha-1}\exp(-z)dz$; $\Gamma'(\alpha)=\int_0^\infty z^{\alpha-1}\exp(-z)\ln z dz$.

Средняя наработка до отказа определяется по формуле

$$T_\Gamma(\alpha,\lambda)=\int_0^\infty \frac{\lambda^\alpha t^\alpha}{\Gamma(\alpha)}\exp(-\lambda t)dt=\frac{\alpha}{\lambda}.$$

Соответствующие производные равны

$$\frac{\partial T_\Gamma(\alpha,\lambda)}{\partial\lambda}=-\frac{\alpha}{\lambda^2};\quad \frac{\partial T_\Gamma(\alpha,\lambda)}{\partial\alpha}=\frac{1}{\lambda}.$$

Интенсивность отказов записывается

$$\lambda_\Gamma(t,\alpha,\lambda)=\frac{\lambda^\alpha t^{\alpha-1}\exp(-\lambda t)}{\left[1-\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^t x^{\alpha-1}\exp(-\lambda x)dx\right]\Gamma(\alpha)}.$$

Производные по параметрам определяются в виде

$$\frac{\partial\lambda_\Gamma(t,\alpha,\lambda)}{\partial\lambda}=\frac{(f_\Gamma(t,\lambda,\alpha))'_\lambda[1-F_\Gamma(t,\lambda,\alpha)]-f_\Gamma(t,\lambda,\alpha)[1-F_\Gamma(t,\lambda,\alpha)]'_\lambda}{[1-F_\Gamma(t,\lambda,\alpha)]^2},$$

где $(f_\Gamma(t,\lambda,\alpha))'_\lambda=\dfrac{\alpha\lambda^{\alpha-1}t^{\alpha-1}\exp(-\lambda t)-\lambda^\alpha t^\alpha\exp(-\lambda t)}{\Gamma(\alpha)}$;

$$[1-F_\Gamma(t,\lambda,\alpha)]'_\lambda=\frac{\alpha\lambda^{\alpha-1}\int_0^t x^{\alpha-1}\exp(-\lambda x)dx-\lambda^\alpha\int_0^t x^\alpha\exp(-\lambda x)dx}{\Gamma(\alpha)};$$

$$\frac{\partial\lambda_\Gamma(t,\alpha,\lambda)}{\partial\alpha}=\frac{(f_\Gamma(t,\lambda,\alpha))'_\alpha[1-F_\Gamma(t,\lambda,\alpha)]-f_\Gamma(t,\lambda,\alpha)[1-F_\Gamma(t,\lambda,\alpha)]'_\alpha}{[1-F_\Gamma(t,\lambda,\alpha)]^2},$$

где

$$(f_\Gamma(t,\lambda,\alpha))'_\alpha=\frac{1}{(\Gamma(\alpha))^2}\Big[\left(\lambda^\alpha\ln\lambda t^{\alpha-1}\exp(-\lambda t)+\lambda^\alpha t^{\alpha-1}\ln t\exp(-\lambda t)\right)-$$
$$-\lambda^\alpha t^{\alpha-1}\exp(-\lambda t)\Gamma'_\alpha(\alpha)\Big];$$

$$[1-F_\Gamma(t,\lambda,\alpha)]'_\alpha=\frac{1}{(\Gamma(\alpha))^2}\Bigg[\Gamma_\alpha(\alpha)\lambda^\alpha\int_0^t x^{\alpha-1}\exp(-\lambda x)dx-$$
$$-\Gamma(\alpha)\left(\lambda^\alpha\ln\lambda\int_0^t x^{\alpha-1}\exp(-\lambda x)dx+\lambda^\alpha\int_0^t x^{\alpha-1}\ln x\exp(-\lambda x)dx\right)\Bigg].$$

Таким образом, получены выражения, позволяющие решать вопросы оценки точности в определении показателей сложных систем. Рассмотрены наиболее часто используемые в системном анализе законы распределения. Получены формулы для определения основных показателей систем и вычислены первые частные производные показателей по параметрам соответствующих законов распределения. Следующим вопросом, который требует решения, является вопрос оценивания параметров выбранного закона распределения. Рассмотрим, как решается данная задача.

7.2. Использование метода максимального правдоподобия для оценивания параметров законов распределения

Метод максимального правдоподобия широко используется при оценивании параметров сложных систем. Этот метод служит основой процедур проверки статистических гипотез и доверительного интервального оценивания. Оценки характеристик, получаемые методом максимального правдоподобия, обладают рядом важных свойств, таких как несмещенность, асимптотическая эффективность, состоятельность [35].

Рассмотрим применение метода максимального правдоподобия для решения задач оценивания параметров законов распределения. Вид закона распределения исследуемой случайной величины будем предполагать известным.

Определим функцию правдоподобия параметра θ как неотрицательную вещественную функцию $L(\theta, t)$, заданную на множестве $\Theta \times T$, пропорциональную функции плотности распределения:

$$L(\theta,t)=\prod_{i=1}^{n} f(\theta,T_i),\ \ i=\overline{1,n}, \qquad (7.1)$$

где Θ – область определения вектора параметров θ; T – область определения наблюдаемой случайной величины t, по результатам наблюдения за которой про-изводится оценивание параметров θ; T_i – реализация случайной величины t; θ – в общем случае вектор параметров закона распределения.

Оценкой максимального правдоподобия для заданной функции правдоподобия $L(\theta, t)$ является функция $\hat{\theta}(T)$, удовлетворяющая соотношению [36]:

$$L\{\hat{\theta}(T),T\}=\sup_{\theta\in\Theta} L(\theta,T).$$

Для нахождения оценки θ решают уравнение

$$\frac{\partial L(\theta,T)}{\partial\theta}=0. \qquad (7.2)$$

Поскольку $\ln L(\theta, t)$ при фиксированных $T_1, T_2, \ldots, T_n$ достигает максимума при том же значении θ, что и $L(\theta, t)$, значения θ можно определять, решая уравнение

$$\frac{\partial \ln L(\theta,T)}{\partial\theta}=0.$$

Для величины $\ln L(\theta, t)$ используют обозначение $l(\theta, t)$.

Одним из важных достоинств метода максимального правдоподобия является то, что он позволяет получить асимптотически нормальные и эффективные оценки параметров функции распределения случайной величины t. Для этого необходимо, чтобы выполнялся ряд условий, называемых условиями регулярности [32].

Для каждого θ, принадлежащего некоторому невырожденному интервалу Θ, существуют производные

$$\frac{\partial \ln f(\theta,t)}{\partial\theta};\ \ \frac{\partial^2 \ln f(\theta,t)}{\partial\theta^2};\ \ \frac{\partial^3 \ln f(\theta,t)}{\partial\theta^3};$$

для каждого θ из Θ имеем

$$\left|\frac{\partial f(\theta,t)}{\partial\theta}\right|<F_1(t);\ \left|\frac{\partial^2 f(\theta,t)}{\partial\theta^2}\right|<F_2(t);\ \left|\frac{\partial^3 \ln f(\theta,t)}{\partial\theta^3}\right|<H(t),$$

где $F_1(t)$, $F_2(t)$, $H(t)$ – некоторые функции, удовлетворяющие следующим условиям:

- $F_1(t)$, $F_2(t)$ интегрируемы на $(-\infty, \infty)$ и интеграл $\int_{-\infty}^{\infty} H(t)f(\theta,t)dt<\beta$,

причем β не зависит от θ;

- для каждого θ из Θ интеграл

$$\int_{-\infty}^{\infty}\left(\frac{\partial \ln f(\theta,t)}{\partial\theta}\right)^2 f(\theta,t)dt$$

конечен и положителен.

Используя метод максимального правдоподобия при наличии больших объемов выборки, можно произвести оценивание дисперсий вектора оценок θ. Для этого составляется информационная матрица Фишера

$$\mathbf{I}=\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \ldots & \ldots & \ldots & \ldots\\ a_{n1} & a_{n2} & \ldots & a_{nn}\end{pmatrix},$$

где $a_{ij}=-M\left[\frac{\partial^2 \ln L}{\partial\theta_i\partial\theta_j}\right]$, θ – вектор параметров функции $F(\theta, t)$, $i=\overline{1,n}$.

Дисперсии и ковариации оценок $\hat{\theta}_i$ и $\hat{\theta}_j$ определяются из ковариационной матрицы **V**, которая является обратной матрице **I**:

$$\mathbf{V}=\mathbf{I}^{-1}=\begin{pmatrix} D_{11} & D_{12} & \dots & D_{1n} \\ D_{21} & D_{22} & \dots & D_{2n} \\ \dots & \dots & \dots & \dots \\ D_{n1} & D_{n2} & \dots & D_{nn} \end{pmatrix},$$

где $D_{ij}=\begin{cases} D(\hat{\theta}_i) \text{ при } i=j; \\ \operatorname{cov}(\hat{\theta}_i,\hat{\theta}_j) \text{ при } i\neq j. \end{cases}$

Приведем формулы для определения дисперсии параметров законов распределения для двухпараметрических плотностей. Пусть α и β – оцениваемые параметры плотности $f(\alpha, \beta, t)$. Пусть определены элементы информационной матрицы

$$\mathbf{I}=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

где

$$a_{11}=-\frac{\partial^2 l(\alpha,\beta,t)}{\partial\alpha^2};\ a_{21}=a_{12}=-\frac{\partial^2 l(\alpha,\beta,t)}{\partial\alpha\partial\beta};\ a_{22}=-\frac{\partial l(\alpha,\beta,t)}{\partial\beta^2}.$$

Дискриминант матрицы равен

$$\mathrm{Dis}=a_{11}a_{22}-a_{12}^2. \tag{7.3}$$

Тогда дисперсию параметров α и β определим по формулам

$$D_\alpha=-\frac{a_{22}}{\mathrm{Dis}}; \tag{7.4}$$

$$D_\beta=-\frac{a_{11}}{\mathrm{Dis}}. \tag{7.5}$$

Ковариация будет вычисляться следующим образом:

$$\operatorname{cov}(\alpha,\beta)=-\frac{a_{12}}{\mathrm{Dis}}. \tag{7.6}$$

Перейдем к рассмотрению примеров применения метода максимального правдоподобия для оценивания параметров некоторых законов распределения, имеющих важное значение в задачах системного анализа.

Экспоненциальное распределение

Рассмотрим имеющее важное прикладное значение экспоненциальное распределение. Например, данное распределение широко используется в теории надежности для описания случайной величины наработки до отказа. Плотность экспоненциального распределения имеет вид $f_э(\lambda, t)$; функция распределения $F_э(\lambda, t) = 1 - \exp(-\lambda t)$. Параметр λ называется интенсивностью отказов. Запишем функцию правдоподобия

$$L_э(\lambda,t)=\prod_{i=1}^{n}\lambda\exp(-\lambda T_i)=\lambda^n\exp\left(-\lambda\sum_{i=1}^{n}T_i\right).$$

Для определения оценки параметра λ необходимо решить уравнение

$$\frac{\partial l(\lambda,t)}{\partial\lambda}=\frac{\partial\left(n\ln\lambda-\lambda\sum_{i=1}^{n}T_i\right)}{\partial\lambda}=0.$$

После дифференцирования получаем оценку максимального правдоподобия параметра экспоненциального закона распределения

$$\hat{\lambda}=\frac{n}{\sum_{i=1}^{n}T_i}.$$

Дисперсия оценки параметра λ характеризует точность этого параметра и равняется

$$D[\hat{\lambda}]=\frac{\lambda^2}{n}.$$

Для интенсивности отказов, зная оценку параметра и ее дисперсию, можно определить доверительный интервал с заданной доверительной вероятностью. Если обозначить верхнюю и нижнюю оценки интенсивности отказов через $\lambda_в$ и $\lambda_н$, то можно определить

$$\lambda_в=\hat{\lambda}+\frac{\hat{\lambda}}{\sqrt{n}}t_\beta,\ \lambda_н=\hat{\lambda}-\frac{\hat{\lambda}}{\sqrt{n}}t_\beta,$$

где t_β – табулированная величина, которая зависит от уровня доверительной вероятности β и определяется обратным интерполированием распределения Стьюдента [37, табл. 6.2].

Определение среднего времени между реализациями событий производится по формуле

$$T_Э = \frac{1}{\hat{\lambda}};$$

интервальные оценки равны

$$T_Н = \frac{1}{\hat{\lambda}_В}, \ T_В = \frac{1}{\hat{\lambda}_Н}.$$

Вероятность безотказной работы определяется следующим образом:

$$P_Э(t,\lambda) = \exp(-\lambda t),$$

соответственно интервальные оценки вычисляются так:

$$P_Н(t,\lambda) = \exp(-\lambda_Н t);$$

$$P_В(t,\lambda) = \exp(-\lambda_В t).$$

Рассмотрим далее пример оценивания параметров нормального закона распределения.

Нормальное распределение

$$f(t,m,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(m-t)^2}{2\sigma^2}\right).$$

Функция правдоподобия имеет вид

$$L_N(m,\sigma,t) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{i=1}^{n}\frac{(T_i-m)^2}{2\sigma^2}\right),$$

$$l_N(m,\sigma,t) = -\frac{n}{2}\left(\ln 2\pi + \ln\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(T_i-m)^2.$$

Для определения оценок максимального правдоподобия параметров m и σ^2 необходимо решить систему уравнений

$$\frac{\partial l(m,\sigma,t)}{\partial m} = \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(T_i-m) = 0;$$

$$\frac{\partial l(m,\sigma,t)}{\partial\sigma^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(T_i-\hat{m})^2 = 0,$$

откуда получаем

$$\hat{m} = \frac{1}{n}\sum_{i=1}^{n}T_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(T_i-\hat{m})^2.$$

Определим точность в оценивании данных параметров. Для этого вычислим вторые производные функции правдоподобия по параметрам:

$$a_{11} = \frac{\partial^2 l(m,\sigma^2,t)}{\partial m^2} = -\frac{n}{\hat{\sigma}^2};\ a_{12} = \frac{\partial^2 l(m,\sigma^2,t)}{\partial m\,\partial\sigma^2} = \frac{-\sum_{i=1}^{n}(T_i-\hat{m})}{\hat{\sigma}^4} = 0;$$

$$a_{22} = \frac{\partial^2 l(m,\sigma^2,t)}{\partial(\sigma^2)^2} = -\frac{n}{2\hat{\sigma}^4}.$$

Дискриминант информационной матрицы будет равен

$$\mathrm{Dis} = \frac{n^2}{2\hat{\sigma}^6},$$

откуда получаем

$$D[m] = \frac{\hat{\sigma}^2}{n};\ D[\sigma^2] = \frac{2\hat{\sigma}^4}{n}.$$

Таким образом, рассмотрен метод оценки параметров законов распределения и определения точности в их оценке. Рассмотренные примеры определения параметров законов распределения имеют важное прикладное значение. Как было указано, экспоненциальное распределение применяется в теории надежности для описания наработок до отказа объектов. Область применения нормального закона распределения еще более широка. Он используется для описания погрешностей, дрейфов параметров, наработок до отказа механических изделий, для которых не удается выделить доминирующей причины, приводящей к отказу, для описания времени обслуживания систем и т.д.

7.3. Оценка вероятностных показателей систем путем обработки цензурированных данных

Постановка задачи при обработке цензурированных выборок формулируется следующим образом. Пусть имеется выборка объема $r = k+\nu$, которая содержит ряд наблюдений за функционированием объектов с реализовавшимся признаком $T_1, T_2, \ldots, T_k$ (полные наработки), и ряд наблюдений с нереализовавшимся признаком $T_1', T_2', ..., T_\nu'$. Пусть известен закон распределения времени до реализации наблюдаемого признака $F(\theta, t)$. Оценим параметры закона распределения. Функция правдоподобия для выборки, содержащей цензурированные наработки при цензурировании справа, запишется следующим образом:

$$L_{\text{п}}(\theta, t) = \prod_{i=1}^{\nu}(1 - F(\theta, T_i')) \prod_{j=1}^{k} f(\theta, T_j). \qquad (7.7)$$

Для цензурированной выборки при цензурировании слева функция правдоподобия имеет вид:

$$L_{\text{л}}(\theta, t) = \prod_{i=1}^{\nu}(F(\theta, T_i')) \prod_{j=1}^{k} f(\theta, T_j). \qquad (7.8)$$

Процедура получения оценок параметров аналогична изложенной в п. 7.2. А именно, необходимо прологарифмировать функцию правдоподобия, взять от нее производную по искомому параметру и приравнять ее нулю. Например, в случае цензурирования справа решение будет выглядеть следующим образом. Логарифм от функции правдоподобия записывается в виде

$$l(\theta, T) = \sum_{j=1}^{k} \ln f(\theta, T_j) + \sum_{i=1}^{\nu} \ln(1 - F(\theta, T_i')).$$

Возьмем производную от данного выражения по искомому параметру

$$\frac{\partial l(\theta, T)}{\partial \theta} = \sum_{j=1}^{k} \frac{\partial \ln f(\theta, T_j)}{\partial \theta} + \sum_{i=1}^{\nu} \frac{\partial \ln(1 - F(\theta, T_i'))}{\partial \theta}$$

и приравняем полученное выражение нулю. Если θ – вектор порядка k, то необходимо взять k частных производных для получения системы уравнений

$$\frac{\partial l(\theta, T)}{\partial \theta_i} = 0, \quad i = \overline{1, k},$$

при решении которой находим эффективные несмещенные оценки параметров закона распределения $F(\theta, t)$.

Рассмотрим примеры оценивания параметров законов распределения.

Экспоненциальное распределение

Запишем функцию распределения $F(\lambda, t) = 1 - \exp(-\lambda t)$ и плотность распределения $f(\lambda, t) = \lambda \exp(-\lambda t)$ величины t. Пусть требуется определить оценку интенсивности отказа с учетом полных и цензурированных справа наработок. Функция правдоподобия для данного случая представления информации имеет вид

$$l(\lambda, t) = \sum_{i=1}^{k} \ln(\lambda \exp(-\lambda T_i)) + \sum_{j=1}^{\nu} \ln(\exp(-\lambda T_j')).$$

Возьмем производную от функции правдоподобия по параметру λ и приравняем ее нулю:

$$\frac{k}{\lambda} - \sum_{i=1}^{k} T_i - \sum_{j=1}^{\nu} T_j' = 0,$$

откуда получаем оценку параметра

$$\hat{\lambda} = \frac{k}{\sum_{i=1}^{k} T_i + \sum_{j=1}^{\nu} T_j'}.$$

Точность определения оценки вычисляется по формуле

$$D(\lambda) = \frac{\hat{\lambda}^2}{k}.$$

Нормальное распределение

Логарифмическая функция правдоподобия для цензурированной справа выборки имеет вид

$$l(t, m, \sigma) = -\frac{k}{2}(\ln 2\pi + \ln \sigma^2) - \sum_{i=1}^{k} \frac{(T_i - m)^2}{2\sigma^2} +$$

$$+ \sum_{j=1}^{\nu} \ln\left[1 - \frac{1}{\sqrt{2\pi}\sigma} \int_0^{T_j'} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx\right].$$

Параметры закона распределения определяются из системы уравнений

$$\frac{\partial l(t, m, \sigma)}{\partial m} = \frac{1}{\sigma^2} \sum_{i=1}^{k} (T_i - m) - \frac{1}{\sqrt{2\pi}\sigma} \sum_{j=1}^{\nu} \frac{\exp\left(-\frac{m^2}{2\sigma^2}\right) - \exp\left(-\frac{(T_j' - m)^2}{2\sigma^2}\right)}{1 - \frac{1}{\sqrt{2\pi}\sigma} \int_0^{T_j'} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx} = 0,$$

$$\frac{\partial l(t,m,\sigma)}{\partial\sigma^2}=-\frac{k}{2\sigma^2}+\frac{\sum_{i=1}^{k}(T_i-m)^2}{2\sigma^4}+$$

$$+\frac{1}{2\sigma^2\sqrt{2\pi}\sigma}\sum_{j=1}^{\nu}\frac{(T_j-m)\exp\left(-\frac{(T_j'-m)^2}{2\sigma^2}\right)+m\exp\left(-\frac{m^2}{2\sigma^2}\right)}{1-\frac{1}{\sqrt{2\pi}\sigma}\int_0^{T_j'}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx}=0.$$

Данная система уравнений является трансцендентной, решается численными методами. В качестве первого приближения при решении системы можно взять следующие оценки:

$$\hat m_1=\frac{\sum_{i=1}^{k}T_i}{k},\quad \hat\sigma_1^2=\frac{\sum_{i=1}^{k}(T_i-m_1)^2}{k}.$$

Элементы информационной матрицы определяются в виде

$$a_{11}=\frac{\partial^2 l(t,m,\sigma)}{\partial m^2}=-\frac{k}{\sigma^2}+\frac{1}{\sqrt{2\pi}\sigma}\sum_{j=1}^{\nu}\frac{B'C-C'B}{C^2};$$

$$B=\exp\left(-\frac{(T_j'-m)^2}{2\sigma^2}\right)-\exp\left(-\frac{m^2}{2\sigma^2}\right);$$

$$B'=\frac{T_j'-m}{\sigma^2}\exp\left(-\frac{(T_j'-m)^2}{2\sigma^2}\right)+\frac{m}{\sigma^2}\exp\left(-\frac{m^2}{2\sigma^2}\right);$$

$$C=1-\frac{1}{\sqrt{2\pi}\sigma}\int_0^{T_j'}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx;$$

$$C'=\frac{1}{\sqrt{2\pi}\sigma}\left[\exp\left(-\frac{(T_j'-m)^2}{2\sigma^2}\right)-\exp\left(-\frac{m^2}{2\sigma^2}\right)\right]=\frac{1}{\sqrt{2\pi}\sigma}B;$$

$$a_{22}=\frac{\partial^2 l(t,m,\sigma)}{\partial(\sigma^2)^2}=\frac{k}{2\sigma^4}-\frac{\sum_{i=1}^{k}(T_i-m)^2}{\sigma^6}-\frac{3}{4\sqrt{2\pi}\sigma^3}\sum_{j=1}^{\nu}E_j+\frac{1}{2\sqrt{2\pi}\sigma^3}\sum_{j=1}^{\nu}E_{jb}';$$

$$E_j=\frac{(T_j'-m)\exp\left(-\frac{(T_j'-m)^2}{2\sigma^2}\right)+m\exp\left(-\frac{m^2}{2\sigma^2}\right)}{1-\frac{1}{\sqrt{2\pi}\sigma}\int_0^{T_j'}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx}=\frac{D}{F};$$

$$E_{jb}'=\frac{D_b'F-F_b'D}{F^2};\ D_b'=\frac{(T_j'-m)^3}{2\sigma^4}\exp\left(-\frac{(T_j'-m)^2}{2\sigma^2}\right)+\frac{m^3}{2\sigma^4}\exp\left(-\frac{m^2}{2\sigma^2}\right);$$

$$F_b'=\frac{T_j'-m}{2\sqrt{2\pi}\sigma^3}\exp\left(-\frac{(T_j'-m)^2}{2\sigma^2}\right)+\frac{m^3}{2\sqrt{2\pi}\sigma^3}\exp\left(-\frac{m^3}{2\sigma^2}\right).$$

$$a_{12}=\frac{\partial^2 l(t,m,\sigma)}{\partial m\,\partial\sigma^2}=0.$$

В этом случае ковариация между параметрами a и σ^2 равна нулю, а дисперсии параметров определяются из выражений

$$D[m]=-\frac{1}{a_{11}};\ D[\sigma^2]=-\frac{1}{a_{22}}.$$

Аналогичные действия выполняются в случае других законов распределения случайной величины, отражающей реализацию наблюдаемого признака. Приведенные примеры иллюстрируют тот факт, что учет цензурированной информации приводит к существенному усложнению процедуры вычисления параметров рассматриваемого закона распределения. Но следует отметить, что сложность решения задачи оценивания компенсируется точностью оценок, которую в итоге удается достичь за счет учета цензурированной информации.

7.4. Оценивание показателей систем по группированным данным

В ходе проведения наблюдений за функционированием систем в ряде случаев исследователь не имеет возможности получать информацию о реализовавшихся значениях наблюдаемой случайной величины. Известными бывают лишь интервалы значений, в которые попал тот или иной результат наблюдения. Наиболее просто и ясно данную ситуацию иллюстрирует пример с организацией и проведением исследований по определению характеристик надежности элементов и систем.

Так при решении задачи анализа надежности реально функционирующих объектов осуществляется сбор информации о поведении объектов в процессе их эксплуатации, в частности фиксируются отказы и соответственно наработки объектов до отказа. Однако в большинстве случаев доступной является лишь информация о том, что отказы произошли в некотором интервале времени. Это связано с тем, что отказы устройств фиксируются не мгновенно, а в некоторые, наперед спланированные моменты контроля исправности функционирования оборудования или даже в моменты проведения плановых профилактических работ. Практически мгновенно отказы выявляются у незначительной группы устройств, имеющих встроенный контроль. В остальных случаях у исследователя имеется информация о том, что отказ произошел в интервале времени между предыдущим и последующим контролем либо в интервале между очередными профилактиками.

Итак, пусть в системе спланированы моменты контроля исправности функ-ционирования оборудования $\xi_1, \xi_2, \dots, \xi_k$, где $0<\xi_1<\xi_2<\dots<\xi_k<<\infty$. В моменты контроля выявляется количество отказавших в интервале времени $[\xi_{i-1}, \xi_i]$ устройств, т.е. наблюдаемыми случайными величинами являются целые неотрицательные числа, характеризующие количество отказавших объектов на рассматриваемом интервале N_1, N_2, ..., N_{k+1} где N_i – число устройств, отказавших в интервале $[\xi_{i-1}, \xi_i]$, $\xi_0 \equiv 0$, $\xi_{k+1} \equiv \infty$.

Данные, представленные таким образом, называются группированными данными, которые являются частным случаем цензурированных данных, причем этот случай называется цензурированием интервалом.

Функция правдоподобия в данном случае имеет вид [39]

$$L(N_1, N_2, \dots, N_{k+1}, \theta) = \prod_{i=1}^{k+1}[F(\theta,\xi_i) - F(\theta,\xi_{i-1})]^{N_i}. \qquad (7.9)$$

Следует обратить особое внимание на граничные точки области определения параметра ξ_i. В левой крайней точке $\xi_0 = 0$ функция распределения тождественно равна нулю: $F(\theta,\xi_0)$, поэтому первый член сомножителя (7.9) имеет вид

$$[F(\theta,\xi_1) - F(\theta,\xi_0)]^{N_1} = [F(\theta,\xi_1)]^{N_1}.$$

В правой крайней точке $\xi_{k+1} = \infty$ функция распределения тождественно равна единице $F(\theta,\xi_{k+1}) \equiv 1$, поэтому последний член произведения (7.9) равен

$$[F(\theta,\xi_{k+1}) - F(\theta,\xi_k)]^{N_{k+1}} = [1 - F(\theta,\xi_k)]^{N_{k+1}}.$$

Интервал $[\xi_k, \infty]$ по своему смыслу представляет собой интервал цензурирования справа. В него попадают те элементы, которые не отказали до последнего момента контроля ξ_k. Предполагается, что далее наблюдения не проводились, и элементы, которые не отказали до момента ξ_k, образуют выборку цензурированных справа наработок. С учетом этого окончательно функцию правдоподобия можно записать следующим образом:

$$L(\theta,\{N_i\}) = [F(\theta,\xi_1)]^{N_1} \prod_{i=1}^{k-1}[F(\theta,\xi_{i+1}) - F(\theta,\xi_i)]^{N_{i+1}} [1 - F(\theta,\xi_k)]^{N_{k+1}}. \qquad (7.10)$$

Прологарифмируем выражение (7.10):

$$l(\theta,\{N_i\}) = N_1 \ln F(\theta,\xi_1) + \sum_{i=1}^{k-1} N_{i+1} \ln[F(\theta,\xi_{i+1}) - F(\theta,\xi_i)] + N_{k+1} \ln[1 - F(\theta,\xi_k)].$$

Решение данного уравнения возможно, как правило, численными методами: если допустить в последнем уравнении $P_i(\theta) = F(\theta,\xi_i) - F(\theta,\xi_{i-1})$, то получим, что оценка максимального правдоподобия будет корнем уравнения правдоподобия

$$\sum_{i=1}^{k+1} N_i \frac{1}{P_i(\theta)} \frac{\partial}{\partial\theta} P_i(\theta) = 0. \qquad (7.11)$$

Покажем процедуру вычисления оценки параметра интенсивности отказа для экспоненциального закона распределения наработки. Функция правдоподобия в этом случае будет выглядеть следующим образом:

$$L(\lambda,\{N_i\}) = \prod_{i=1}^{k+1}[\exp(-\lambda\xi_{i-1}) - \exp(-\lambda\xi_i)]^{N_i}.$$

Подставляя в (7.11) $P_i(\lambda) = \exp(-\lambda\xi_{i-1}) - \exp(\lambda\xi_i)$, получаем уравнение для определения параметра λ экспоненциального закона распределения. Получить решение данного уравнения можно численными методами. Для приближенного решения может быть применен итеративный метод Ньютона-Рафсона с начальным значением

$$\lambda_1 = (\theta_1)^{-1} = \left(\sum_{i=1}^{k+1} N_i t_i / n\right)^{-1},$$

где $n = \sum_{i=1}^{k+1} N_i$, $t_i = \frac{1}{2}(\xi_{i-1} + \xi_i)$, $i = \overline{1, k}$, $t_{k+1} = \xi_k + 1$.

Как видно из последнего выражения, даже в самом простом случае экспоненциального распределения наработки до отказа параметры распределения приходится оценивать численными методами. Поэтому можно считать выражение (7.11) окончательным, дальнейшее преобразование которого нецелесообразно. Процедура оценивания параметров закона распределения реализуется исключительно численными методами.

7.5. Примеры оценки показателей законов распределения

В данном параграфе приведем результаты вычисления параметров законов распределения и определения точности в их оценке для ряда законов, имеющих наиболее широкое применение в системном анализе.

Вначале подведем некоторые итоги. Остановимся еще раз на обозначениях, используемых при расчете показателей:

θ – параметр закона распределения; $\{T_i\}$ – реализации наблюдаемой случайной величины (наработки до отказов); $F(\theta,T_i)$ – функция распределения случайной величины, $f(\theta,T_i)$ – плотность распределения. Функция правдоподобия обозначается через $L(\theta,T)$, логарифмическая функция правдоподобия $l(\theta,T)$, причем $l(\theta,T) = \ln L(\theta,T)$; k – объем выборки полных наработок, ν – количество цензурированных данных, т.е. $T_1, T_2,\ldots, T_k$ – реализации полных и $T_1',\ldots,T_\nu'$ – цензурированных наработок.

Для группированных данных имеем следующую информацию: моменты контроля исправности функционирования оборудования $\xi_1, \xi_2,\ldots, \xi_k$, где $0 < \xi_1 < \xi_2 < \ldots < \xi_k < \infty$. В моменты контроля выявляется количество отказавших в интервале времени $(\xi_{i-1}, \xi_i]$ устройств, т.е. наблюдаемыми случайными величинами являются целые неотрицательные числа, характеризующие количество отказавших объектов на рассматриваемом интервале: $N_1, N_2,\ldots, N_{k+1}$; N_i – число устройств, отказавших в интервале $(\xi_{i-1}, \xi_i]$. Как и ранее функцию правдоподобия для полных наработок будем писать без индекса

$$L = \prod_{i=1}^{k} f(\theta,T_i);$$

для данных, имеющих наряду с полными цензурированные справа наработки, будем использовать обозначение

$$L_{\text{П}} = \prod_{i=1}^{k} f(\theta,T_i)\prod_{j=1}^{\nu}\left(1-F(\theta,T_j')\right);$$

аналогично, для имеющих цензурированные слева наработки будем писать

$$L_{\text{Л}} = \prod_{i=1}^{k} f(\theta,T_i)\prod_{j=1}^{\nu} F(\theta,T_j');$$

для цензурированных интервалом или группированных данных

$$L_{\text{Г}} = \prod_{i=1}^{k+1}[F(\theta,\xi_i)-F(\theta,\xi_{i-1})]^{N_i}.$$

После сделанных обозначений приведем результаты оценивания параметров законов распределения и оценок дисперсии данных показателей.

Экспоненциальное распределение

1. Для полных наработок имеем:

$$L(\theta,T) = \prod_{i=1}^{k}\theta\exp(-\theta T_i) = \theta^k\exp\left(-\theta\sum_{i=1}^{k}T_i\right);$$

$$\frac{\partial l(\theta,T)}{\partial\theta} = \frac{\partial\left(k\ln\theta-\theta\sum_{i=1}^{k}T_i\right)}{\partial\theta} = \frac{k}{\theta}-\sum_{i=1}^{k}T_i = 0; \quad \hat\theta = \frac{k}{\sum_{i=1}^{k}T_i}.$$

Таким образом, в данном случае решение получается в явном виде.

2. Для данных, содержащих полные и цензурированные справа наработки:

$$L_{\text{П}}(\theta,T) = \prod_{j=1}^{\nu}\exp(-\theta T_j')\prod_{i=1}^{k}\theta\exp(-\theta T_i) = \theta^k\exp\left(-\theta\sum_{j=1}^{\nu}T_j'\right)\exp\left(-\theta\sum_{i=1}^{k}T_i\right);$$

$$l(\theta,T) = k\ln\theta-\theta\left(\sum_{i=1}^{k}T_i+\sum_{j=1}^{\nu}T_j'\right);$$

$$\frac{\partial l(\theta,T)}{\partial\theta} = \frac{k}{\theta}-\left(\sum_{i=1}^{k}T_i+\sum_{j=1}^{\nu}T_j'\right) = 0; \qquad \hat\theta = \frac{k}{\left(\sum_{i=1}^{k}T_i+\sum_{j=1}^{\nu}T_j'\right)}.$$

Решение так же как и в предыдущем случае получается в явном виде.

3. Для данных, содержащих полные и цензурированные слева наработки:

$$L_{л}(\theta,T)=\prod_{j=1}^{\nu}\left(1-\exp(-\theta T_j')\right)\prod_{i=1}^{k}\theta\exp(-\theta T_i)=\theta^k\exp\left(-\theta\sum_{i=1}^{k}T_i\right)\prod_{j=1}^{\nu}\left(1-\exp(-\theta T_j')\right)$$

$$l(\theta,T)=k\ln\theta-\theta\sum_{i=1}^{k}T_i+\sum_{j=1}^{\nu}\ln\left(1-\exp\left(-\theta T_j'\right)\right);$$

$$\frac{\partial l(\theta,T)}{\partial\theta}=\frac{k}{\theta}-\sum_{i=1}^{k}T_i+\sum_{j=1}^{\nu}\frac{T_j'\exp\left(-\theta T_j'\right)}{1-\exp\left(-\theta T_j\right)}=$$

$$=\frac{k}{\theta}-\sum_{i=1}^{k}T_i-\sum_{j=1}^{\nu}\frac{T_j'\left(1-\exp\left(-\theta T_j'\right)\right)-T_j'}{1-\exp\left(-\theta T_j'\right)}=\frac{k}{\theta}-\sum_{i=1}^{k}T_i-\sum_{j=1}^{\nu}T_j'+\sum_{j=1}^{\nu}\frac{T_j'}{1-\exp\left(-\theta T_j'\right)}=0.$$

Данное уравнение в явном виде не имеет представления. Программная реализация решения подобных уравнений требует применения численных методов.

4. В случае, когда выборка содержит только группированные данные, решение имеет следующий вид:

$$L_{Г}(N_1,N_2,...,N_{k+1},\theta)=\prod_{i=1}^{k+1}[F(\theta,\xi_i)-F(\theta,\xi_{i-1})]^{N_i};$$

$$l=\ln L_{Г}=\sum_{i=1}^{k+1}N_i\ln[F(\theta,\xi_i)-F(\theta,\xi_{i-1})]=\sum_{i=1}^{k+1}N_i\ln[\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)];$$

$$\frac{\partial l}{\partial\theta}=\sum_{i=1}^{k+1}\frac{N_i[\xi_{i-1}\exp(-\theta\xi_{i-1})-\xi_i\exp(-\theta\xi_i)]}{\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)}=0.$$

При равных интервалах цензурирования решение можно представить в следующем виде:

$$L_{Г}(N_1,N_2,...,N_k,\theta)=\prod_{i=1}^{k}[\exp(\theta\xi_{i-1})-\exp(\theta\xi_i)]^{N_i};$$

$$l=\sum_{i=1}^{k}N_i\ln[\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)];$$

$$\frac{\partial l}{\partial\theta}=\sum_{i=1}^{k}\frac{N_i[\xi_i\exp(-\theta\xi_i)-\xi_{i-1}\exp(-\theta\xi_{i-1})]}{\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)}.$$

Если представить $\xi_i=\Delta i$, $\xi_{i-1}=\Delta(i-1)$, где Δ – временной интервал группирования, будем иметь

$$\frac{\partial l}{\partial\theta}=\sum_{i=1}^{k}\frac{N_i[\Delta i(\exp(-\theta\xi_i)-\exp(-\theta\xi_{i-1}))+\Delta\exp(-\theta\xi_{i-1})]}{\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)}=$$

$$=-\sum_{i=1}^{k}N_i\Delta i+\sum_{i=1}^{k}\frac{\Delta\exp(-\theta\xi_{i-1})}{\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)}=0;$$

$$\sum_{i=1}^{k}N_i i=\sum_{i=1}^{k}\frac{\exp(-\theta\xi_{i-1})}{\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)}=\sum_{i=1}^{k}\frac{\exp(-\theta\Delta(i-1))}{\exp(-\theta\Delta(i-1))[1-\exp(-\theta\Delta)]};$$

$$\sum_{i=1}^{k}N_i i=\sum_{i=1}^{k}\frac{1}{1-\exp(-\theta\Delta)}=\frac{k}{1-\exp(-\theta\Delta)};$$

$$\hat\theta=-\frac{\ln\left(1-k\Big/\sum_{i=1}^{k}N_i i\right)}{\Delta}.$$

Приведем результаты определения точности оценки параметра экспоненциального закона распределения.

Дисперсионная матрица для вектора параметров определяется путем транспонирования информационной матрицы, элементы которой имеют вид

$$a_{ij}=-M\left[\frac{\partial^2\ln L(\theta,t)}{\partial\theta_i\partial\theta_j}\right],$$

в нашем случае оценки одного параметра необходимо определить вторую производную по параметру:

$$D^{-1}=-\frac{\partial^2 l}{\partial\theta^2}.$$

В итоге будем иметь следующие результаты.

1. Для полных наработок:

$$\frac{\partial l(\theta,T)}{\partial\theta}=\frac{\partial\left(k\ln\theta-\theta\sum_{i=1}^{k}T_i\right)}{\partial\theta}=\frac{k}{\theta}-\sum_{i=1}^{k}T_i;$$

$$D^{-1}=-\frac{\partial^2 l}{\partial\theta^2}=-\left(-\frac{k}{\hat{\theta}^2}\right);\ D=\frac{\hat{\theta}^2}{k}.$$

2. Для выборки, содержащей полные и цензурированные справа наработки:

$$\frac{\partial l(\theta,T)}{\partial\theta}=\frac{k}{\theta}-\left(\sum_{i=1}^{k}T_i+\sum_{j=1}^{\nu}T_j'\right)=0;\ D^{-1}=-\frac{\partial^2 l}{\partial\theta^2}=-\left(-\frac{k}{\hat{\theta}^2}\right);\ D=\frac{\hat{\theta}^2}{k}.$$

3. Для выборки, содержащей полные и цензурированные слева наработки:

$$\frac{\partial l(\theta,T)}{\partial\theta}=\frac{k}{\theta}-\sum_{i=1}^{k}T_i-\sum_{i=1}^{\nu}T_j'+\sum_{j=1}^{\nu}\frac{T_j'\exp(-\theta T_j')}{1-\exp(-\theta T_j')};$$

$$\frac{\partial^2 l(\theta,T)}{\partial\theta^2}=-\frac{k}{\theta^2}+\sum_{j=1}^{\nu}\frac{-(T_j')^2\exp\left(-\theta T_j'\right)\left(1-\exp\left(-\theta T_j'\right)\right)-T_j'\exp\left(-\theta T_j'\right)T_j'\exp\left(-\theta T_j'\right)}{\left(1-\exp\left(-\theta T_j'\right)\right)^2}=$$

$$=-\frac{k}{\theta^2}+\sum_{j=1}^{\nu}\frac{-\left(T_j'\right)^2\exp\left(-\theta T_j'\right)+\left(T_j'\right)^2\exp\left(-2\theta T_j'\right)-\left(T_j'\right)^2\exp\left(-2\theta T_j'\right)}{\left(1-\exp(-\theta T_j')\right)^2}=$$

$$=-\frac{k}{\theta^2}-\sum_{j=1}^{\nu}\frac{-(T_j')^2\exp\left(-\theta T_j'\right)}{\left(1-\exp\left(-\theta T_j'\right)\right)^2};$$

$$D^{-1}=-\frac{\partial^2 l}{\partial\theta^2};\ D=\frac{1}{\dfrac{k}{\hat{\theta}^2}+\sum_{j=1}^{\nu}\dfrac{(T_j')^2\exp(-\hat{\theta}T_j')}{\left(1-\exp(-\hat{\theta}T_j')\right)^2}}.$$

4. Для группированных данных:

$$\frac{\partial l}{\partial\theta}=\sum_{i=1}^{k}\frac{N_i\left[\xi_i\exp(-\theta\xi_i)-\xi_{i-1}\exp(-\theta\xi_{i-1})\right]}{\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)};$$

$$\frac{\partial^2 l}{\partial\theta^2}=\sum_{i=1}^{k}\frac{1}{[\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)]^2}\Big[N_i\left(\xi_{i-1}^2\exp(-\theta\xi_{i-1})-\xi_i{}^2\exp(-\theta\xi_i)\right)\times$$

$$\times\left(\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)\right)-\left(\xi_i\exp(-\theta\xi_i)-\xi_{i-1}\exp(-\theta\xi_{i-1})\right)\times$$

$$\times\left(\xi_i\exp(-\theta\xi_i)-\xi_{i-1}\exp(-\theta\xi_{i-1})\right)=$$

$$=\sum_{i=1}^{k}\frac{N_i}{[\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)]^2}\Big[\xi_{i-1}^2\exp\left(-2\theta\xi_{i-1}\right)+\xi_i{}^2\exp\left(-2\theta\xi_i\right)-\xi_i{}^2\exp\left(-\theta(\xi_i+\xi_{i-1})\right)-$$

$$-\xi_{i-1}{}^2\exp\left(-\theta(\xi_i+\xi_{i-1})\right)-\left[\xi_i^2\exp\left(-2\theta\xi_i\right)+\xi_{i-1}{}^2\exp\left(-2\theta\xi_{i-1}\right)-2\xi_{i-1}\xi_i\exp\left(-\theta(\xi_i+\xi_{i-1})\right)\right]\Big]=$$

$$=\sum_{i=1}^{k}N_i\frac{\exp\left(-\theta\left(\xi_i+\xi_{i-1}\right)\right)(-1)\left(\xi_i-\xi_{i-1}\right)^2}{\left[\exp(-\theta\xi_{i-1})-\exp(-\theta\xi_i)\right]^2}.$$

$$D=\frac{1}{\sum_{i=1}^{k}N_i\dfrac{(\xi_i-\xi_{i-1})^2\exp(-\hat{\theta}(\xi_i+\xi_{i-1}))}{[\exp(-\hat{\theta}\xi_{i-1})-\exp(-\hat{\theta}\xi_{i-1})]^2}}.$$

В случае равных интервалов группирования:

$$\frac{\partial l}{\partial\theta}=-\sum_{i=1}^{k}N_i\Delta i+\sum_{i=1}^{k}\frac{\Delta\exp(-\theta\Delta(i-1))}{\exp(-\theta\Delta(i-1))-\exp(-\theta\Delta i)}=-\sum_{i=1}^{k}N_i\Delta i+\Delta\sum_{i=1}^{k}\frac{1}{[1-\exp(-\theta\Delta)]};$$

$$\frac{\partial^2 l}{\partial\theta^2}=-\Delta^2\sum_{i=1}^{k}\frac{\exp(-\theta\Delta)}{[1-\exp(-\theta\Delta)]^2};\ D=1\Big/\left[\Delta^2\sum_{i=1}^{k}\frac{\exp(-\hat{\theta}\Delta)}{[1-\exp(-\hat{\theta}\Delta)]^2}\right],$$

имеем в виду, что:

$$\exp(-\hat{\theta}\Delta)=\exp\left(\ln\left[1-\frac{k}{\sum N_i i}\right]\right)=1-\frac{k}{\sum N_i i},$$

тогда:

$$D=\frac{1}{\Delta^2\sum\limits_{i=1}^{k}\frac{1-\frac{k}{\sum N_i i}}{\left(\frac{k}{\sum N_i i}\right)^2}}=\frac{1}{\Delta^2\sum\limits_{i=1}^{k}\left[\frac{\left(\sum N_i i\right)^2}{k^2}-\frac{\sum N_i i}{k}\right]};\ D=\frac{k}{\Delta^2\left(\sum N_i i\right)\left[\sum N_i i-k\right]}.$$

Нормальное распределение

Плотность и функция распределения для нормального закона распределения имеют вид:

$$f(t,m,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right);$$

$$F(t,m,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{t}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx.$$

В данных формулах приняты обозначения: m – математическое ожидание; σ – среднее квадратическое отклонение. При получении оценок параметров и определении точности в их оценке будем также в качестве математического ожидания использовать обозначение θ_1, в качестве среднего квадратического отклонения – θ_2. Приведем результаты вычислений.

1. В случае полных наработок имеем:

$$L(\theta_1,\theta_2,\{T_i\})=\frac{1}{(\sqrt{2\pi}\theta_2)^k}\exp\left(-\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)^2}{2\theta_2^2}\right);$$

$$l=\ln L=-k\ln(\sqrt{2\pi}\theta_2)-\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)^2}{2\theta_2^2}=-k\ln(\sqrt{2\pi}\theta_2)-\frac{k\theta_1^2-2\theta_1\sum T_i+\sum T_i^2}{2\theta_2^2};$$

$$\frac{\partial l}{\partial\theta_1}=-\frac{\theta_1}{\theta_2^2}k+\frac{\sum T_i}{\theta_2^2}=0;$$

откуда получаем

$$\hat{\theta}_1=\sum_{i=1}^{k}T_i\Big/k.$$

Производя аналогичные действия для второго параметра, получим

$$\frac{\partial l}{\partial\theta_2}=-\frac{k}{\theta_2}+\frac{k\theta_1^2-2\theta_1\sum T_i+\sum T_i^2}{\theta_2^3}=0;$$

$$k\theta_2^2=k\theta_1^2-2\theta_1\sum T_i+\sum T_i^2=\sum(T_i-\theta_1)^2$$

в итоге получаем следующую формулу

$$\hat{\theta}_2^2=\frac{1}{k}\sum_{i=1}^{k}(T_i-\hat{\theta}_1)^2.$$

2. Для выборок, содержащих полные и цензурированные справа наработки, функция правдоподобия имеет вид

$$L_\Pi(\theta_1,\theta_2,\{T_i\})=\frac{1}{(\sqrt{2\pi}\theta_2)^k}\exp\left(-\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)^2}{2\theta_2^2}\right)\prod_{i=1}^{\nu}\left(1-\int\limits_{-\infty}^{T_i}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)\right)dt,$$

логарифмическая функция правдоподобия:

$$l=-\frac{k}{2}\ln(2\pi)-k\ln\theta_2-\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)^2}{2\theta_2^2}+\sum_{j=1}^{\nu}\ln\left[1-\int\limits_{-\infty}^{T_j'}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(\frac{-(t-\theta_1)^2}{2\theta_2^2}\right)dt\right];$$

и, наконец, частные производные определяются выражениями

$$\frac{\partial l}{\partial\theta_1}=\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)}{\theta_2^2}+\sum_{j=1}^{\nu}\frac{-\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(T_j'-\theta_1)^2}{2\theta_2^2}\right)}{1-\int\limits_{-\infty}^{T_j'}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt}=0;$$

$$\frac{\partial l}{\partial\theta_2}=\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)^2}{\theta_2^3}-\frac{k}{\theta_2}+\sum_{j=1}^{\nu}\frac{-\frac{1}{\sqrt{2\pi}\theta_2^2}\exp\left(-\frac{(T_j'-\theta_1)^2}{2\theta_2^2}\right)(T_j'-\theta_1)}{1-\int\limits_{-\infty}^{T_j'}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt}=0.$$

В данном случае результат в явном виде получить не удается, поэтому необходимо решать трансцендентные уравнения.

3. Для выборок, содержащих полные и цензурированные слева наработки, функция правдоподобия записывается

$$L_{л}(\theta_1,\theta_2,\{T_i\})=\frac{1}{(\sqrt{2\pi}\theta_2)^k}\exp\left(-\frac{\sum_{i=1}^{k}(T_i-\theta_1)^2}{2\theta_2^2}\right)\prod_{j=1}^{\nu}\int_{-\infty}^{T_j'}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt;$$

логарифмическая функция правдоподобия:

$$l=\ln l=-\frac{k}{2}\left(\ln 2\pi+\ln\theta_2^2\right)-\frac{\sum(T_i-\theta_1)^2}{2\theta_2^2}+\sum_{j=1}^{\nu}\ln\int_{-\infty}^{T_2}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt;$$

частные производные:

$$\frac{\partial l}{\partial\theta_1}=\frac{\sum_{i=1}^{k}(T_i-\theta_1)}{\theta_2^2}+\sum_{j=1}^{\nu}\frac{-\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(T_j'-\theta_1)^2}{2\theta_2^2}\right)}{\int_{-\infty}^{T_j'}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt}=0;$$

$$\frac{\partial l}{\partial\theta_2}=\frac{\sum_{i=1}^{k}(T_i-\theta_1)^2}{\theta_2^3}-\frac{k}{\theta_2}+\sum_{j=1}^{\nu}\frac{-\frac{1}{\sqrt{2\pi}\theta_2^2}\exp\left(-\frac{(T_j'-\theta_1)^2}{2\theta_2^2}\right)(T_j'-\theta_1)}{\int_{-\infty}^{T_j'}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt}=0.$$

Для выборок, содержащих цензурированные слева наработки, делаем тот же вывод, что и в предыдущем случае, а именно, решение необходимо искать численными методами.

4. Для группированных данных итоговые оценки получаются таким образом. Функция правдоподобия:

$$L_{Г}(\theta_1,\theta_2,\{T_i\})=\prod_{i=1}^{N}\left(\int_{-\infty}^{T_{i-1}}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt-\int_{-\infty}^{T_i}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt\right)^{N_i};$$

логарифмическая функция правдоподобия:

$$l=\sum_{i=1}^{N}N_i\ln\left[\int_{-\infty}^{T_{i-1}}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt-\int_{-\infty}^{T_i}\frac{1}{\sqrt{2\pi}\theta_2}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt\right];$$

производные от нее по параметрам:

$$\frac{\partial l}{\partial\theta_1}=\sum_{i=1}^{N}N_i\frac{-\frac{1}{\sqrt{2\pi}\theta_2}\left[\exp\left(-\frac{(T_{i-1}-\theta_1)^2}{2\theta_2^2}\right)-\exp\left(-\frac{(T_i-\theta_1)^2}{2\theta_2^2}\right)\right]}{\frac{1}{\sqrt{2\pi}\theta_2}\left[\int_{-\infty}^{T_{i-1}}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt-\int_{-\infty}^{T_i}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt\right]}=0;$$

$$\frac{\partial l}{\partial\theta_2}=\sum_{i=1}^{N}N_i\frac{\frac{-1}{\sqrt{2\pi}\theta_2^2}\left[\exp\left(-\frac{(T_{i-1}-\theta_1)^2}{2\theta_2^2}\right)(T_{i-1}-\theta_1)-\exp\left(-\frac{(T_i-\theta_1)^2}{2\theta_2^2}\right)(T_i-\theta_1)\right]}{\frac{1}{\sqrt{2\pi}\theta_2}\left[\int_{-\infty}^{T_{i-1}}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt-\int_{-\infty}^{T_i}\exp\left(-\frac{(t-\theta_1)^2}{2\theta_2^2}\right)dt\right]}=0.$$

Рассмотрим вопрос вычисления точности в полученных оценках. Определим вторые производные для случая, когда имеются в наличии полные наработки.

$$\frac{\partial^2 l}{\partial\theta_1^2}=\left(\frac{\sum_{i=1}^{k}(T_i-\theta_1)}{\hat{\theta}_2^2}\right)'_{\theta_1}=-\frac{k}{\hat{\theta}_2^2};\quad D_{\theta_1}=\frac{\hat{\theta}_2^2}{k};$$

$$\frac{\partial^2 l}{\partial\theta_2^2}=\left(-\frac{k}{\theta_2}+\frac{\sum_{i=1}^{k}(T_i-\hat{\theta}_1)^2}{\theta_2^3}\right)'_{\theta_2}=-\frac{k}{\theta_2^2}-3\frac{\sum_{i=1}^{k}(T_i-\hat{\theta}_1)^2}{\theta_2^4};$$

$$\frac{1}{D_{\theta_2}}=3\frac{\sum_{i=1}^{k}(T_i-\hat{\theta}_1)^2}{\theta_2^4}+\frac{k}{\theta_2^2}.$$

Для всех остальных типов данных при расчете дисперсии получаются результаты, не имеющие представления в явном виде. Поиск решения осуществляется численными методами, поэтому итоговые формулы не приводятся.

Усеченное нормальное распределение

Плотность усеченного нормального распределения имеет следующий вид:

$$f(x)=\begin{cases}\frac{c}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), & x>0;\\ 0, & x\le 0,\end{cases}$$

здесь $c = \dfrac{1}{\int\limits_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(x-m)^2}{2\sigma^2}\right) dx}$ – нормировочная константа.

Функция распределения равна

$$F(x) = \frac{c}{\sqrt{2\pi}\sigma} \int\limits_0^t \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

Будем обозначать θ_1 – математическое ожидание, θ_2 – среднее квадратическое отклонение. Рассмотрим последовательно вычисление оценок параметров для различных схем наблюдений.

1. Функция правдоподобия для полных наработок имеет вид

$$L = \frac{c^k}{(\sqrt{2\pi}\theta_2)^k} \exp\left(-\frac{\sum\limits_{i=1}^{k} (T_i - \theta_1)^2}{2\theta_2^2}\right).$$

Соответствующая логарифмическая функция правдоподобия

$$l = -\frac{\sum\limits_{i=1}^{k} (T_i - \theta_1)^2}{2\theta_2^2} - k \ln \int\limits_0^{\infty} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right) dx.$$

Обозначим: $A(x) = \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)$, тогда

$$\frac{\partial l}{\partial \theta_1} = \frac{\sum\limits_{i=1}^{k} (T_i - \theta_1)}{\theta_2^2} + k \frac{\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right)}{\int\limits_0^{\infty} A(x)dx} = 0;$$

$$\frac{\partial l}{\partial \theta_2} = \frac{\sum\limits_{i=1}^{k} (T_i - \theta_1)^2}{\theta_2^3} - k \frac{\frac{\theta_1}{\theta_2} \exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right) + \frac{1}{\theta_2} \int\limits_0^{\infty} A(x)dx}{\int\limits_0^{\infty} A(x)dx} = 0.$$

Решая данную систему уравнений, получаем оценки параметров усеченного нормального закона распределения для случая полных наработок.

2. Для выборок, содержащих полные и цензурированные справа наработки, функцию правдоподобия можно записать

$$L = \frac{c^k}{(\sqrt{2\pi}\theta_2)^k} \exp\left(-\frac{\sum\limits_{i=1}^{k} (T_i - \theta_1)^2}{2\theta_2^2}\right) \prod_{j=1}^{v} \left(1 - \frac{c}{\sqrt{2\pi}\theta_2} \int\limits_0^{T_j'} A(x)dx\right),$$

логарифмическая функция правдоподобия равна

$$l = -\frac{\sum\limits_{i=1}^{k} (T_i - \theta_1)^2}{2\theta_2^2} - k \ln \int\limits_0^{\infty} A(x)dx + \sum_{j=1}^{v} \ln\left(1 - \frac{\int\limits_0^{T_j'} A(x)dx}{\int\limits_0^{\infty} A(x)dx}\right).$$

И, наконец, производные по параметрам определим следующим образом:

$$\frac{\partial l}{\partial \theta_1} = \frac{\sum\limits_{i=1}^{k} (T_i - \theta_1)}{\theta_2^2} + k \frac{\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right)}{\int\limits_0^{\infty} A(x)dx} +$$

$$+ \sum_{j=1}^{v} \frac{\left[\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right) - A(T_j')\right] \int\limits_0^{\infty} A(x)dx + \exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right) \int\limits_0^{T_j'} A(x)dx}{\left[\int\limits_0^{\infty} A(x)dx\right]^2 - \int\limits_0^{\infty} A(x)dx \int\limits_0^{T_j'} A(x)dx} = 0;$$

$$\frac{\partial l}{\partial \theta_2} = \frac{\sum\limits_{i=1}^{k} (T_i - \theta_1)^2}{\theta_2^3} - k \frac{\frac{\theta_1}{\theta_2} \exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right) + \frac{1}{\theta_2} \int\limits_0^{\infty} A(x)dx}{\int\limits_0^{\infty} A(x)dx} +$$

$$+ \sum_{j=1}^{v} \frac{\frac{\theta_1}{\theta_2} \exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right) \left[\int\limits_0^{\infty} A(x)dx - \int\limits_0^{T_j'} A(x)dx\right] - \frac{T_j' - \theta_1}{\theta_2} A(T_j') \int\limits_0^{\infty} A(x)dx}{\left[\int\limits_0^{\infty} A(x)dx\right]^2 - \int\limits_0^{\infty} A(x)dx \int\limits_0^{T_j'} A(x)dx} = 0.$$

3. Для выборок, содержащих полные и цензурированные слева наработки, функция правдоподобия записывается

$$L=\frac{c^k}{(\sqrt{2\pi}\theta_2)^k}\exp\left(-\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)^2}{2\theta_2^2}\right)\frac{c^\nu}{(\sqrt{2\pi}\theta_2)^\nu}\prod_{j=1}^{\nu}\int\limits_0^{T_j'}A(x)dx.$$

Далее вычисляем логарифмическую функцию

$$l=-\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)^2}{2\theta_2^2}-(k+\nu)\ln\int\limits_0^{\infty}A(x)dx+\sum_{j=1}^{\nu}\ln\int\limits_0^{T_j'}A(x)dx$$

и производные для вычисления оценок параметров

$$\frac{\partial l}{\partial\theta_1}=\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)}{\theta_2^2}+(k+\nu)\frac{\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right)}{\int\limits_0^{\infty}A(x)dx}+\sum_{j=1}^{\nu}\frac{\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right)-\exp\left(-\frac{(T_j'-\theta_1)^2}{2\theta_2^2}\right)}{\int\limits_0^{T_j'}A(x)dx}=0;$$

$$\frac{\partial l}{\partial\theta_2}=\frac{\sum\limits_{i=1}^{k}(T_i-\theta_1)^2}{\theta_2^3}-(k+\nu)\frac{\frac{\theta_1}{\theta_2}\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right)+\frac{1}{\theta_2}\int\limits_0^{\infty}A(x)dx}{\int\limits_0^{\infty}A(x)dx}+$$

$$+\sum_{j=1}^{\nu}\frac{\frac{1}{\theta_2}\int\limits_0^{T_j'}A(x)dx-\frac{T_j'-\theta_1}{\theta_2}A(T_j')-\frac{\theta_1}{\theta_2}\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right)}{\int\limits_0^{T_j'}A(x)dx}=0.$$

4. Для группированных данных итоговые оценки получаются таким образом.

Функция правдоподобия имеет вид

$$L=\prod_{i=1}^{N}\left(\frac{c}{\sqrt{2\pi}\theta_2}\left[\int\limits_0^{T_i'}A(x)dx-\int\limits_0^{T_{i-1}'}A(x)dx\right]\right)^{N_i};$$

логарифмическая функция правдоподобия:

$$l=\left(\sum_{i=1}^{N}N_i\right)\left(-\ln\int\limits_0^{\infty}A(x)dx+\ln\left[\int\limits_0^{T_i}A(x)dx-\int\limits_0^{T_{i-1}}A(x)dx\right]\right).$$

Производные по параметрам равны

$$\frac{\partial l}{\partial\theta_1}=\left(\sum_{i=1}^{N}N_i\right)\frac{\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right)}{\int\limits_0^{\infty}A(x)dx}+\left(\sum_{i=1}^{N}N_i\right)\frac{A(T_i)-A(T_{i-1})}{\int\limits_0^{T_i}A(x)dx-\int\limits_0^{T_{i-1}}A(x)dx}=0;$$

$$\frac{\partial l}{\partial\theta_2}=-\left(\sum_{i=1}^{N}N_i\right)\frac{\frac{\theta_1}{\theta_2}\exp\left(-\frac{\theta_1^2}{2\theta_2^2}\right)+\frac{1}{\theta_2}\int\limits_0^{\infty}A(x)dx}{\int\limits_0^{\infty}A(x)dx}+$$

$$+\left(\sum_{i=1}^{N}N_i\right)\sum_{j=1}^{\nu}\frac{-\frac{T_i-\theta_1}{\theta_2}A(T_i)+\int\limits_0^{T_i}A(x)dx+\frac{T_{i-1}-\theta_1}{\theta_2}A(T_{i-1})-\int\limits_0^{T_{i-1}}A(x)dx}{\int\limits_0^{T_i}A(x)dx-\int\limits_0^{T_{i-1}}A(x)dx}=0.$$

Как видно из приведенных выражений, для определения параметров усеченного нормального закона распределения необходимо решать систему уравнений численными методами.

Логарифмически нормальиое распределеиие

Логарифмически нормальному закону распределения подчиняется случайная величина t, логарифм которой распределен по нормальному закону.

Плотность и функция распределения имеют вид

$$f(t)=\frac{1}{t\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln t-\mu)^2}{2\sigma^2}\right);\ F(t)=\int\limits_{-\infty}^{\ln t}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx;$$

Пусть $\mu=\theta_1$, $\sigma=\theta_2$.

1. Результаты расчетов для полных наработок следующие:

$$L=\frac{1}{\sigma^k(2\pi)^{k/2}\prod\limits_{i=1}^{k}T_i}\exp\left(-\frac{\sum\limits_{i=1}^{k}(\ln T_i-\mu)^2}{2\sigma^2}\right);$$

$$l=\ln\left(\frac{1}{\theta_2^{k}(2\pi)^{k/2}\prod\limits_{i=1}^{k}T_i}\exp\left(-\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{2\theta_2^{2}}\right)\right)=$$

$$=-\sum_{i=1}^{k}\ln T_i-\frac{k}{2}\ln(2\pi)-k\ln\theta_2-\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{2\theta_2^2};$$

$\dfrac{\partial l}{\partial\theta_1}=\dfrac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)}{\theta_2^2}=0$, откуда получаем $\hat\theta_1=\dfrac{\sum\limits_{i=1}^{k}\ln T_i}{k}$;

$$\frac{\partial l}{\partial\theta_2}=-\frac{k}{\theta_2}+\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{\theta_2^3}=0;\ \ \hat\theta_2^2=\frac{\sum\limits_{i=1}^{k}(\ln T_i-\hat\theta_1)^2}{k}.$$

2. Для выборок, содержащих полные и цензурированные справа наработки:

$$L=\frac{1}{\theta_2^k(\sqrt{2\pi})^k\prod\limits_{i=1}^{k}T_i}\exp\left(-\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{2\theta_2^2}\right)\prod_{j=1}^{\nu}\left[1-\int_{-\infty}^{\ln T_j'}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx\right];$$

$$l=-\sum_{i=1}^{k}\ln T_i-\frac{k}{2}\ln 2\pi-k\ln\theta_2-\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{2\theta_2^2}+\sum_{j=1}^{\nu}\ln\left[1-\int_{-\infty}^{\ln T_j'}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx\right];$$

$$\frac{\partial l}{\partial\theta_1}=\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)}{\theta_2^2}+\sum_{j=1}^{\nu}\frac{\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(\ln T_j'-\theta_1)^2}{2\theta_2^2}\right)}{1-\int\limits_{-\infty}^{\ln T_j'}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx}=0;$$

$$\frac{\partial l}{\partial\theta_2}=-\frac{k}{\theta_2}+\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{\theta_2^3}+\sum_{j=1}^{\nu}\frac{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\ln T_j'-\theta_1)^2}{2\theta_2^2}\right)\left(-\frac{\ln T_j'-\theta_1}{\theta_2^2}\right)}{1-\int\limits_{-\infty}^{\ln T_j'}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx}=0.$$

3. Для выборок, содержащих полные и цензурированные слева наработки:

$$L=\frac{1}{\theta_2^k(\sqrt{2\pi})^k\prod\limits_{i=1}^{k}T_i}\exp\left(-\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{2\theta_2^2}\right)\prod_{j=1}^{\nu}\int_{-\infty}^{\ln T_j'}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx;$$

$$\frac{\partial l}{\partial\theta_1}=\left(-\sum_{i=1}^{k}\ln T_i-\frac{k}{2}\ln(2\pi)-k\ln\theta_2-\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{2\theta_2^2}+\sum_{j=1}^{\nu}\ln\int_{-\infty}^{\ln T_j'}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx\right)'_{\theta_1}=$$

$$=\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)}{\theta_2^2}-\sum_{j=1}^{\nu}\frac{\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(\ln T_j'-\theta_1)^2}{2\theta_2^2}\right)}{\int\limits_{-\infty}^{\ln T_j'}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx}=0;$$

$$\frac{\partial l}{\partial\theta_2}=-\frac{k}{\theta_2}+\frac{\sum\limits_{i=1}^{k}(\ln T_i-\theta_1)^2}{\theta_2^3}-\sum_{j=1}^{\nu}\frac{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\ln T_j'-\theta_1)^2}{2\theta_2^2}\right)\left(-\frac{\ln T_j'-\theta_1}{\theta_2^2}\right)}{\int\limits_{-\infty}^{\ln T_j'}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx}=0.$$

4. Для группированных данных:

$$L=\prod_{i=1}^{N}\left[\int_{-\infty}^{\ln T_i}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx-\int_{-\infty}^{\ln T_{i-1}}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx\right];$$

$$\frac{\partial l}{\partial\theta_1}=\left(\sum_{i=1}^{N}N_i\ln\left[\int_{-\infty}^{\ln T_i}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx-\int_{-\infty}^{\ln T_{i-1}}\frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx\right]\right)'_{\theta_1}=$$

$$=\left(-\sum_{i=1}^{N}\left[N_i\ln\theta_2-\frac{N_i}{2}\ln(2\pi)\right]+\sum_{i=1}^{N}N_i\ln\left[\int_{-\infty}^{\ln T_i}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx-\int_{-\infty}^{\ln T_{i-1}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)dx\right]\right)'_{\theta_1}=$$

$$= \frac{1}{\theta_2} \sum_{i=1}^{N} \frac{N_i \left[-\exp\left(-\frac{(\ln T_i - \theta_1)^2}{2\theta_2^2} \right) + \exp\left(-\frac{(\ln T_{i-1} - \theta_1)^2}{2\theta_2^2} \right) \right]}{\int\limits_{-\infty}^{\ln T_i} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2} \right) dx - \int\limits_{-\infty}^{\ln T_{i-1}} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2} \right) dx} = 0;$$

$$\frac{\partial l}{\partial \theta_2} = \frac{1}{\theta_2} \sum_{i=1}^{N} \frac{-\frac{\ln T_i - \theta_1}{\theta_2^2} \exp\left(-\frac{(\ln T_i - \theta_1)^2}{2\theta_2^2} \right) + \frac{\ln T_{i-1} - \theta_1}{\theta_2^2} \exp\left(-\frac{(\ln T_{i-1} - \theta_1)^2}{2\theta_2^2} \right)}{\int\limits_{-\infty}^{\ln T_i} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2} \right) dx - \int\limits_{-\infty}^{\ln T_{i-1}} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2} \right) dx} = 0.$$

Распределение Вейбулла

Плотность и функция распределения для распределения Вейбулла:

$$f(t) = a(\lambda t)^{a-1} \exp\left(-(\lambda t)^a\right); \quad F(t) = 1 - \exp\left(-(\lambda t)^a\right),$$

где α – параметр формы; λ – параметр масштаба. Будем обозначать $\alpha = \theta_1$, $\lambda = \theta_2$. Приведем итоговые результаты для различных схем получаемых данных.

1. Для случая полных наработок:

$$L(\theta_1, \theta_2, \{T_i\}) = \theta_1{}^k \theta_2{}^{\theta_1 k} \exp\left(-\theta_2{}^{\theta_1} \sum_{i=1}^{k} T_i^{\theta_1} \right) \prod_{i=1}^{k} (T_i)^{\theta_1 - 1};$$

$$\frac{\partial l}{\partial \theta_1} = \left(k \ln \theta_1 + k\theta_1 \ln \theta_2 + (\theta_1 - 1) \sum_{i=1}^{k} \ln T_i - \theta_2{}^{\theta_1} \sum_{i=1}^{k} T_i^{\theta_1} \right)'_{\theta_1} = 0;$$

$$\frac{\partial l}{\partial \theta_2} = \left(k \ln \theta_1 + k\theta_1 \ln \theta_2 + (\theta_1 - 1) \sum_{i=1}^{k} \ln T_i - \theta_2{}^{\theta_1} \sum_{i=1}^{k} T_i^{\theta_1} \right)'_{\theta_2} = 0;$$

$$\frac{k}{\theta_1} + k \ln \theta_2 + \sum_{i=1}^{k} \ln T_i - \sum_{i=1}^{k} (T_i \theta_2)^{\theta_1} \cdot \ln(T_i \theta_2) = 0;$$

$$\frac{k}{\theta_2} - \theta_2^{\theta_1 - 1} \theta_1 \sum_{i=1}^{k} T_i^{\theta_1} = 0.$$

2. Для выборок, содержащих полные и цензурированные справа наработки:

$$L(\theta_1, \theta_2, \{T_i\}) = \theta_1{}^k \theta_2{}^{\theta_1 k} \prod_{i=1}^{k} (T_i)^{\theta_1 - 1} \exp\left(-\theta_2{}^{\theta_1} \sum_{j=1}^{\nu} (T_j')^{\theta_1} \right) \exp\left(-\theta_2^{\theta_1} \sum_{i=1}^{k} T_i^{\theta_1} \right);$$

$$\frac{\partial l}{\partial \theta_1} = \left(k \ln \theta_1 + k \ln \theta_2 + \theta_1 \sum_{i=1}^{k} \ln T_i - \theta_2 \left[\sum_{i=1}^{k} T_i^{\theta_1} + \sum_{j=1}^{\nu} (T_j')^{\theta_1} \right] \right)'_{\theta_1} = 0;$$

$$\frac{\partial l}{\partial \theta_2} = \left(k \ln \theta_1 + k \ln \theta_2 + \theta_1 \sum_{i=1}^{k} \ln T_i - \theta_2 \left[\sum_{i=1}^{k} T_i^{\theta_1} + \sum_{j=1}^{\nu} (T_j')^{\theta_1} \right] \right)'_{\theta_2} = 0.$$

После элементарных преобразований

$$\frac{k}{\theta_1} + k \ln \theta_2 + \sum_{i=1}^{k} \ln T_i - \theta_2^{\theta_1} \ln \theta_2 \left[\sum_{i=1}^{k} T_i^{\theta_1} + \sum_{j=1}^{\nu} (T_j')^{\theta_1} \right] -$$

$$-\theta_2^{\theta_1} \left[\sum_{i=1}^{k} T_i^{\theta_1} \ln T_i + \sum_{j=1}^{\nu} (T_j')^{\theta_1} \ln T_j' \right] = 0;$$

$$k\theta_1 - \theta_1 \theta_2^{\theta_1 - 1} \left(\sum_{i=1}^{k} T_i^{\theta_1} + \sum_{j=1}^{\nu} (T_j')^{\theta_1} \right) = 0.$$

3. Для выборок, содержащих полные и цензурированные слева наработки:

$$L(\theta_1, \theta_2, \{T_i\}) = \theta_1{}^k \theta_2{}^{\theta_1 k} \prod_{i=1}^{k} (T_i)^{\theta_1 - 1} \exp\left(-\theta_2{}^{\theta_1} \sum_{i=1}^{k} T_i^{\theta_1} \right) \prod_{j=1}^{\nu} \left(1 - \exp\left(-(\theta_2 T_j')^{\theta_1} \right) \right);$$

$$\frac{\partial l}{\partial \theta_1} = \frac{k}{\theta_1} + k \ln \theta_2 + \sum_{i=1}^{k} \ln T_i - \sum_{i=1}^{k} \left[(T_i \theta_2)^{\theta_1} \ln(T_i \theta_2) \right] +$$

$$+ \sum_{j=1}^{\nu} \frac{\exp\left(-(T_j' \theta_2)^{\theta_1} \right) (T_j' \theta_2)^{\theta_1} \ln(T_j' \theta_2)}{1 - \exp\left(-(T_j' \theta_2)^{\theta_1} \right)} = 0;$$

$$\frac{\partial l}{\partial \theta_2} = k\theta_1 - \theta \sum_{i=1}^{k} T_i^{\theta_1} - \sum_{i=1}^{k} \left[(T_i \theta_2)^{\theta_1} \ln(T_i \theta_2) \right] +$$

$$+ \sum_{j=1}^{\nu} \frac{\exp\left(-(T_j' \theta_2)^{\theta_1} \right) (T_j' \theta_2)^{\theta_1} \ln(T_j' \theta_2)}{1 - \exp\left(-(T_j' \theta_2)^{\theta_1} \right)} = 0.$$

4. Для группированных данных:

$$L(\theta_1, \theta_2, \{T_i\}) = \prod_{i=1}^{N} \left(\exp\left(-\theta_2 T_{i-1} \right)^{\theta_1} - \exp\left(-\theta_2 T_i \right)^{\theta_1} \right)^{N_i};$$

$$\frac{\partial l}{\partial\theta_1}=\left[\sum_{i=1}^{N}N_i\ln\left[\exp\left(-(T_{i-1}\theta_2)^{\theta_1}\right)-\exp\left(-(T_i\theta_2)^{\theta_1}\right)\right]\right]'_{\theta_1}=$$

$$=\sum_{i=1}^{N}N_i\frac{\exp\left(-(T_i\theta_2)^{\theta_1}\right)(T_i\theta_2)^{\theta_1}\ln(T_i\theta_2)-\exp\left(-(T_{i-1}\theta_2)^{\theta_1}\right)(T_{i-1}\theta_2)^{\theta_1}\ln(T_{i-1}\theta_2)}{\exp\left(-(T_{i-1}\theta_2)^{\theta_1}\right)-\exp\left(-(T_i\theta_2)^{\theta_1}\right)}=0;$$

$$\frac{\partial l}{\partial\theta_2}=\sum_{i=1}^{N}N_i\frac{\exp\left(-(T_i\theta_2)^{\theta_1}\right)T_i^{\theta_1}\theta_1\theta_2^{\theta_1-1}-\exp\left(-(T_{i-1}\theta_2)^{\theta_1}\right)T_{i-1}^{\theta_1}\theta_1\theta_2^{\theta_1-1}}{\exp\left(-(T_{i-1}\theta_2)^{\theta_1}\right)-\exp\left(-(T_i\theta_2)^{\theta_1}\right)}=0.$$

Гамма-распределение

Плотность и функция распределения для гамма-закона записываются в виде

$$f(t,\lambda,\alpha)=\frac{\lambda^{\alpha}t^{\alpha-1}\exp(-\lambda t)}{\Gamma(\alpha)};\quad F(t,\lambda,\alpha)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^t x^{\alpha-1}\exp(-\lambda x)dx.$$

Введем обозначения $\theta_1=\alpha$; $\theta_2=\lambda$ и приведем результаты вычисления оценок параметров.

1. Для случая полных наработок:

$$L=\left(\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\right)^k\prod_{i=1}^{k}T_i^{\theta_1-1}\exp\left(-\theta_2\sum_{i=1}^{k}T_i\right);$$

$$l=k\theta_1\ln\theta_2-k\ln\Gamma(\theta_1)+\left(\sum_{i=1}^{k}\ln T_i\right)(\theta_1-1)-\theta_2\sum_{i=1}^{k}T_i;$$

$$\frac{\partial l}{\partial\theta_1}=k\ln\theta_2-\frac{k\Gamma'(\theta_1)}{\Gamma(\theta_1)}+\sum_{i=1}^{k}\ln T_i=0;\quad \frac{\partial l}{\partial\theta_2}=\frac{k\theta_1}{\theta_2}-\sum_{i=1}^{k}T_i=0.$$

2. Для выборок, содержащих полные и цензурированные справа наработки:

$$L=\left(\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\right)^k\prod_{i=1}^{k}T_i^{\theta_1-1}\exp\left(-\theta_2\sum_{i=1}^{k}T_i\right)\prod_{j=1}^{\nu}\left(1-\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\int_0^{T_j'}x^{\theta_1-1}\exp(-\theta_2 x)dx\right);$$

$$l=k\theta_1\ln\theta_2-k\ln\Gamma(\theta_1)+(\theta_1-1)\sum_{i=1}^{k}\ln T_i-\theta_2\sum_{i=1}^{k}T_i+\sum_{j=1}^{\nu}\ln\left[1-\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\int_0^{T_j'}x^{\theta_1-1}\exp(-\theta_2 x)dx\right].$$

Введем следующие обозначения:

$$A(x)=x^{\theta_1-1}\exp(-\theta_2 x);$$
$$B(x)=x^{\theta_1-1}\ln x\exp(-\theta_2 x);$$
$$C(x)=\left(-x^{\theta_1}\right)\exp(-\theta_2 x)$$

и получим

$$\frac{\partial l}{\partial\theta_1}=k\ln\theta_2+\sum_{i=1}^{k}\ln T_i-k\frac{\Gamma'(\theta_1)}{\Gamma(\theta_1)}-$$

$$-\sum_{j=1}^{\nu}\frac{\Gamma(\theta_1)[\theta_2^{\theta_1}\ln\theta_2\int_0^{T_j'}A(x)dx+\theta_2^{\theta_1}\int_0^{T_j'}B(x)dx]-\Gamma'(\theta_1)\theta_2^{\theta_1}\int_0^{T_j'}A(x)dx}{\Gamma^2(\theta_1)\left[1-\frac{\theta_2^{\theta_1}}{\tilde{A}(\theta_1)}\int_0^{T_j'}x^{\theta_1-1}\exp(-\theta_2 x)dx\right]}=0;$$

$$\frac{\partial l}{\partial\theta_2}=\frac{k\theta_1}{\theta_2}-\sum_{i=1}^{k}T_i-\sum_{j=1}^{\nu}\frac{\theta_2^{\theta_1}\theta_1\int_0^{T_j'}A(x)dx+\theta_2^{\theta_1}\int_0^{T_j'}C(x)dx}{\left[1-\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\int_0^{T_j'}A(x)dx\right]\Gamma(\theta_1)}=0.$$

3. Для выборок, содержащих полные и цензурированные слева наработки:

$$L=\left(\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\right)^{k+\nu}\prod_{i=1}^{k}T_i^{\theta_1-1}\exp\left(-\theta_2\sum_{i=1}^{k}T_i\right)\prod_{j=1}^{\nu}\int_0^{T_j'}x^{\theta_1-1}\exp(-\theta_2 x)dx;$$

$$l=(k+\nu)\theta_1\ln\theta_2-(k+\nu)\ln\Gamma(\theta_1)+(\theta_1-1)\sum_{i=1}^{k}\ln T_i-\theta_2\sum_{i=1}^{k}T_i+\sum_{j=1}^{\nu}\ln\int_0^{T_j'}A(x)dx;$$

$$\frac{\partial l}{\partial\theta_1}=(k+\nu)\ln\theta_2-\frac{k+\nu}{\Gamma(\theta_1)}\Gamma'(\theta_1)+\sum_{i=1}^{k}\ln T_i+\sum_{j=1}^{\nu}\frac{\int_0^{T_j'}B(x)dx}{\int_0^{T_j'}A(x)dx}=0;$$

$$\frac{\partial l}{\partial \theta_2} = \frac{(k+\nu)\theta_1}{\theta_2} - \sum_{i=1}^{k} T_i + \sum_{j=1}^{\nu} \frac{\int_0^{T_j'} C(x)dx}{\int_0^{T_j'} A(x)dx} = 0.$$

4. Для группированных данных:

$$L = \prod_{i=1}^{N} \left[\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} \left(\int_0^{T_i} A(x)dx - \int_0^{T_{i-1}} A(x)dx \right) \right]^{N_i};$$

$$l = \sum_{i=1}^{N} N_i \left[\theta_1 \ln \theta_2 - \ln \Gamma(\theta_1) \right] + \sum_{i=1}^{N} N_i \ln \left(\int_0^{T_i} A(x)dx - \int_0^{T_{i-1}} A(x)dx \right);$$

$$\frac{\partial l}{\partial \theta_1} = \sum_{i=1}^{N} N_i \ln \theta_2 - \frac{\sum_{i=1}^{N} N_i}{\Gamma(\theta_1)} \Gamma'(\theta_1) + \sum_{i=1}^{N} N_i \frac{\int_0^{T_i} B(x)dx - \int_0^{T_{i-1}} B(x)dx}{\int_0^{T_i} A(x)dx - \int_0^{T_{i-1}} A(x)dx} = 0;$$

$$\frac{\partial l}{\partial \theta_2} = \frac{\sum_{i=1}^{N} N_i \theta_1}{\theta_2} + \sum_{i=1}^{N} N_i \frac{\int_0^{T_i} C(x)dx - \int_0^{T_{i-1}} C(x)dx}{\int_0^{T_i} A(x)dx - \int_0^{T_{i-1}} A(x)dx} = 0.$$

Подводя итог можно сказать, что для большинства законов распределения и схем формирования информации получаются результаты, не имеющие аналитического выражения. Их вычисление может быть проведено только с применением численных методов.

Глава 8

ПОВЫШЕНИЕ ДОСТОВЕРНОСТИ ОЦЕНИВАНИЯ ЗА СЧЕТ ИСПОЛЬЗОВАНИЯ АПРИОРНОЙ ИНФОРМАЦИИ

8.1. Формулировка теоремы Байеса для событий

При построении моделей систем важным вопросом является вопрос обеспечения адекватности модели и системы, для которой строится данная модель. От качества модели зависит достоверность результатов анализа, проводимого с помощью модели. Известен тезис системотехников и системных аналитиков: «Что в модель заложили, то и получили на выходе данной модели!». Одним из параметров, обеспечивающих качество модели, является точность задания показателей и характеристик составных частей модели, описывающих характер функционирования этих частей в процессе жизненного цикла системы и используемых в результате расчетов моделей.

Как уже отмечалось, выборки экспериментальных данных о функционировании анализируемых систем и их составных частей присутствуют в весьма ограниченном объеме. Поэтому, для получения оценок показателей объектов анализа с высокой степенью достоверности необходимо использовать дополнительную информацию. В качестве такого вида информации может применяться информация о функционировании объектов-аналогов, различного рода субъективная информация, учитывающая опыт и квалификацию персонала и т.п. Учет такой информации может быть осуществлен с помощью подхода, основанного на применении формулы Байеса.

В настоящее время перед исследователями стоит задача оценивания элементов и системы в целом с анализом точности и достоверности оценивания, построением доверительных интервалов на оценки.

Рассмотрим схему оценивания, согласно которой предполагается, что у исследователя имеется априорная информация об исследуемых показателях объектов-аналогов. Изложим метод, использующий формулу Байеса и позволяющий проводить оценивание показателей элементов на основании текущей (эксплуатационной) информации с учетом результатов наблюдений, полученных на этапе априорных исследований объектов-аналогов. Байесовские методы находят широкое применение при решении задач оценивания показателей сложных систем.

Формула или теорема Байеса – одна из центральных теорем теории вероятностей. В настоящее время область применения этой теоремы чрезвычайно широка. Это и учет априорной информации в задачах оценивания и применение формулы в самообучающихся системах, в системах диагностики, и, наконец, в экспертных системах.

В простейшем случае формула выводится следующим образом. Пусть имеются два зависимых события A и B. По определению условной вероятности наступления события A при условии, что произошло событие B, имеем

$$P(A/B)=\frac{P(AB)}{P(B)}, \tag{8.1}$$

где $P(AB)$ – вероятность совместного наступления событий A и B; $P(B)$ – вероятность события B.

Аналогично можно записать

$$P(B/A)=\frac{P(AB)}{P(A)}. \tag{8.2}$$

Определив из равенства (8.2) $P(AB)$ и поставив данное значение в (8.1), получим простейший вариант формулы Байеса [40]

$$P(A/B)=\frac{P(A)P(B/A)}{P(B)}.$$

Распространим данную формулу на группу несовместных событий A_i, $i=\overline{1,n}$. Пусть событие B может осуществиться с одним и только одним из n несовместных событий A_i, т.е. $B=\sum_{i=1}^{n}BA_i$. Множество A образует полную группу событий, т.е. $\sum_{i=1}^{n}P(A_i)=1$, BA_i и BA_j – попарно несовместимые события для любых $i=\overline{1,n}$, $j=\overline{1,n}$ и $i\neq j$. Тогда для этих событий можно записать формулу полной вероятности

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B/A_i).$$

Используя эту формулу и записав выражение (8.1), (8.2) для события A, получаем формулу Байеса в виде

$$P(A_i/B)=\frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n}P(A_i)P(B/A_i)}. \tag{8.3}$$

Данная формула в теории вероятностей называется также формулой вероятности гипотез. Формула в виде (8.3) справедлива для событий.

Продемонстрируем возможность применения формулы Байеса в задачах, имеющих непосредственное отношение к задачам системного анализа и связанных с принятием решения. На предприятии (например АЭС) готовятся к проведению реконструкции. Для успешного проведения работ по реконструкции отдельных систем необходимо приобрести некоторое оборудование и поставить его вместо отработавшего свой ресурс. Естественно, что реконструкцию имеет смысл проводить только в том случае, если показатели надежности вновь поставляемого оборудования будут выше показателей надежности устройств, которые собираются заменить. Предприятию предлагают приобрести требуемое оборудование, причем завод-изготовитель утверждает, что показатели надежности находятся на достаточно высоком уровне. Например, утверждается, что вероятность безотказной работы (ВБР) изделия в течение требуемого времени не менее, чем $P_{з.и}$. Однако из опыта эксплуатации аналогичных устройств на других объектах известно, что показатель ВБР меньше, чем утверждает завод-изготовитель, и находится на уровне опытных значений $P_{оп}$. Если надежность оборудования такая, как утверждает завод-изготовитель, то предприятию выгодно приобрести его и провести реконструкцию. Если же надежность такая, как сообщает предприятие, уже эксплуатирующее данное оборудование, то реконструкцию проводить нецелесообразно и, следовательно, покупать оборудование нет необходимости.

Пусть заказчик сомневается как в заявлении завода-изготовителя, так и в информации предприятия, эксплуатирующего устройства. (Если бы заказчик был уверен, что кто-то из них прав, то решение задачи было бы тривиально.) Заказчик может сформулировать свою неуверенность следующим образом: «Вероятность того, что прав завод-изготовитель, равна p_1; вероятность того, что верно заявление предприятия, равна $p_2=1-p_1$». Перед тем, как принять решение о покупке изделий, заказчик намерен провести их испытания. Пусть он провел испытания k изделий в течение времени $T_р$, в результате которых m изделий отказало.

Покажем, как сформулировать решение о приобретении устройств, учитывая информацию завода-изготовителя и предприятия, эксплуатирующего аналогичные устройства. Метод вычисления вероятностей, используемых при принятии решения, дает теорема Байеса. Вероятность того, что прав завод-изготовитель, при условии, что при испытании k устройств в течение времени $T_р$ m из них отказало, равна

$$P(p = P_{з.и} \ / \ \text{отказало } m \text{ из } k \text{ устройств}) =$$

$$= \frac{P(\text{отказало } m \text{ из } k \text{ устройств} \ / \ p = P_{з.и})P(p = P_{з.и})}{P(\text{отказало } m \text{ из } k \text{ устройств})}.$$

Покажем, как определять величины, стоящие в данном выражении:

$$P(\text{отказало } m \text{ из } k \text{ устройств} \ / \ p = P_{з.и}) = C_k^m (1 - P_{з.и})^m P_{з.и}^{(k-m)},$$

где C_k^m – число сочетаний из m по k; $P(p = P_{з.и})$ – первоначальное мнение заказчика о том, что информация завода-изготовителя верна – $P(p = P_{з.и}) = p_1$. Полная вероятность события, состоящего в том, что отказало m устройств из k испытываемых, вычисляется следующим образом:

$$P(\text{отказало } m \text{ из } k \text{ устройств}) =$$

$$= P(\text{отказало } m \text{ из } k \text{ устройств} \ / \ p = P_{з.и})P(p = P_{з.и}) +$$

$$+ P(\text{отказало } m \text{ из } k \text{ устройств} \ / \ p = P_{оп})P(p = P_{оп}).$$

Здесь $P(\text{отказало } m \text{ из } k \text{ устройств} \ / \ p = P_{оп}) = C_k^m (1 - P_{оп})^m P_{оп}^{(k-m)}$;

$P(p = P_{оп}) = p_2 = 1 - p_1$.

Рассмотрим числовой пример решения задачи о целесообразности покупки оборудования. Пусть завод-изготовитель утверждает, что ВБР изделия равна 0,98. Предприятие, имеющее опыт эксплуатации указанных изделий, оценивает ВБР на уровне 0,9. Заказчик считает, что вероятность того, что завод-изготовитель верно оценил надежность оборудования, равна 0,4; вероятность того, что верно заявление предприятия, равна 0,6. Далее предприятие-заказчик производит испытания двух объектов в течение времени $T_р$ и оба объекта за этот период отказывают.

Подсчитаем вероятность события, состоящего в том, что утверждение предприятия, эксплуатирующего аналогичные объекты, верно:

$$P(p = 0{,}9 \ / \ \text{отказало 2 из 2}) = \frac{P(\text{отказало 2 из 2} \ / \ p = 0{,}9)P(p = 0{,}9)}{P(\text{отказало 2 из 2})}.$$

Произведем расчет каждого из сомножителей:

$$P(\text{отказало 2 из 2} \ / \ p = 0{,}9) = (1 - 0{,}9)^2 = 0{,}01; \ P(p = 0{,}9) = 0{,}6;$$

$$P(\text{отказало 2 из 2}) = P(\text{отказало 2 из 2} \ / \ p = 0{,}9)P(p = 0{,}9) +$$

$$+ P(\text{отказало 2 из 2} \ / \ p = 0{,}98)P(p = 0{,}98) =$$

$$= 0{,}01 \cdot 0{,}6 + 0{,}0004 \cdot 0{,}4 = 0{,}006016.$$

Окончательно получаем

$$P(p = 0{,}9 \ / \ \text{отказало 2 из 2}) = \frac{0{,}006}{0{,}006016} \approx 0{,}97.$$

Таким образом, после проведения независимых испытаний можно принять решение о нецелесообразности закупки оборудования, так как с вероятностью 0,97 верны выводы предприятия, эксплуатирующего указанные объекты.

Другой возможный случай. Оба объекта успешно проходят испытания. Обратимся опять к формуле Байеса и посмотрим, как в этом случае изменится вероятность того, что оценки предприятия верны:

$$P(p = 0{,}9 \ / \ \text{отказов нет}) = \frac{P(\text{отказов нет} \ / \ p = 0{,}9)P(p = 0{,}9)}{P(\text{отказов нет})},$$

где

$$P(\text{отказов нет} \ / \ p = 0{,}9) = 0{,}9^2 = 0{,}81; \ P(p = 0{,}9) = 0{,}6;$$

$$P(\text{отказов нет}) = 0{,}81 \cdot 0{,}6 + 0{,}9604 \cdot 0{,}4 \approx 0{,}87.$$

Окончательно получаем

$$P(p = 0{,}9 \ / \ \text{отказов нет}) = \frac{0{,}486}{0{,}87} \approx 0{,}56.$$

После проведения независимой серии испытаний, закончившихся успешно, вероятность того, что выводы предприятия, имеющего опыт эксплуатации, верны, несколько снизилась, а соответственно вероятность того, что верны утверждения завода-изготовителя, несколько возросла и стала равна $p_1 = 1 - p_2 = 0{,}44$. Однако значения данных вероятностей таковы, что по ним трудно принять решение и отдать предпочтение выводам какого-либо партнера. Наилучшее решение в данном случае будет заключаться в продолжении испытаний объектов (если это экономически целесообразно).

8.2. Теорема Байеса для непрерывных случайных величин

Изложенный вариант теоремы Байеса предполагает простейшую схему оценивания показателей сложных систем. В данной схеме исследователь оперирует с точечными оценками показателей, не затрагивая вопросы точности их определения, доверия к полученному результату.

Рассмотрим более общий вариант теоремы Байеса, позволяющий применять ее для оценивания характеристик, определяемых по результатам наблюдения за реализациями непрерывных случайных величин. Введем ряд предположений.

1. Производятся наблюдения за непрерывной случайной величиной $t \in T$, имеющей распределение $F(\theta, t)$. Функция $F(\theta, t)$ дифференцируема, т.е. существует плотность распределения случайной величины $t \sim f(\theta, t)$.

2. Параметр $\theta \in \Theta$ – число или вектор с заданной априорной плотностью распределения $h(\theta)$.

3. Оценка d параметра θ задана на множестве возможных решений D.

4. Функция потерь $u(\theta, d)$ определена на (Θ, D) и выражает потери, обусловленные ошибочным решением.

В общем виде функция потерь выглядит следующим образом:

$$u(\theta,d) = \lambda(\theta) W(|\, d-\theta\,|),$$

где $W(0)=0$; $W(x)$ – монотонно возрастающая функция, $x > 0$; $\lambda(\theta)$ – положительно определенная конечная функция.

Применяя формулу Байеса, можно записать выражение для апостериорной плотности распределения параметра θ при условии, что в результате проведения опыта реализовалась случайная величина T [39]:

$$h(\theta/\{T\}) = \frac{f(\{T\},\theta)h(\theta)}{\int\limits_{\Theta} f(\{T\},\tau)h(\tau)d\tau}, \qquad (8.4)$$

где $h(\theta)$ – априорная плотность распределения искомого параметра θ, $f(\{T, \theta\})$ – совместная плотность распределения величин T и θ.

Для определения оценки d параметра θ введем апостериорную функцию риска:

$$R(\theta,d) = \int\limits_{\Theta} \lambda(\theta)\ W(|\, d(t)-\theta\,|)\ h(\theta/\{T\})\ d\theta\,.$$

При квадратичной функции потерь $W(|\, d(t)-\theta\,|) = (d(t)-\theta)^2$ и $\lambda(\theta)=1$ функция риска примет вид

$$R(\theta,d) = \int\limits_{\Theta} (d(t)-\theta)^2 h(\theta/\{T\})d\theta\,. \qquad (8.5)$$

Минимизируя данную функцию риска, определяем оценку параметра θ. Возьмем производную от функции (8.5) и приравняем ее нулю:

$$d(t)\int\limits_{\Theta} h(\theta/\{T\})d\theta - \int\limits_{\Theta} \theta h(\theta/\{T\})d\theta = 0\,.$$

Значение $\int\limits_{\Theta} h(\theta/\{T\})d\theta$ равно 1, так как представляет собой интеграл от плотности распределения по всей области определения параметра θ, тогда можно записать

$$d(t) = \int\limits_{\Theta} \theta h(\theta/\{T\})d\theta\,. \qquad (8.6)$$

Выражение для дисперсии оцениваемого параметра имеет вид

$$\sigma_\theta^2 = \int\limits_{\Theta} \theta^2 h(\theta/\{T\})d\theta - d^2(t)\,. \qquad (8.7)$$

В данных рассуждениях используется понятие априорной плотности распределения оцениваемого параметра. В общем случае обоснование вида априорной плотности является сложной задачей. При традиционном байесовском подходе априорная плотность распределения формируется исходя из опыта и научной интуиции исследователя. Сформированные таким образом суждения получили название субъективной вероятности [40, 41].

Схема проведения исследований при байесовском подходе следующая. Исследования проводит высококвалифицированный в данной области системных исследований специалист. До проведения испытаний у него сформировалось определенное мнение относительно предмета исследования. Исследователь проводит серию испытаний и в своем окончательном выводе учитывает как априорные суждения, сформулированные до испытаний, так и результаты проведенных экспериментов.

Априорная информация может быть сформирована на основании анализа коллективного мнения группы экспертов. При этом группа экспертов формируется из числа высококвалифицированных специалистов в области, к которой относятся организуемые исследования. В данном случае априорная информация будет более объективна, так как представляет собой результат обработки коллективного мнения специалистов.

Приведем пример оценивания показателей надежности элементов применительно к описанной схеме. Пусть в результате длительного опыта эксплуатации элемента в составе изделия у специалиста, обслуживающего данные изделия, имеется мнение, что надежность изделия достаточно высока.

Например, свое мнение он может выразить следующим образом: вероятность безотказной работы элемента за время его эксплуатации $T_р$ (от одной плановой профилактики до другой) не менее некоторой величины $p_и$. Или же мнение может состоять в том, что ВБР за время $T_р$ лежит в интервале ($p_и$, $p_в$). На указанном интервале значений ВБР ($p_и$, $p_в$) специалист не может выделить наиболее вероятное значение, т.е. можно сказать, что в данном интервале все значения p равновероятны. Тогда априорная плотность распределения

$$h(p)=\begin{cases}0 \text{ при } p<p_н;\\ \dfrac{1}{p_в-p_н} \text{ при } p_н\le p\le p_в;\\ 0 \text{ при } p>p_в.\end{cases}$$

Пусть далее проводятся испытания по схеме Бернулли, в результате которых из k испытываемых элементов за время $T_р$ отказывает m объектов. Вероятность события, происшедшего при испытаниях

$$f(p/m,k)=C_k^m p^{k-m}(1-p)^m,$$

где $C_k^m=\dfrac{k!}{m!(k-m)!}$ – число сочетаний из m по k.

Подставляя данное выражение и выражение для априорной плотности распределения в формулу Байеса, получаем

$$h(\hat{p}/p)=\frac{p^{k-m}(1-p)^m}{\int_{p_н}^{p_в} p^{k-m}(1-p)^m\,dp}.$$

Байесовская оценка ВБР

$$\hat{p}_0=\frac{\int_{p_н}^{p_в} p^{k-m+1}(1-p)^m\,dp}{\int_{p_н}^{p_в} p^{k-m}(1-p)^m\,dp}. \tag{8.8}$$

Точность байесовской оценки

$$\sigma_p^2=\frac{\int_{p_н}^{p_в} p^{k-m+2}(1-p)^m\,dp}{\int_{p_н}^{p_в} p^{k-m}(1-p)^m\,dp}-\hat{p}_0^2. \tag{8.9}$$

Вычисление интегралов, входящих в (8.8) и (8.9), после подстановки численных значений k и m не вызывает особых затруднений.

Излагаемые до настоящего момента байесовские процедуры касались исследования методов совместного учета информации, полученной в результате текущих и априорных наблюдений. Попытаемся сформулировать некоторые способы формирования соответствующих плотностей, входящих в формулу Байеса. Отметим, что более правильно для величин, входящих в формулу Байеса, применять термин «обобщенная вероятностная плотность». Это понятие включает в себя как понятие плотности распределения вероятностей, используемое в записи (8.4), так и понятие функции вероятностей, используемое при формулировке теоремы Байеса для дискретных случайных событий (8.3) [41]. Естественно, что наиболее общие и интересные задачи оценивания связаны с применением теоремы Байеса для непрерывных случайных величин.

Приведем методику формирования функции правдоподобия, в которой концентрируется текущая информация. Пусть на этапе текущих исследований зафиксирована выборка $T_1, T_2,\dots, T_k$, где T_i – независимые случайные величины. Каждая величина T_i распределена согласно некоторой плотности $f(\theta, t)$. В этом случае совместная плотность распределения величин $\{\theta;\ T_1, T_2,\dots, T_k\}$ будет выражаться следующим образом [39]:

$$f(\theta;T_1,\dots,T_k)=\prod_{i=1}^{k} f(\theta,t).$$

Данное выражение называется функцией правдоподобия (см. гл.7).

Рассмотрим несколько конкретных примеров

1. Пусть T_i – элементы выборки наблюдаемой случайной величины, распределенной по экспоненциальному закону с параметром λ. Тогда функция правдоподобия запишется следующим образом:

$$f(\lambda,\{T_i\}) = \prod_{i=1}^{k} \lambda \exp(-\lambda T_i) = \lambda^k \exp\left(-\lambda \sum_{i=1}^{k} T_i\right).$$

Выражая сумму, стоящую в показателе экспоненты, через среднее арифметическое, получаем

$$f(\lambda,\{T_i\}) = \lambda^k \exp(-k\lambda m_т),$$

где $m_т = \dfrac{\sum_{i=1}^{k} T_i}{k}$.

2. Пусть T_i – элементы выборки случайной величины, распределенной по нормальному закону с параметрами m и σ. Тогда функция правдоподобия

$$f(m,\sigma,\{T_i\}) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(T_i - m)^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{k/2}} \exp\left(-\frac{\sum_{i=1}^{k}(T_i - m)^2}{2\sigma^2}\right).$$

3. T_i – элементы выборки случайной величины, подчиняющиеся гамма-распределению

$$f(\lambda,a,t) = \frac{\lambda^\alpha t^\alpha}{\Gamma(a)} \exp(-\lambda t).$$

В этом случае функция правдоподобия

$$f(\lambda,a,\{T_i\}) = \frac{\lambda^{k\alpha}}{[\Gamma(a)]^k} \exp\left(-\sum_{i=1}^{k} T_i \lambda\right) \prod_{i=1}^{k} T_i^{\alpha-1}.$$

Здесь λ и α – параметры масштаба и формы гамма-распределения; $\Gamma(\alpha)$ – гамма-функция.

Таким образом, не возникает принципиальных трудностей формирования совместной плотности распределения, содержащей текущую информацию об исследуемом процессе или объекте. Методы формирования априорной плотности будут рассмотрены в следующих параграфах.

8.3. Вычисление апостериорной плотности при последовательном накоплении информации

Рассмотрим схему оценивания, когда наблюдения за функционированием объектов проводятся в несколько этапов, и после завершения каждого этапа необходимо сделать заключение о достигнутом значении исследуемого показателя сложной системы. Данная ситуация весьма характерна для эксплуатации оборудования ЯЭУ. На практике ведутся наблюдения за функционированием оборудования ЯЭУ и по истечении определенного периода (например, каждый год) информация поступает в вышестоящую организацию для обработки. Вновь поступающую информацию необходимо анализировать для определения значения достигнутых показателей надежности элементов, систем, устройств ЯЭУ. Однако при проведении расчетов необходимо иметь в виду, что данные элементы уже эксплуатировались в составе ЯЭУ и имеется информация об их поведении за прошлые периоды функционирования. Учет вновь поступающей информации наряду с уже имеющейся существенно повышает достоверность и точность оценок исследуемых характеристик.

Математическая постановка задачи, решаемой в данном параграфе, формулируется следующим образом.

Пусть исследователь производит испытания объектов в несколько этапов. В результате первой серии испытаний получена статистика $\{T_i,\ i=\overline{1,\ k}\}$. Априорная плотность распределения параметра θ имеет вид $h(\theta)$. Апостериорная плотность распределения параметра θ определяется соотношением (8.4). Во второй серии испытаний исследователь зафиксировал статистику $\{T_j,\ j=\overline{1,\ l}\}$. Необходимо получить оценку характеристик надежности, учитывая как априорную информацию, так и результаты последовательно проводящихся серий испытаний.

Покажем, как следует подходить к решению задачи в такой постановке. После проведения второй серии испытаний плотность распределения (8.4) можно рассматривать как априорную плотность по отношению к новой статистике $\{T_j,\ j=\overline{1,\ l}\}$. После применения формулы Байеса получим новую апостериорную плотность, учитывающую обе серии наблюдений:

$$h_{апост}(\theta/\{T_i,\ i=\overline{1,\ k}\};\{T_j,\ j=\overline{1,\ l}\}) = \frac{f_1(\{T_j\},\theta)\ h(\theta/\{T_i\})}{\int_\Theta f_1(\{T_j\},\theta)\ h(\theta/\{T_i\})d\theta}$$

или, подставив вместо $h(\theta/\{T_i\})$ выражение (8.4), получим

$$h_{\text{апост}}(\theta/\{T_i\};\{T_j\}) = \frac{f_1(\{T_i\},\theta)\ f_2(\{T_j\},\theta)\ h(\theta)}{\int\limits_\Theta f_1(\{T_i\},\theta)\ f_2(\{T_j\},\theta)\ h(\theta)d\theta}.$$

Если проводится несколько серий испытаний (например m), то после m-й серии наблюдений выражение для апостериорной плотности будет иметь вид

$$h_{\text{апост}}(\theta/\{T_{i1}\};\{T_{i2}\};...;\{T_{im}\}) = \frac{f_1(\{T_{i1}\},\theta)...f_m(\{T_{im}\},\theta)\ h(\theta)}{\int\limits_\Theta f_1(\{T_{i1}\},\theta)...f_m(\{T_{im}\},\theta)\ h(\theta)d\theta}.$$

Это означает, что если наблюдения проводятся в несколько этапов, то апостериорное распределение можно вычислять на каждом этапе, беря в качестве априорного распределения для последующего этапа апостериорное распределение, полученное на предыдущем, т.е. оценивание можно проводить последовательно. Далее можно сделать следующее заключение: если апостериорное распределение параметра θ вычисляется в два приема и более, то окончательный результат не зависит от того, какая выборка получена сначала.

Процесс оценивания параметра θ теперь может быть описан в следующем виде. В каждый заданный момент времени исследователь располагает вероятностным распределением параметра θ. С течением времени к исследователю поступает информация о θ в виде статистики $\{T\}$, и он использует эту информацию для корректировки распределения θ. В те моменты времени, когда исследователю необходимо оценить θ, он применяет процедуру (8.6), (8.7) и получает решение, оптимальное относительно распределения θ в данный текущий момент.

8.4. Байесовское оценивание и несобственная плотность распределения

При изложении байесовской процедуры оценивания до настоящего времени предполагалось, что у системного аналитика до проведения исследований над объектами имеется некоторая априорная информация. Процедура оценивания состоит в том, что на основании априорной информации формируется некоторая априорная оценка искомого параметра и затем по мере поступления информации эта оценка уточняется.

Однако при решении задач системного анализа нередки случаи, когда у исследователя кроме информации, полученной в результате текущих наблюдений, никаких других сведений нет. Известен подход, дающий возможность применять процедуру байесовского оценивания в ситуации полного отсутствия априорной информации, основанный на использовании несобственной плотности распределения.

Вначале изложим пример из области оценивания характеристик надежности. Итак, пусть требуется оценить показатель ВБР. Предположим, что имеется априорная информация о показателе надежности изделия p, согласно которой данный параметр имеет β-распределение:

$$h(p) = C_{\alpha+\beta-1}^{\alpha-1} p^{\alpha-1}(1-p)^{\beta-1}.$$

Текущая информация представлена в виде результатов испытаний группы однотипных объектов, в ходе которых из k испытываемых изделий m объектов отказало.

Результат испытаний можно записать в виде распределения Бернулли:

$$f(p,k,m) = C_k^m p^{k-m}(1-p)^m. \tag{8.10}$$

Апостериорное распределение ВБР изделия запишется в виде

$$h_{\text{апост}}(p/\hat{p}) = \frac{p^{\alpha+k-m-1}\ (1-p)^{\beta+m-1}}{\int\limits_0^1 p^{\alpha+k-m-1}\ (1-p)^{\beta+m-1}\ dp} = C(\alpha,\beta,m,k)\ p^{\alpha+k-m-1}\ (1-p)^{\beta+m-1},$$

где $C(\alpha, \beta, m, k)$ – постоянная, зависящая от параметров α, β, m, k.

Таким образом, вновь получили β-распределение с параметрами $(\alpha + k - m - 1)$ и $(\beta + m - 1)$. В выражении (8.10), описывающем результаты текущих исследований, параметры распределения представляют собой m – количество отказавших изделий и $(k - m)$ – количество изделий, испытания которых прошли успешно. По аналогии с этим можно интерпретировать априорное распределение как эквивалент наблюдений выборки объема $\alpha + \beta$, в которой β элементов отказало за время проведения исследований и α элементов прошло испытания успешно. После того, как провели такую аналогию, естественно при полном отсутствии априорной информации логично положить значения α и β равными нулю, т.е. априорное распределение должно быть эквивалентно рассмотрению выборки объема 0, в которой 0 элементов отказало.

При этом апостериорное распределение примет вид

$$h_{\text{апост}}(p/\alpha=0,\beta=0) = C(k,m)\ p^{k-m-1}\ (1-p)^{m-1}.$$

Оценки параметра ВБР и дисперсии оценки ВБР будут равны

$$\hat{p} = (k-m)/k;\ \sigma_p^2 = \frac{(k-m)m}{k^2(k+1)} = \frac{(1-\hat{p})}{k+1},$$

т.е. получили результат, состоящий в том, что апостериорные среднее и дисперсия оказываются зависящими только от текущей информации.

Способ, с помощью которого произвели оценивание показателя ВБР при полном отсутствии априорной информации, формально дает приемлемый результат. Единственное затруднение, которое возникает при этом, заключается в том, что, если α и β приравнять нулю, то получается априорная плотность, не удовлетворяющая условиям нормировки. А именно интеграл от этой плотности по всей области определения может оказаться равным бесконечности, независимо от выбора масштабного множителя.

Функцию с параметрами α и β, одновременно обращающимися в нуль, называют несобственной функцией.

Приведем другой пример оценивания показателей надежности. Необходимо оценить наработку на отказ механических элементов. Известно, что наработки объектов с механическими компонентами хорошо описываются гауссовским законом распределения. Пусть имеется априорная информация о параметре наработки на отказ в виде априорной плотности распределения

$$h(\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\theta-m_{\text{а}})^2}{2\sigma^2}\right),$$

где σ^2 – дисперсия априорной оценки наработки на отказ $m_{\text{а}}$.

Текущая информация представлена в виде наработок объектов до отказа $T_1, T_2, \ldots, T_k$. Функция правдоподобия в данном случае

$$f(\theta,\{T_i\}) = \frac{1}{\sqrt{2\pi}S}\exp\left(-\frac{(m_{\text{т}}-\theta)^2}{2S^2}\right),$$

где S^2 – дисперсия наработки на отказ, определенная на основании текущей информации, а именно

$$m_{\text{т}} = \frac{\sum_{i=1}^{k} T_i}{k};\quad S^2 = \frac{\sum_{i=1}^{k}(T_i - m_{\text{т}})}{(k-1)k}.$$

Апостериорная плотность распределения будет определяться следующим образом:

$$h_{\text{апост}}(\theta/\{T_i\}) = \frac{\exp\left(\left[\frac{(m_{\text{т}}-\theta)^2}{2S^2} + \frac{(\theta-m_{\text{а}})^2}{2\sigma^2}\right]\right)}{2\pi S\sigma\int_{-\infty}^{\infty} f(\theta,\{T_i\})h(\theta)d\theta}.$$

Преобразуем показатель степени у экспоненты. Для этого прибавим и вычтем величину $\dfrac{(m_{\text{а}}+m_{\text{т}})^2}{2S^2}$.

Перегруппировывая члены, получаем

$$\frac{(m_{\text{т}}-\theta)^2}{2S^2} + \frac{(\theta-m_{\text{а}})^2}{2\sigma^2} = \frac{S^2+\sigma^2}{2S^2\sigma^2}\left[\theta^2 - \frac{2S^2m_{\text{т}}\theta + 2\sigma^2m_{\text{а}}\theta}{S^2+\sigma^2} + \right.$$

$$\left. + \left(\frac{S^2m_{\text{а}}+\sigma^2m_{\text{т}}}{S^2+\sigma^2}\right)^2\right] + b = \frac{S^2+\sigma^2}{2S^2\sigma^2}\left(\theta - \frac{S^2m_{\text{а}}+\sigma^2m_{\text{т}}}{S^2+\sigma^2}\right)^2 + b.$$

Здесь $b = \dfrac{1}{2S^2}\left[\sigma^2m_{\text{т}} + S^2m_{\text{а}} - \dfrac{(S^2m_{\text{а}}+\sigma^2m_{\text{т}})^2}{S^2+\sigma^2}\right]$.

Апостериорное распределение для θ можно записать теперь в виде

$$h_{\text{апост}}(\theta/\{T_i\}) = A\ \exp\left\{-\left[\theta - \frac{S^2m_{\text{а}}+\sigma^2m_{\text{т}}}{S^2+\sigma^2}\right]\Big/\frac{2S^2\sigma^2}{S^2+\sigma^2}\right\}.$$

Так как, согласно условию нормировки, интеграл по области определения параметра θ должен равняться 1, то

$$A = \frac{1}{\sqrt{2\pi S^2\sigma^2/(S^2+\sigma^2)}}.$$

Отсюда видно, что апостериорная оценка наработки на отказ будет определяться по формуле

$$\hat{\theta} = \frac{S^2m_{\text{а}}+\sigma^2m_{\text{т}}}{S^2+\sigma^2}.$$

Дисперсия оценки определяется из выражения

$$D(\hat{\theta}) = \frac{S^2\sigma^2}{S^2+\sigma^2}.$$

Естественно положить, что чем меньше у наблюдателя сведений об априорной оценке $m_а$, тем больше дисперсия σ^2, так как она характеризует степень неопределенности в оценивании данного параметра. Отсутствие априорной информации равнозначно абсолютной неопределенности в априорной оценке $m_а$. Устремив σ^2 к бесконечности, получим, что апостериорное распределение преобразуется к виду

$$h_{\text{апост}}(\theta/\{T_i\})_{\sigma^2\to\infty} = \frac{1}{\sqrt{2\pi}S}\exp\left(-\frac{(\theta-m_т)^2}{2S^2}\right).$$

Иными словами, апостериорное распределение зависит исключительно от информации, полученной на этапе текущих исследований.

В данном случае априорная плотность распределения с бесконечной дисперсией так же, как и в примере с β-распределением, является несобственной плотностью распределения.

С вводом понятия несобственной плотности распределения получен формальный метод оценивания вероятностных характеристик сложных систем в случае полного отсутствия априорной информации. В основе метода, как и прежде, лежит байесовский подход.

Таким образом, при отсутствии априорной информации на первом этапе оценивания можно воспользоваться несобственной функцией и получить оценки искомых параметров, а затем, воспользовавшись моделью последовательного накопления информации, изложенной в п. 8.3, получать все более точные значения оцениваемого показателя. Данный подход особенно актуален в условиях автоматизированного анализа характеристик надежности, когда большое значение имеет единообразие методик расчета.

8.5. Достаточные статистики

В ряде задач системного анализа исследователю для проведения работы не обязательно хранить всю информацию о функционировании объектов, т.е. не нужно при расчетах иметь выборку о реализациях наблюдаемой случайной величины, на основании которой оценивают параметры системы для включения в дальнейшем их в модель.

Объем требуемой для расчетов информации можно существенно сократить, если вычислить заранее значения некоторого количества числовых характеристик. При этом необходимо убедиться, что рассчитанные значения характеристик содержат всю информацию, имевшуюся в первоначальных данных.

Рассмотрим следующую модель оценивания. Пусть имеется случайная величина или случайный вектор **T**, который принимает значения $T_1, T_2, \ldots, T_n$. Требуется оценить некоторый вектор $\boldsymbol{\theta}$. При этом предполагается, что оценивание параметра θ производится по наблюдениям $T_1, T_2, \ldots, T_n$. Случайная величина T и параметр θ связаны условной плотностью распределения $f(T/\theta)$ при $\theta = \hat{\theta}$.

Введем статистику $M(T)$, которая является функцией наблюдаемой случайной величины.

Приведем байесовское определение достаточной статистики. Согласно [41] статистику M называют достаточной, если при любом априорном распределении параметра θ его апостериорное распределение зависит от значения T только через $M(T)$. Статистику с такими свойствами называют достаточной потому, что для вычисления апостериорного распределения θ, исходя из любого априорного распределения, исследователю достаточно знать лишь значение $M(T)$. При этом нет необходимости сохранять значения самого случайного вектора **T**, который может иметь большую размерность.

На практике это обстоятельство является важным при проведении автоматизированных расчетов, так как использование вместо массива случайных величин **T** достаточных статистик $M(T)$ резко сокращает объем требуемой памяти ЭВМ.

Приведем теорему, которая дает простой способ распознавания достаточных статистик.

Теорема. Статистика M достаточна для семейства плотностей распределения $f(T/\theta)$ тогда и только тогда, когда функцию $f(T/\theta)$ можно представить в виде произведения следующим образом:

$$f(T/\theta) = u(T)\nu[M(T),\theta]$$

для всех $T \in \Im$ и $\theta \in \Theta$.

Здесь функция u положительна и не зависит от θ; функция ν неотрицательна и зависит от T только через $M(T)$.

Доказательство теоремы приведено в [41].

Пример 1. Пусть $T_1, T_2, \ldots, T_n$ – выборка, подчиняющаяся нормальному распределению с неизвестными значениями математического ожидания m и среднего квадратического отклонения σ.

Совместная плотность распределения $f_n(t;\, m, \sigma^2)$ выборки $T_1, T_2, \ldots, T_n$ задается следующим образом:

$$f_n(T_1, T_2, ..., T_n; m, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(T_i - m)^2\right). \quad (8.11)$$

Введем обозначение $\hat{m} = \frac{1}{n}\sum_{i=1}^{n} T_i$.

Произведем некоторые преобразования:

$$\sum_{i=1}^{n}(T_i - m)^2 = \sum_{i=1}^{n}(T_i - \hat{m})^2 + n(\hat{m} - m)^2,$$

и получим, что плотность распределения (8.11) зависит от результатов наблюдений $T_1, T_2, \ldots, T_n$ только через величины $\hat{m}$ и $\sum_{i=1}^{n}(T_i - \hat{m})^2$. Таким образом, для данной плотности распределения двумерная векторная статистика

$$M(T_1, T_2, ..., T_n) = \left[\hat{m}, \sum_{i=1}^{n}(T_i - \hat{m})^2\right]$$

будет являться достаточной статистикой.

Пример 2. В теории надежности широкое распространение при расчетах получил экспоненциальный закон распределения. Пусть $T_1, T_2, \ldots, T_n$ – выборка наработок до отказа, подчиняющаяся экспоненциальному закону распределения с неизвестным параметром λ. Совместная плотность распределения $f_n(t;\, \lambda)$ выборки $T_1, T_2, \ldots, T_n$ задается формулой

$$f_n(T_1, T_2, ..., T_n; \lambda) = \lambda^n \exp\left(-\lambda\sum_{i=1}^{n} T_i\right).$$

Введем обозначение $\hat{m} = \frac{1}{n}\sum_{i=1}^{n} T_i$.

Откуда следует

$$f_n(T_1, T_2, ..., T_n; \lambda) = \lambda^n \exp(-\lambda \hat{m} n).$$

В данном выражении совместная плотность зависит только от величины $\hat{m}$, которая является для экспоненциального закона распределения достаточной статистикой.

Из примеров 1 и 2 видно, что имеется достаточная статистика, которую можно представить одной или двумя скалярными функциями от результатов наблюдения $\{T_1, T_2, \ldots, T_n\}$. Независимо от объема выборки наблюдаемой случайной величины размерность достаточной статистики одна и та же. Обработка результатов наблюдений и анализ характеристик сложной системы существенно упрощаются, если наблюдения описываются функцией распределения, обладающей достаточной статистикой фиксированной размерности.

8.6. Сопряженные распределения

Будем считать, что выборку случайных величин можно описать функцией распределения, параметры которой обладают свойствами достаточной статистики. Для рассматриваемого случая сформулирован вывод, согласно которому объем выборки не влияет на размерность достаточной статистики. Другими словами, если для выборки $\{T_1, T_2, \ldots, T_n\}$ получена статистика $M(T)$, то при увеличении объема наблюдений свойства достаточной статистики не изменяются; изменения объема приведут только к изменению абсолютного значения величины $M(T)$. Распределения, оценки параметров которых выражаются через достаточные статистики, обладают следующим свойством. Если априорное распределение параметра θ принадлежит некоторому семейству распределений, то при любых значениях наблюдений в выборке и при любом объеме выборки апостериорное распределение параметра θ должно также принадлежать этому семейству. Семейство распределений с этим свойством называют сопряженным семейством распределений ввиду особой связи, которая должна существовать между семейством распределений параметра θ и семейством распределений наблюдений t. Таким образом, когда существует достаточная статистика фиксированной размерности, у исследователя, занимающегося анализом надежности, появляется возможность работать только с априорными и апостериорными распределениями из сравнительно узкого семейства сопряженных распределений.

Проанализируем примеры, изложенные в п. 8.4.

1. Результат испытаний группы однотипных объектов, в ходе которых из k испытываемых объектов отказало m, описывается распределением Бернулли

$$f(p, k, m) = C_k^m p^{k-m}(1-p)^m.$$

Предположим, что априорное распределение параметра p есть β-распределение:

$$h(p) = C_{\alpha+\beta-1}^{\alpha-1} p^{\alpha-1}(1-p)^{\beta-1}.$$

Тогда апостериорное распределение параметра p будет иметь вид

$$h_{\text{апост}}(p/\hat{p}) = C_{k+\alpha+\beta-1}^{m+\alpha-1} p^{k-m+\alpha-1}(1-p)^{m+\beta-1}, \qquad (8.12)$$

где $\hat{p}$ – оценка параметра p.

Таким образом, из (8.12) видно, что апостериорное распределение параметра p есть β-распределение с параметрами $(m+\beta)$ и $(k-m+\alpha)$. Иными словами получен следующий результат: семейство β-распределений и семейство распределений Бернулли являются семействами сопряженных распределений.

2. Если случайная выборка $T_1, T_2, \ldots, T_n$ есть выборка из нормального распределения с неизвестным значением среднего m и заданной дисперсией σ^2 и априорное распределение m – нормальное, то апостериорное распределение m также нормальное распределение. То есть, произведя оценивание параметра математического ожидания нормально распределенной выборки случайных величин получаем, что априорное и апостериорное распределения оцениваемой величины принадлежат классу нормальных распределений.

3. Рассмотрим семейство Г-распределений. Пусть случайная выборка $T_1, T_2, \ldots, T_n$ есть выборка из Г-распределения с неизвестным значением параметра масштаба λ и известным параметром формы α:

$$f(t,\theta,\alpha) = \frac{\theta(\theta t)^{\alpha-1}}{\Gamma(\alpha)}\exp(-\theta t).$$

Пусть априорное распределение λ также относится к классу гамма-распределений:

$$h_\lambda(\theta) = \frac{\lambda_{\text{a}}(\lambda_{\text{a}}\theta)^{\alpha_{\text{a}}-1}}{\Gamma(\alpha_{\text{a}})}\exp(-\lambda_{\text{a}}\theta).$$

Тогда можно определить апостериорную плотность распределения параметра λ. Определим вначале совместную плотность распределения результатов наблюдений:

$$f(\{T_i\};\theta,\alpha) = \frac{\theta^{\alpha n}}{\{\Gamma(\alpha)\}^n}\exp\left(-\theta\sum_{i=1}^{n}T_i\right)\prod_{i=1}^{n}T_i.$$

Апостериорная плотность в этом случае имеет вид

$$h_{\text{апост}}(\theta/\{T_i\}) = \frac{\theta^{\alpha_{\text{a}}+n\alpha-1}}{\Gamma(\alpha_{\text{a}}+n\alpha)}\exp\left(-\theta\left(\lambda_{\text{a}}+\sum_{i=1}^{n}T_i\right)\right),$$

т.е. получили, что априорное и апостериорное распределения принадлежат к классу гамма-распределений.

4. Пусть $Q_1, Q_2, \ldots, Q_n$ – выборка дискретных случайных величин из распределения Пуассона с неизвестным значением среднего q. Тогда совместную плотность результатов наблюдения можно записать в виде

$$f(q,\hat{q}) = \frac{(kq)^m}{m!}\exp(-kq),$$

где $m = \sum_{i=1}^{k} Q_i$.

Предположим, что априорное распределение параметра q есть гамма-распределение с параметрами λ и α:

$$h_q(q,\lambda,\alpha) = \frac{\lambda(\lambda q)^{\alpha-1}}{\Gamma(\lambda)}\exp(-\lambda q).$$

Определим апостериорную плотность распределения параметра q:

$$h_{\text{апост}}(q/\hat{q}) = \frac{q^{\alpha+m-1}\exp(-q(\lambda+k))}{\int_{\Theta} q^{\alpha+m-1}\exp(-q(\alpha+k))dq},$$

где Θ – область определения параметра q. Если q представляет собой вероятностную характеристику, например, вероятность отказа, то $\Theta = [0, 1]$.

Получен следующий результат: распределение Пуассона и Г-распределение являются сопряженными распределениями.

Исследование сопряженности распределений позволяет обоснованно подходить к выбору априорной плотности распределения оцениваемого параметра. Если из практики наблюдения за функционированием объекта на этапе текущих исследований удалось восстановить плотность распределения наблюдаемой случайной величины, то при наличии априорной информации априорную плотность распределения оцениваемого параметра следует выбирать из класса сопряженных распределений к плотности распределения наблюдаемой величины.

8.7. Формирование априорной плотности распределения оцениваемого параметра

Подведем некоторые итоги, касающиеся применимости байесовских методов. Во-первых, в п. 8.2 представлена процедура формирования обобщенной вероятностной плотности, учитывающей текущую информацию. Решение этой задачи не представляет трудности. Во-вторых, сформулирована методика, позволяющая применять байесовский подход, основанная на использовании в качестве априорной плотности несобственной плотности распределения.

В общем случае задание априорной плотности распределения – сложная задача. Развитие байесовских методов исторически связано с использованием в качестве априорной информации мнения лица, принимающего решение. В примерах пп. 8.1 и 8.2 априорная информация формировалась на основании суждения исследователя, проводящего системный анализ. Вероятность, сформированная на основании учета мнения одного или нескольких экспертов по проблеме, относительно которой принимается решение, называется субъективной вероятностью. В простейшем случае субъективная вероятность отражает мнение одного лица. Так, в примере п. 8.1 заказчик с вероятностью p_1 оценил истинность заявления завода-изготовителя, которое состояло в том, что ВБР изделия в течение требуемого времени не хуже, чем $p_{з.н}$, а с вероятностью p_2 – заявление предприятия, имеющего опыт эксплуатации аналогичных объектов и утверждавшего, что ВБР изделия ниже требуемого уровня.

Использовать субъективные вероятности для формирования априорной плотности распределения целесообразно в задачах принятия решения. В задачах же оценивания характеристик сложных систем необходима методика, позволяющая учитывать объективную информацию о функционировании объектов.

В качестве такой методики можно предложить следующий подход. Пусть имеются результаты эксплуатации или испытаний объектов-аналогов на различных установках. Результаты каждой серии испытаний представлены в виде некоторой статистики $\{T_{i_m},\ i_m = \overline{1,\ k_m}\}$, где T_{i_m} реализация наблюдаемой случайной величины i-го объекта в m-м испытании. Предполагается, что закон распределения случайной величины T известен. Необходимо оценить некоторый параметр θ, связанный с T.

Методика решения задачи будет заключаться в следующем. Поскольку необходимо пользоваться только объективными данными о функционировании объектов, то на первом шаге выберем в качестве

априорной плотности распределения параметра θ несобственную априорную плотность распределения. Вид плотности целесообразно выбирать из класса сопряженных распределений. Такой выбор вида априорной плотности обеспечит удобство работы в ходе байесовского оценивания. Однако это требование не является обязательным. Позднее будет показано, что при увеличении объема информации о наблюдаемом процессе или явлении влияние вида априорной плотности на конечный результат оценивания сказывается все меньше и меньше.

Пример. Пусть случайная величина T распределена по нормальному закону с известным параметром среднего квадратического отклонения и неизвестным математическим ожиданием, которое и требуется оценить. Возьмем на первом шаге в качестве априорной плотности несобственное распределение из класса нормальных распределений с параметрами $m = 0$ и $\sigma \to \infty$, как это сделано в п. 8.4. После подстановки априорной плотности и функции правдоподобия в формулу Байеса получим апостериорное распределение в виде

$$h_{\text{апост}}(\theta / \{T_i\}) = \frac{1}{\sqrt{2\pi}S} \exp\left(-\frac{(\theta - m_{\text{т}})^2}{2S^2}\right), \tag{8.13}$$

где $m_{\text{т}}$ и S – оценки математического ожидания и среднего квадратического отклонения, полученные по одной из выборок $\{T_{i_m},\ i_m = \overline{1,\ k_m}\}$. Методика получения приведенного результата изложена в п. 8.4. На следующем шаге объединения информации апостериорная плотность (8.13) принимается в качестве априорной по отношению к следующей выборке. В п. 8.3 было показано, как производить учет последовательно накапливаемой информации. Там же отмечено, что окончательный результат оценивания не зависит от того, какая выборка получена сначала. На основании изложенных до настоящего момента методов сформулируем методику применения байесовских процедур в задачах оценивания показателей сложных систем, закладываемых при составлении моделей.

1. Определяется вид закона распределения наблюдаемой случайной величины, характеризующей функционирование объекта системного исследования.

2. Из класса сопряженных к сформированной плотности распределения определяется вид априорной плотности.

3. В качестве априорной плотности распределения принимается несобственная плотность вида, установленного на этапе 2.

4. Результаты наблюдений, полученные на этапе априорных и текущих исследований, приводятся к схеме последовательного накопления

информации. На основании информации, полученной на первом этапе наблюдений, формируется функция правдоподобия и по формуле (8.4) определяется апостериорная плотность. В качестве априорной плотности используется несобственная плотность из класса сопряженных распределений, а в качестве функции правдоподобия – функция, сформированная на основании информации, полученной на первом этапе наблюдений.

5. Формируется функция правдоподобия на основании текущей информации (второй этап наблюдений). Определяется апостериорная плотность распределения, в которой в качестве априорной плотности используется апостериорная, полученная на этапе 4, в качестве функции правдоподобия – функция, сформированная на основании информации второго этапа наблюдений.

6. На основании апостериорной плотности определяются байесовская оценка параметра (8.6) и точность в определении байесовской оценки (8.7).

7. При поступлении новой информации о функционировании объекта исследования заново строится функция правдоподобия и определяется апостериорная плотность, где в качестве априорной плотности используется апостериорная плотность, полученная на этапе 5. Затем вновь выполняется этап 6.

Другой метод формирования априорной плотности распределения, основанный на использовании реальной информации о функционировании объектов, впервые был изложен в [42]. Для обоснованного применения предлагаемого метода необходимо сделать одно предположение, которое заключается в следующем: вектор параметров θ заменяем вектором оценок $\hat{\theta}$ и работаем в дальнейшем с оценками параметров θ. В этом случае вектор представляет собой случайный вектор и может идти речь об определении его плотности распределения статистическими методами. Такая замена вектора параметров θ вектором оценок $\hat{\theta}$ является основным элементом эмпирического байесовского подхода. Это предположение уже использовалось при изложении метода байесовского оценивания, когда в качестве априорной плотности распределения оцениваемого параметра использовалось несобственное распределение. В этом методе после однократного применения формулы Байеса получают плотность распределения параметра θ, которая строится исключительно по результатам выборки $\{T_i,\ i=\overline{1,k}\}$. По своей сути построенная таким образом плотность представляет собой не что иное, как плотность распределения оценки вектора параметров $\hat{\theta}$. На следующем шаге учета информации апостериорную плотность, полученную на предыдущем этапе, берем в качестве априорной. Таким образом, после каждого применения теоремы Байеса получается плотность распределения оценки, учитывающая вновь поступающую информацию.

Итак, под априорной плотностью распределения будем понимать плотность распределения оценки θ, полученную на предварительном этапе системных исследований.

Изложим метод восстановления априорной плотности оцениваемого параметра применительно к следующей схеме наблюдений. В результате проведения эксперимента наблюдается последовательность $\{\tilde{T}_1, \hat{\theta}_1\}, \{\tilde{T}_2, \hat{\theta}_2\}, \ldots$ независимых случайных чисел. После n-го наблюдения T_n оценивается параметр $\hat{\theta}$. Поскольку $\hat{\theta}_1, \ldots, \hat{\theta}_{n-1}$ порождены тем же самым априорным распределением, что и $\hat{\theta}_n$, то $\hat{\theta}_1, \ldots, \hat{\theta}_{n-1}$ содержат также некоторую информацию о $\hat{\theta}_n$. Поэтому оценку $\hat{\theta}_n$ будем искать как функцию от случайных величин $\{\tilde{T}_n; \tilde{T}_1, \ldots, \tilde{T}_{n-1}\}$. Основываясь на этих данных, можно получить плотность распределения оценки $\hat{\theta}_n$. Эту процедуру можно провести, например, когда оценки $\hat{\theta}_n$ выражаются через статистику $M(T)=\sum_{i=1}^{n}\tilde{T}_i$ или статистику $D(T)=\sum_{i=1}^{n}\left(\tilde{T}_i-\frac{1}{n}M(T)\right)^2$, где $\tilde{T}_i$ – реализации случайной величины t, зафиксированные на этапе априорных исследований.

Рассмотрим метод определения плотности распределения $h(\theta)$ по известной плотности распределения наработки до отказа $f(\theta, t)$ в случае, когда параметр закона распределения выражается через достаточную статистику $M(T)=\sum_{i=1}^{n}\tilde{T}_i$. При переходе от плотности распределения наработки до отказа $f(\theta, t)$ случайной величины t к априорной плотности $h(\theta)$ воспользуемся аппаратом характеристических функций.

Пусть $\varphi_т(y)$ – характеристическая функция случайной величины t, имеющей плотность распределения $f(\theta, t)$. Известно, что характеристическая функция суммы случайных величин $M(T)$ в случае, когда все T_i имеют одну и ту же плотность распределения $f(\theta, t)$, определяется из соотношения

$$\varphi_м(y)=\{\varphi_т(y)\}^n .$$

Применяя к данному выражению обратное фурье-преобразование, получаем плотность распределения случайной величины $\xi = M(T)$. Если параметр закона распределения θ выражается через статистику $M(T)$, то в этом случае легко перейти от плотности распределения $f_м(\xi)$ к плотности $h(\theta)$, а именно:

17 — 4355

$$h(\theta) = R\, f_м(\xi),$$

где R = ξ/θ.

Проиллюстрируем возможности применения данного метода определения априорной плотности на примере оценивания параметра масштаба λ гамма-распределения $\Gamma(t, \lambda, \alpha)$. Будем считать, что параметр формы α известен. На первом этапе необходимо определить априорную плотность $h_\lambda(\theta)$ распределения параметра λ. Выразим неизвестный параметр через достаточную статистику $M(T)$. Известно, что для гамма-распределения выполняются следующие соотношения [30]:

$$M(t) = \frac{\alpha}{\lambda};\ D(t) = \frac{\alpha}{\lambda^2};\ M(t) = \frac{1}{n}M(T);\ D(t) = \frac{1}{n-1}D(T).$$

Из этих соотношений можно получить $\lambda = M(t)/D(t)$. Раскрываем знак математического ожидания, а дисперсию заменяем ее оценкой S^2, в результате получаем

$$\lambda = \frac{\sum_{i=1}^{n}\tilde{T}_i}{nS^2} = \theta. \tag{8.14}$$

Все случайные величины T_i имеют одну и ту же плотность распределения

$$\Gamma(t,\alpha,\lambda) = \frac{{}^{\alpha}t^{\alpha-1}}{\Gamma(\alpha)}\exp(-\hat{\lambda}t). \tag{8.15}$$

Характеристическая функция случайной величины t получается путем применения к плотности (8.15) преобразования Фурье и имеет вид

$$\varphi_t(y) = \left(\frac{\hat{\lambda}}{\hat{\lambda} - iy}\right)^{\alpha}.$$

Характеристическая функция случайной величины x, равной сумме n одинаково распределенных случайных величин $t\left(X = \sum_{i=1}^{n} T_i\right)$, определяется из соотношения

$$\varphi_x(y) = \left(\frac{\hat{\lambda}}{\hat{\lambda} - iy}\right)^{n\alpha}. \tag{8.16}$$

Применяя обратное фурье-преобразование к (8.16), получаем плотность распределения случайной величины x:

$$f(x) = \frac{\hat{\lambda}^{n\alpha}}{\Gamma(n\alpha)}x^{n\alpha-1}\exp(-\hat{\lambda}x).$$

Перейдем от плотности случайной величины x к плотности случайной величины θ. Из (8.14) получаем $x = nS^2\theta$. Для случайных величин, связанных друг с другом функционально, возможно перейти от плотности распределения одной из них к плотности распределения другой [32]. Получаем, что априорная плотность распределения параметра θ имеет вид

$$h_\lambda(\theta) = \frac{\hat{\lambda}^{n\alpha}}{\Gamma(n\alpha)}nS^2(nS^2\theta)^{n\alpha-1}\exp\left(-\hat{\lambda}nS^2\theta\right). \tag{8.17}$$

Рассмотрим методику определения априорной плотности параметра закона распределения в случае, когда этот параметр выражается через статистику $D(T) = \sum_{i=1}^{n}\left(\tilde{T}_i - \frac{1}{n}M(T)\right)^2$. Произведем замену переменных $r = [t - M(T)]^2$. Величина r всегда неотрицательна, и при $r \geq 0$ соотношение $r \leq R$ эквивалентно соотношению $-\sqrt{R} \leq t \leq \sqrt{R}$; без ограничения общности считаем $M(t)=0$. Если это условие не выполняется, то в качестве t можно использовать величину $\tilde{t} = [t - M(t)]$. Тогда случайная величина r будет иметь функцию распределения

$$G(r) = \begin{cases} 0, & r < 0; \\ F(\sqrt{r}) - F(-\sqrt{r}), & r \geq 0. \end{cases}$$

где $F(t)$ – функция распределения случайной величины t.

Если плотность вероятности $f(t) = F'(t)$ существует для всех t, то r имеет плотность вероятности

$$g(r) = G'(r) = \begin{cases} 0 \text{ при } r < 0; \\ \frac{1}{2\sqrt{r}}\left[f(\sqrt{r}) + f(-\sqrt{r})\right] \text{ при } r \geq 0. \end{cases} \tag{8.18}$$

Далее методика определения априорной плотности аналогична методике, когда параметр выражается через статистику $M(T)$. Статистика $D(T)$ выражается через сумму случайных величин R_i с известной плотностью распределения $g(r)$. Следовательно, для определения плотности распределения величины $D(T)$ можно также воспользоваться аппаратом характеристических функций.

Применение данного метода рассмотрим на примере оценивания параметра σ^2 нормального закона распределения. Пусть требуется оценить дисперсию нормального закона распределения σ^2 при известном значении математического ожидания m.

На первом этапе оценивания плотности $h_\sigma(\theta)$ положим $\sigma = 1$. Тогда, согласно формуле (8.18), плотность случайной величины r будет иметь вид

$$g(r) = \begin{cases} 0 \text{ при } t \le 0; \\ \dfrac{1}{\sqrt{2\pi r}} \exp\left(-\dfrac{r}{2}\right) \text{ при } t > 0. \end{cases}$$

Соответствующая этой функции плотности характеристическая функция равна

$$\int_0^\infty \exp(iyr)\frac{1}{\sqrt{2\pi r}} \exp\left(-\frac{r}{2}\right) dr = (1-2iy)^{-\frac{1}{2}}.$$

Каждая величина r имеет характеристическую функцию $(1-2iy)^{-1/2}$, следовательно, их сумма имеет характеристическую функцию $\varphi_D(y) = (1-2iy)^{-n/2}$.

Соответствующая плотность распределения будет иметь вид

$$K(r) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)} r^{\frac{n}{2}-1} \exp\left(-\dfrac{r}{2}\right) \text{ при } r \ge 0; \\ 0 \text{ при } r < 0. \end{cases}$$

Если среднее квадратическое отклонение не равно 1, а есть конечная величина $S_\text{а}^2$, то плотность распределения суммы $\sum_{i=1}^{n} R_i^2$ имеет вид

$$\frac{1}{S_\text{а}^2} K\left(\frac{r}{S_\text{а}^2}\right) = \frac{1}{2^{n/2} S_\text{а}^2 \Gamma\left(\dfrac{n}{2}\right)} r^{\frac{n}{2}-1} \exp\left(-\frac{r}{2S_\text{а}^2}\right).$$

Отсюда получаем, что случайная величина $z = n\sigma^2/S_\text{а}^2$ имеет распределение χ^2 с $(n-1)$ степенями свободы: $S_\text{а}^2$ – априорная оценка дисперсии. Считая σ^2 случайной величиной и полагая $\theta_N = \sigma^2$, получаем

$$h_{\sigma^2}(\theta_N) = n/S_\text{а}^2 \frac{\left(\theta_N\, n/S_\text{а}^2\right)^{\frac{n-1}{2}}}{\Gamma\left(\dfrac{n-1}{2}\right) 2^{\frac{n-1}{2}}} \exp\left(-\frac{\theta_N\, n/S_\text{а}^2}{2}\right). \quad (8.19)$$

Рассмотренные два случая определения априорной плотности параметров законов распределения (параметра λ гамма-распределения, а также m и σ^2 нормального закона) охватывают широкий класс распределений и имеют большое прикладное значение в задачах системного анализа.

Изложенный метод применим лишь в тех случаях, когда оценка искомого параметра выражается через достаточную статистику. Например, при оценивании параметров распределения Вейбулла данный метод не применим. Однако преимущество данного метода по сравнению с методом, основанным на использовании в качестве априорной плотности распределения несобственного распределения, заключается в том, что он не требует знания вида сопряженного распределения. Этот метод может быть использован для установления вида сопряженных распределений.

В заключение данного параграфа приведем теорему Бернштейна-Мизеса в той формулировке, которая дана в [41]:

«Если априорная плотность распределения параметра θ непрерывна, то по мере возрастания числа наблюдений апостериорное распределение, задаваемое формулой Байеса, стремится к пределу, не зависящему от априорного распределения». Согласно этой теореме, при оценивании параметра θ при достаточно большом объеме наблюдений выбор вида априорной плотности распределения не имеет существенного значения. Однако заметим, что при малом числе наблюдений желательно в качестве априорной плотности использовать плотность из класса сопряженных распределений. Такой выбор вида априорной плотности гарантирует относительную простоту процедуры оценивания.

8.8. Оценивание параметров нормального закона распределения

При решении задач системного анализа на основании наблюдений за случайной величиной, распределенной по нормальному закону, возможны следующие ситуации:

1) известно значение математического ожидания и среднего квадратического отклонения;

2) известно значение среднего квадратического отклонения, требуется оценить математическое ожидание;

3) известно значение математического ожидания, требуется оценить среднее квадратическое отклонение;

4) оценивается и среднее квадратическое отклонение, и математическое ожидание.

Случай 1. Тривиален, поскольку о параметрах распределения все известно.

Случай 2. Пусть в результате системных исследований группы однотипных объектов зафиксирована выборка $\{T_i\}$, $i=\overline{1,k}$, где T_i – параметр, характеризующий исследуемый показатель сложной системы или анализируемого объекта. Исследователь располагает априорной информацией об анализируемом параметре объекта, однотипного с исследуемыми. Пусть $\{\tilde{T}_j\}$, $j=\overline{1,n}$ – выборка, зафиксированная на этапе априорных исследований. Будем считать, что случайные величины T_i и $\tilde{T}_j$ имеют одну и ту же функцию распределения, т.е. априорная и текущая информации однородны. В данном случае функция распределения нормальна и ее плотность имеет вид

$$f_N(t;m,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

Основываясь на результатах априорных исследований, определим вид априорной плотности распределения параметра m. Характеристическая функция каждой величины $\tilde{T}_j$ равна

$$\varphi_T(y)=\exp\left(m_a iy-\frac{1}{2}S^2y^2\right).$$

Характеристическую функцию суммы n независимых случайных величин определяем из соотношения

$$\varphi_M(y)=\exp\left(n\,m_a iy-\frac{n}{2}S^2y^2\right).$$

Тогда плотность распределения суммы независимых случайных величин будет иметь вид

$$f_M(x,m_a,S)=\frac{1}{\sqrt{2\pi n}\,S}\exp\left(-\frac{(x-\hat{m}_a n)^2}{2nS^2}\right).$$

Перейдем к переменной $\theta=\frac{x}{n}$, получим распределение математического ожидания:

$$f_m(\theta,m_a,S)=\frac{1}{\sqrt{\frac{2\pi}{n}}\,S}\exp\left(-\frac{(\theta-\hat{m}_a)^2}{2S^2/n}\right).$$

Величина S^2/n представляет собой дисперсию оценки параметра математического ожидания. Обозначим ее через S_m^2. Таким образом, получаем, что оценка математического ожидания имеет нормальное распределение:

$$f_m(\theta,m_a,S_m)=\frac{1}{\sqrt{2\pi}\,S_m}\exp\left(-\frac{(\theta-\hat{m}_a)^2}{2S_m^2}\right)=h(\theta). \tag{8.20}$$

Данную плотность распределения примем в качестве априорной плотности распределения оцениваемого параметра, так как эта плотность построена на основании априорной информации об анализируемом параметре объекта.

Определим апостериорную плотность распределения оцениваемого параметра. Функция правдоподобия формируется на основании текущей информации и в случае нормального распределения случайной величины, характеризующей исследуемый показатель, она имеет вид

$$f(\theta,\{T_i\})=\frac{k}{\sqrt{2\pi}\,\sigma_т}\exp\left(-\frac{(\theta-\hat{m}_т)^2k}{2\sigma_т^2}\right).$$

Согласно формуле Байеса апостериорная плотность распределения

$$h_{апост}(\theta/\{T_i\})=\frac{\exp(-a)}{\sqrt{\frac{2\pi S^2\sigma_т^2}{nk}}\int\limits_{-\infty}^{\infty}f(\theta,\{T_i\})h(\theta)\,d\theta},$$

где $a=\frac{k(\hat{m}_т-\theta)^2}{2\sigma_т^2}+\frac{(\theta-\hat{m}_a)^2n}{2S^2}$.

Преобразуем данное выражение, приведя его к общему знаменателю и раскрывая скобки, получаем

$$a=\frac{n\sigma_т^2+kS^2}{2\sigma_т^2S^2}\left(\theta-\frac{n\sigma_т^2\hat{m}_a+kS^2\hat{m}_т}{n\sigma_т^2+kS^2}\right)+b,$$

где $b = \dfrac{(\hat{m}_т - \hat{m}_а)^2}{2(n\sigma_т^2 + kS^2)/kn}$.

Апостериорное распределение для θ можно записать теперь в виде

$$h_{апост}(\theta/\{T_i\}) = A\exp\left(-\left(\theta - \frac{n\sigma_т^2\hat{m}_а + kS^2\hat{m}_т}{n\sigma_т^2 + kS^2}\right)\left(\frac{n\sigma_т^2 + kS^2}{2\sigma_т^2 S^2}\right)\right).$$

Так как интеграл от этого выражения по области определения параметра θ должен равняться единице, т.е. должно соблюдаться условие нормировки, то

$$A = \frac{\sqrt{n\sigma_т^2 + kS^2}}{\sqrt{2\pi}\,S\sigma_т}.$$

Отсюда видно, что апостериорное распределение математического ожидания случайной величины также является нормальным. При этом байесовская оценка параметра θ определяется выражением

$$\hat{m}_б = \frac{n\sigma_т^2\hat{m}_а + kS^2\hat{m}_т}{n\sigma_т^2 + kS^2},$$

точность в определении оценки

$$S_{m_б}^2 = \frac{\sigma_т^2 S^2}{n\sigma_т^2 + kS^2}.$$

Определим выигрыш в точности байесовской оценки математического ожидания по сравнению с оценкой этого параметра на основании только лишь текущей информации. Понятие выигрыша в точности показывает, во сколько раз байесовская оценка точнее оценки, полученной только лишь на основании текущей информации. Выигрыш в точности определяется из соотношения

$$\eta = D\{\hat{\theta}_т\}/D\{\hat{\theta}_б\}.$$

В рассматриваемом случае получаем

$$\eta = \frac{S_{m_т}^2}{S_{m_б}^2} = \frac{\sigma_т^2}{k}\frac{n\sigma_т^2 + kS^2}{\sigma_т^2 S^2} = \frac{n\sigma_т^2}{kS^2} + 1.$$

Таким образом, получен следующий результат: использование априорной информации при оценивании математического ожидания нормального закона распределения всегда приводит к выигрышу в точности по сравнению с результатом, получаемым только на основании текущей информации.

Случай 3. Значение математического ожидания известно, требуется оценить дисперсию распределения. Как было показано в п. 8.7, априорная плотность распределения оцениваемого параметра выражается формулой (8.19). Функция правдоподобия в этом случае имеет вид

$$f_N(\theta, \{T_i\}) = (2\pi)^{-\frac{k}{2}}\theta^{\frac{k}{2}}\exp\left(-\frac{k\theta}{2\sigma_т^2}\right).$$

Подставляя выражение для априорной плотности (8.19) и функции правдоподобия в формулу Байеса (8.4), получаем

$$h_{апост}(\theta/\{T_i\}) = \frac{h_{\sigma^2}(\theta)\, f_N(\theta, \{T_i\})}{\int\limits_0^\infty h_{\sigma^2}(\theta)\, f_N(\theta, \{T_i\})d\theta}.$$

Вычислим интеграл, входящий в данное выражение:

$$\int\limits_0^\infty h_{\sigma^2}(\theta)\, f_N(\theta, \{T_i\})d\theta = \frac{n/S_а^2\,(2\pi)^{-\frac{k}{2}}}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}}\left(n/S_а^2\right)^{\frac{n-3}{2}}\times$$

$$\times\int\limits_0^\infty \theta^{\frac{k+n-3}{2}}\exp\left(-\frac{\theta}{2}\left(n/S_а^2 + k/\sigma_т^2\right)\right)d\theta = \frac{\left(n/S_а^2\right)^{\frac{n-1}{2}}(2\pi)^{-\frac{k}{2}}\,\Gamma\left(\frac{k+n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}\left(\frac{n/S_а^2 + k/\sigma_т^2}{2}\right)^{\frac{k+n-1}{2}}}.$$

Подставив вычисленное значение интеграла в выражение для апостериорной плотности распределения параметра θ, получим

$$h_{апост}(\theta/\{T_i\}) = \frac{\left(n/S_а^2 + k/\sigma_т^2\right)^{\frac{k+n-1}{2}}}{\Gamma\left(\frac{k+n-1}{2}\right)}\theta^{\frac{k+n-3}{2}}\exp\left(-\theta\frac{n/S_а^2 + k/\sigma_т^2}{2}\right). \qquad (8.21)$$

Определим байесовскую оценку параметра θ. Подставим выражение (8.21) в (8.6). Результат вычисления имеет вид

$$\theta = \sigma_т^2 S_а^2 \Big/ \frac{n\sigma_т^2 + kS_а^2}{k+n-1}. \qquad (8.22)$$

Точность байесовской оценки в определении параметра θ получим подстановкой (8.21) в (8.7):

$$D(\theta)=\sigma_\text{т}^4 S_\text{а}^4 \Big/ \frac{(n\sigma_\text{т}^2+kS_\text{а}^2)^2}{2(k+n-1)}.$$

Байесовская оценка параметра σ^2 имеет вид

$$\hat{\sigma}_\text{б}^2=\frac{\sigma_\text{т}^2 S_\text{а}^2}{\dfrac{n\sigma_\text{т}^2+kS_\text{а}^2}{k+n-1}}.$$

Таким образом, получили байесовскую оценку параметра σ^2 нормального закона распределения.

Случай 4. Значения математического ожидания и дисперсии неизвестны.

Определим совместную плотность распределения параметров нормального закона $h(\theta_1, \theta_2)$, где $\theta_1 = m$, $\theta_2 = \sigma^2$.

Так как параметры θ_1 и θ_2 независимы, можно записать

$$h(\theta_1,\theta_2)=h_1(\theta_1)h_2(\theta_2).$$

Априорная плотность распределения параметра θ_2 описывается выражением (8.19). Априорная плотность распределения параметра θ_1 имеет вид (8.20), но так как значение дисперсии по условию неизвестно, то априорная плотность параметра θ_1 запишется следующим образом:

$$h_1(\theta_1)=\sqrt{\frac{\theta_2}{2\pi}}\exp\left(-\frac{n(\hat{m}_\text{а}-\theta_1)^2\theta_2}{2}\right). \tag{8.23}$$

Функция правдоподобия в данном случае определяется по формуле

$$f(\theta_1,\theta_2,\{T_i\})=(2\pi)^{-\frac{k}{2}}\theta_2^{\frac{k}{2}}\exp\left(-\frac{k\sigma_\text{т}^2+k(\hat{m}_\text{т}-\theta_1)^2}{2}\theta_2\right). \tag{8.24}$$

После подстановки выражений (8.19), (8.23) и (8.24) в формулу Байеса и проведения элементарных преобразований апостериорная плотность распределения

$$h_\text{апост}(\theta_1,\theta_2/\{T_i\})=\frac{\theta_2^{\frac{k+n-1}{2}}\exp\left(-\frac{\theta_2}{2}\left(k\sigma_\text{т}^2+nS_\text{а}^2+k(\hat{m}_\text{т}-\theta_1)^2+n(\hat{m}_\text{а}-\theta_1)^2\right)\right)}{\int_{-\infty}^{\infty}\int_0^{\infty}\theta_2^{\frac{k+n-1}{2}}\exp\left(-\frac{\theta_2}{2}\left(k\sigma_\text{т}^2+nS_\text{а}^2+k(\hat{m}_\text{т}-\theta_1)^2+n(\hat{m}_\text{а}-\theta_1)^2\right)\right)d\theta_1 d\theta_2}.$$

Выражения для определения исходных параметров можно получить в виде

$$\hat{\theta}_1=\frac{\int_0^{\infty}\theta_1\exp\left(-\frac{\theta_2}{2}\left(k(\hat{m}_\text{т}-\theta_1)^2+n(\hat{m}_\text{а}-\theta_1)^2\right)\right)d\theta_1}{C_1};$$

$$\hat{\theta}_2=\frac{\int_{-\infty}^{\infty}\theta_2^{\frac{n+k}{2}}\exp\left(-\frac{\theta_2}{2}\left(k\sigma_\text{т}^2+nS_\text{а}^2+k(\hat{m}_\text{т}-\theta_1)^2+n(\hat{m}_\text{а}-\theta_1)^2\right)\right)d\theta_2}{C_2},$$

где C_1 и C_2 – некоторые нормирующие константы.

Опустив промежуточные выкладки, приведем конечный результат для оценок параметров нормального распределения:

$$\hat{\theta}_1=\frac{k\hat{m}_\text{т}+n\hat{m}_\text{а}}{k+n};\quad \hat{\theta}_2=\frac{(n+k-1)\sigma_\text{т}^2S_\text{а}^2}{k\sigma_\text{т}^2+nS_\text{а}^2+k(\hat{m}_\text{т}-\theta_1)^2+n(\hat{m}_\text{а}-\theta_1)^2}.$$

Таким образом получили простую систему уравнений для определения неизвестных параметров нормального закона распределения.

В заключение данного параграфа рассмотрим методику оценивания моментов логарифмически нормального закона распределения. Напомним, что случайная величина t подчиняется логарифмически нормальному закону распределения, если ее логарифм $z=\ln t$ распределен нормально. Для получения байесовских оценок моментов m_L и σ_L логарифмически нормального закона необходимо от случайных величин t перейти к случайным величинам z: $z_i = \ln T_i$. Далее, используя формулы для байесовского оценивания параметров нормального закона распределения, находим оценки параметров $\hat{m}_z$ и $\hat{\sigma}_z$. Затем по соотношениям

$$m_L=\exp\left(m_z+\frac{\sigma_z^2}{2}\right);\quad \sigma_L^2=\left[\exp\left(\sigma_z^2\right)-1\right]\exp\left(2m_z+\sigma_z^2\right)$$

производим обратный переход к оценкам логарифмически нормального закона распределения.

8.9. Оценивание параметров семейства гамма-распределений

Применение байесовской методики в задаче оценивания параметра масштаба гамма-распределения состоит в следующем. Предположим, что параметр формы известен. Априорная плотность распределения параметра масштаба была получена в п. 8.7 (8.17). Функция правдоподобия определяется на основании текущей информации и имеет вид

$$f(\theta,\{T_i\})=\frac{\theta^{k\alpha}}{[\Gamma(\alpha)]^k}\exp(-k\tau_k\theta)\prod_{i=1}^{k}T_i^{\alpha-1}.$$

Подставляя это выражение и выражение (8.17) в формулу Байеса, получаем апостериорную плотность распределения оцениваемого параметра:

$$h_{\text{апост}}(\theta/\{T_i\})=(nS^2\hat{\lambda}+k\tau_k)^{n\alpha+k\alpha}\,\frac{\theta^{n\alpha+k\alpha-1}\exp\left(-\theta(nS^2\hat{\lambda}+k\tau_k)\right)}{n\alpha+k\alpha-1},\qquad(8.25)$$

где τ_k – математическое ожидание случайной величины t, полученное на этапе текущих наблюдений; k – объем выборки текущих значений наблюдаемой случайной величины.

Определим байесовскую оценку параметра λ и точность в определении этого параметра. Для этого подставим выражение для апостериорной плотности распределения параметра λ (8.25) в формулу (8.6), тогда

$$\theta_r=\frac{n\alpha+k\alpha-1}{nS^2\hat{\lambda}+k\tau_k}.\qquad(8.26)$$

Для оценки точности в определении параметра λ подставим выражение (8.25) в (8.7) и получим

$$D(\theta_r)=\frac{n\alpha+k\alpha-1}{(nS^2\hat{\lambda}+k\tau_k)^2}.\qquad(8.27)$$

Таким образом, получены формула (8.26) для объединения априорной и текущей информации о параметрах объектов, наблюдаемая случайная величина которых имеет гамма-распределение, и формула (8.27) для определения точности байесовской оценки параметра λ. Известно, что ряд законов распределения являются частными случаями гамма-распределения. Так, экспоненциальное распределение можно представить следующим образом: $E(t,\lambda)=\Gamma(t,1,\lambda)$;

распределение Рэлея: $R(t,\sigma^2)=\Gamma\left(t,\frac{1}{2},\frac{1}{2\sigma^2}\right)$;

χ^2-распределение: $k(t,l)=\Gamma\left(t,\frac{1}{2},\frac{1}{2}l\right)$.

Подставляя в выражения (8.26) и (8.27) соответствующие коэффициенты ($\alpha=1$ – для экспоненциального закона; $\alpha=1/2$, $\lambda=1/2\sigma^2$ – для закона Рэлея; $\alpha=1/2$, $\lambda=(1/2)l$ – для распределения χ^2), можно получить выражения для оценивания параметров этих законов распределения. Результаты вычисления данных параметров представлены в табл. 8.1.

Таблица 8.1

Вид закона распределения	Оценка параметра	Оценка дисперсии параметра
Гамма-распределенне $\Gamma(t,\alpha,\lambda)$	$\lambda_б=\frac{n\alpha+k\alpha-1}{nS^2\hat{\lambda}+k\tau_k}$	$D\{\lambda_б\}=\frac{n\alpha+k\alpha-1}{(nS^2\hat{\lambda}+k\tau_k)^2}$
Экспоненциальное распределенне $E(t,\lambda)$	$\lambda_б=\frac{n+k-1}{n\tau_n+k\tau_k}$	$D\{\lambda_б\}=\frac{n+k-1}{(n\tau_n+k\tau_k)^2}$
χ^2-распределенне	$l=\frac{n+k-2}{n\tau_n+k\tau_k}$	$D\{l\}=\frac{n+k-2}{(n\tau_n+k\tau_k)^2}$
Распределенне Рэлея $R(t,\sigma^2)$	$\frac{1}{\sigma^2}=\frac{n+k-2}{n\tau_n+k\tau_k}$	$D\left\{\frac{1}{\sigma^2}\right\}=\frac{n+k-2}{(n\tau_n+k\tau_k)^2}$

Примечание. τ_n – математическое ожидание случайной величины t, полученное на этапе априорных наблюдений; n – объем выборки априорных значений наблюдаемой случайной величины

Таким образом, полученный результат оценивания параметра λ гамма-распределения может быть распространен на большую группу законов распределения.

8.10. Байесовское оценивание параметров по многократно цензурированным данным

До настоящего времени излагались модели байесовского оценивания, основанные на довольно простых планах испытаний (эксплуатации). В частности, в предыдущих параграфах описана схема обработки результатов наблюдений, полученных в предположении, что в каждом испытании реализуется наблюдаемый признак. Например, если решается задача анализа надежности, то описанная схема предполагает, что все объекты, находящиеся под наблюдением, доведены до отказа.

На практике при эксплуатации элементов и устройств наблюдается иная картина. Как уже отмечалось в предыдущей главе, эксплуатационная информация, поступающая на обработку, бывает представлена в виде многократно цензурированных, группированных данных. Рассмотрим последовательность применения процедуры байесовского оценивания в указанных ситуациях.

Пусть текущая информация представлена в виде выборки объема $r = k+\nu$, которая содержит ряд элементов с реализовавшимся наблюдаемым признаком $T_1, T_2, \ldots, T_k$ и ряд элементов с не реализовавшимся признаком, т.е. цензурированные данные $T'_1, T'_2, \ldots, T'_\nu$.

Известна плотность распределения наблюдаемой случайной величины, которая имеет вид $f(\theta, t)$, где θ – вектор параметров. Пусть $h(\theta)$ – априорная плотность распределения вектора θ.

В такой постановке оценивание вектора параметров будем проводить следующим образом. Как следует из результатов, изложенных в гл. 7, функция правдоподобия для выборки с элементами, для которых реализовался признак, и элементами, содержащими цензурированные справа данные, записывается в следующем виде:

$$f(\theta,\{T_i\}) = \prod_{l=1}^{k} f(\theta, T_l) \prod_{j=1}^{\nu} \left(1 - F(\theta, T'_j)\right).$$

Далее процедура оценивания не отличается от уже изложенной ранее. Апостериорная плотность распределения записывается следующим образом:

$$h_{\text{апост}}(\theta/\{T_i\}) = \frac{h(\theta)\prod_{l=1}^{k} f(\theta, T_l)\prod_{j=1}^{\nu}\left(1 - F(\theta, T'_j)\right)}{\int_{\Theta} h(\tau)\prod_{l=1}^{k} f(\tau, T_l)\prod_{j=1}^{\nu}\left(1 - F(\tau, T'_j)\right)d\tau}.$$

Оценки вектора параметров и точности в их определении рассчитываются по (8.6), (8,7).

Получить решение данной задачи в явном виде, по всей вероятности, не удастся ни для одного распределения. Решение необходимо искать численными методами.

Покажем, какие выражения получаются в самом простейшем случае, когда наблюдаемая случайная величина распределена по экспоненциальному закону. В этом случае функция правдоподобия

$$f(\theta,\{T_i\}) = \prod_{i=1}^{k} \theta \exp(-\theta T_i) \prod_{j=1}^{\nu}\left(1 - \exp(-\theta T'_j)\right).$$

Апостериорная плотность распределения параметра θ с учетом (8.17) запишется следующим образом:

$$h_{\text{апост}}(\theta/\{T_i\}) = \frac{\theta^{k+n-1}\exp\left(-\theta\left(nS^2\hat{\lambda} + k\tau_k\right)\right)\prod_{j=1}^{\nu}\left(1 - \exp\left(-\theta T'_j\right)\right)}{\int_0^\infty \tau^{k+n-1}\exp\left(-\tau\left(nS^2\hat{\lambda} + k\tau_k\right)\right)\prod_{j=1}^{\nu}\left(1 - \exp\left(-\tau T'_j\right)\right)d\tau},$$

где $\tau_k = \dfrac{\sum_{l=1}^{k} T_l}{k}$.

Для экспоненциального распределения известно, что математическое ожидание равняется среднеквадратическому отклонению и равно обратной величине интенсивности отказа, следовательно, выполняется соотношение $M(T) = S = 1/\lambda$, и выражение для апостериорной плотности можно переписать как

$$h_{\text{апост}}(\theta/\{T_i\}) = \frac{\theta^{k+n-1}\exp\left(-\theta\left(n\tau_n + k\tau_k\right)\right)\prod_{j=1}^{\nu}\left(1 - \exp\left(-\theta T'_j\right)\right)}{\int_0^\infty \tau^{k+n-1}\exp\left(-\tau\left(n\tau_n + k\tau_k\right)\right)\prod_{j=1}^{\nu}\left(1 - \exp\left(-\tau T'_j\right)\right)d\tau},$$

где $\tau_n = \dfrac{\sum_{i=1}^{n} T_i}{n}$ оценивается по априорной информации.

Оценка параметра экспоненциального распределения

$$\hat{\theta}_б = \frac{\int_0^\infty \theta^{k+n}\exp\left(-\theta\left(n\tau_n + k\tau_k\right)\right)\prod_{j=1}^{\nu}\left(1 - \exp\left(-\theta T'_j\right)\right)d\theta}{\int_0^\infty \theta^{k+n-1}\exp\left(-\theta\left(n\tau_n + k\tau_k\right)\right)\prod_{j=1}^{\nu}\left(1 - \exp\left(-\theta T'_j\right)\right)d\theta},$$

при этом точность в определении параметра θ имеет вид

$$\sigma_б^2 = \frac{\int_0^\infty \theta^{k+n+1}\exp\left(-\theta\left(n\tau_n + k\tau_k\right)\right)\prod_{j=1}^{\nu}\left(1 - \exp\left(-\theta T'_j\right)\right)d\theta}{\int_0^\infty \theta^{k+n-1}\exp\left(-\theta\left(n\tau_n + k\tau_k\right)\right)\prod_{j=1}^{\nu}\left(1 - \exp\left(-\theta T'_j\right)\right)d\theta} - \hat{\theta}_б^2.$$

Определение данных параметров может быть осуществлено путем численного интегрирования.

Не перегружая в дальнейшем материал формулами, отметим, что в случае, когда у исследователя имеется информация, содержащая данные, цензурированные слева или интервалом, необходимо воспользоваться соответствующей функцией правдоподобия, приведенной в гл. 7. Далее функцию правдоподобия, учитывающую цензурированную информацию, подставляют в выражение для апостериорной плотности распределения и на основании ее получают оценки параметров законов распределения и вычисляют точность в определении данных параметров.

8.11. Байесовское оценивание вероятностных показателей сложных систем

До настоящего времени были изложены задачи оценивания, касающиеся определения значений параметров того или иного закона распределения случайной величины. Зная вид закона распределения и оценив его параметры, можно перейти к определению вероятностных показателей сложных систем.

Так, например, если в качестве наблюдаемой случайной величины рассматривается наработка до отказа, то, определив закон распределения наработки, можно вычислить характеристики надежности, например, наработку на отказ:

$$T_{\text{ср}} = \int_0^{\infty} t\, f(t,\hat{\theta})\, dt\,;$$

вероятность безотказной работы объекта за время $T_{\text{р}}$

$$P(T_{\text{р}}) = 1 - \int_0^{T_{\text{р}}} f(t,\hat{\theta})\, dt\,;$$

интенсивность отказа

$$\lambda(t,\hat{\theta}) = f(t,\hat{\theta}) / F(t,\hat{\theta})$$

и ряд других характеристик.

Однако при решении задач системного анализа встречаются случаи, когда можно произвести оценивание тех или иных вероятностных характеристик объектов системного анализа, не определяя вид и параметры закона распределения наработки до отказа. Проиллюстрируем

возможность решения такого рода задач на примере оценивания показателей надежности.

Рассмотрим некоторые из этих задач.

1. При эксплуатации высоконадежных систем и устройств имеют место ситуации, когда априорная информация представлена в виде оценки ВБР устройства за время $T_{\text{р}} - \hat{p}_{\text{а}}(T_{\text{р}})$, известен доверительный интервал $(p_{\text{н}}, p_{\text{в}})$ с доверительной вероятностью $p_{\text{д}} = P(p_{\text{н}} \le p \le p_{\text{в}})$, здесь $p_{\text{н}}$ и $p_{\text{в}}$ – соответственно нижняя и верхняя границы доверительного интервала. Информация в таком виде почти всегда присутствует в паспортах и технических условиях на устройства и системы.

Пусть текущая информация получена в результате проведения некоторого числа испытаний k, при которых зафиксировано m отказов. На основании теоремы Байеса можно произвести оценивание ВБР с учетом априорной и текущей информации.

Будем считать, что искомая характеристика надежности p – случайная величина. Предположим, что известна априорная плотность распределения $h(p)$. По результатам текущих исследований определяем плотность $f(p,\hat{p})$. Здесь $f(p,\hat{p})$ – совместная плотность распределения опытного значения искомого показателя надежности и истинного его значения, равного p.

Формула Байеса в данном случае будет иметь вид

$$h_{\text{апост}}(p/\hat{p}) = \frac{h(p)\, f(p,\hat{p})}{\int_0^1 h(p)\, f(p,\hat{p})\, dp}.$$

Тогда байесовская оценка ВБР может быть получена методом моментов

$$\hat{p}_{\text{б}} = \int_0^1 p\, h_{\text{апост}}(p/\hat{p})\, dp\,, \qquad (8.28)$$

а точность в определении байесовской оценки

$$\sigma_{\text{р}}^2 = \int_0^1 p\, h_{\text{апост}}(p/\hat{p})\, dp - \hat{p}_{\text{б}}^2\,. \qquad (8.29)$$

Описана достаточно общая формулировка задачи оценивания ВБР при наличии априорной информации, заданной в виде доверительного интервала. Перейдем к рассмотрению частных случаев оценивания.

2. Постановка задачи. Пусть априорная информация задана в виде доверительного интервала $(p_{\text{н}}, p_{\text{в}})$ с известной доверительной вероят-

18 — 4355

ностью $p_д = P(p_н \le p \le p_в)$. Пусть также известна оценка ВБР объекта за время $T_р - \hat{p}_а(T_р)$. Будем считать, что границы $p_н$ и $p_в$ симметричны относительно величины $\hat{p}_а$.

В данном случае в качестве закона распределения для величины p можно принять нормальный закон распределения. Оценка ВБР представляет собой сумму n независимых одинаково распределенных случайных величин

$$p_i = \begin{cases} 0, & t_i < T_р; \\ 1, & t_i \ge T_р, \end{cases}$$

и, согласно центральной предельной теореме, при достаточно большом n ее закон распределения близок к нормальному. На практике [32] даже при относительно небольшом числе слагаемых (примерно 10 – 20) закон распределения суммы можно приближенно считать нормальным. Итак, будем считать, что величина p имеет нормальный закон распределения. Априорную плотность распределения запишем в виде

$$h(p) = n(\hat{p}_а, \sigma^2_{\hat{p}_а}, p), \tag{8.30}$$

где $\hat{p}_а$ и $\sigma^2_{\hat{p}_а}$ – среднее значение ВБР и среднее квадратическое отклонение.

Пусть в результате текущих наблюдений зафиксирована выборка $(p_1, p_2, \ldots, p_k)$

$$p_i = \begin{cases} 0, & \text{если за время } T_р \text{ объект отказал;} \\ 1, & \text{если отказа не было.} \end{cases}$$

В условиях, рассматриваемых в данном пункте, случайная величина p m раз приняла значение 0 и $k - m$ раз значение 1. Среднее значение этой выборки равно $\hat{p}_т = \frac{1}{k}\sum_{i=1}^{k} p_i$ и является нормально распределенной случайной величиной. Для любого фиксированного значения математического ожидания величины p функция правдоподобия при реализовавшемся среднем значении $\hat{p}_т$ будет выражаться в виде

$$f(p, \hat{p}) = n\left(p, \hat{p}_т, \frac{\sigma^2_{p_т}}{k}\right). \tag{8.31}$$

Применяя теорему Байеса и подставляя в нее выражения (8.30) и (8.31), получаем апостериорное распределение для p:

$$h(p/\beta) = \frac{1}{\sqrt{2\pi\sigma^2_{p_т}/(c+k)}} \exp\left(-\left(p - \frac{c\hat{p}_а + k\hat{p}_т}{c+k}\right)\left(\frac{c+k}{2\sigma^2_{p_т}}\right)\right),$$

где $c = n\sigma^2_{p_т}/\sigma^2_{p_а}$; n – объем испытаний на этапе априорных исследований, т.е. апостериорное распределение ВБР является нормальным распределением, а байесовская оценка ВБР

$$p_б = \frac{c\hat{p}_а + k\hat{p}_т}{c+k}. \tag{8.32}$$

В формулах (8.30)–(8.32) необходимо определить величину $\sigma^2_{p_а}$. В силу симметрии доверительного интервала относительно величины можно записать, что $p_в = \hat{p}_а + \varepsilon$, $p_н = \hat{p}_а - \varepsilon$. Применяя формулу для доверительного интервала нормально распределенной случайной величины, получаем

$$p_д = P(|\hat{p}_а - p| < \varepsilon) = \Phi\left(\frac{\varepsilon}{\sqrt{2}\,\sigma_{p_а}}\right).$$

Из данного уравнения находим значение $\sigma_{p_а}$:

$$\sigma_{p_а} = \frac{\varepsilon}{\sqrt{2}\,\Phi^{-1}(p_д)} = \frac{p_в - p_н}{2\sqrt{2}\,\Phi^{-1}(p_д)},$$

где $\Phi^{-1}(p_д)$ – функция, обратная функции Лапласа. Таким образом, определены все величины, входящие в (8.30) – (8.32).

3. Пусть априорная информация присутствует в виде оценки ВБР объекта за время $T_р - \hat{p}_а(T_р)$. Известна нижняя доверительная граница с доверительной вероятностью $p_д = P(p_н \le p)$. Такой вид задания априорной информации имеет место для высоконадежных объектов, когда ВБР близка к единице. Текущая информация получена в результате испытаний объектов за время $T_р$, в ходе которых из k испытываемых устройств m устройств отказывает. Вместо характеристики вероятности безотказной работы p будем пользоваться характеристикой вероятности отказа (ВО) – q. При этом нижний доверительный интервал для ВБР заменим на верхний доверительный интервал для ВО $q_в = 1 - p_н$ с доверительной вероятностью

$$P(q \le q_в) = P(p \ge p_н),\ \hat{q}_а = 1 - \hat{p}_а.$$

В качестве априорной плотности распределения вероятности отказа в этом случае можно использовать гамма-распределение

$$h(q)=\frac{\lambda(\lambda q)^{\alpha-1}}{\Gamma(\alpha)}\exp(-\lambda q),$$

а в качестве функции правдоподобия – распределение Пуассона

$$f(q,m,k)=\frac{(kq)^m}{m!}\exp(-kq). \qquad (8.33)$$

Тогда апостериорная плотность распределения ВО

$$h_{\text{апост}}(q/\hat{q})=\frac{q^{\alpha+m-1}\exp\left(-q(\lambda+k)\right)}{\int\limits_0^1 q^{\alpha+m-1}\exp\left(-q(\lambda+k)\right)dq}. \qquad (8.34)$$

Подставляя в (8.28), (8.29) выражение для апостериорной плотности (8.34), получаем байесовскую оценку ВО

$$\hat{q}_{\text{б}}=\frac{\int\limits_0^1 q^{\alpha+m}\exp\left(-q(\lambda+k)\right)dq}{\int\limits_0^1 q^{\alpha+m-1}\exp\left(-q(\lambda+k)\right)dq}; \qquad (8.35)$$

точность оценки ВО

$$D(\hat{q}_{\text{б}})=\frac{\int\limits_0^1 q^{\alpha+m+1}\exp\left(-q(\lambda+k)\right)dq}{\int\limits_0^1 q^{\alpha+m-1}\exp\left(-q(\lambda+k)\right)dq}-\hat{q}_{\text{б}}^2. \qquad (8.36)$$

В общем случае интегралы, стоящие в (8.34)–(8.36), в элементарных функциях не представляются; решение получают численными методами. Для целых α интегралы принимают следующий вид:

$$I=\int\limits_0^1 q^{\alpha}\exp(-qb)dq=\exp(-b)\left[-\frac{1}{b}-\frac{\alpha}{b^2}-\ldots-\frac{\alpha!}{b^{\alpha}}\right]+\frac{\alpha!}{b^{\alpha+1}}\left(1-\exp(-b)\right),$$

где $b=\lambda+k$;

$$\alpha=\begin{cases} m+\alpha-1 \ \text{для} \ I_1=\int\limits_0^1 q^{m+\alpha-1}\exp\left(-q(\lambda+k)\right)dq; \\ m+\alpha \ \text{для} \ I_2=\int\limits_0^1 q^{m+\alpha}\exp\left(-q(\lambda+k)\right)dq; \\ m+\alpha+1 \ \text{для} \ I_3=\int\limits_0^1 q^{m+\alpha+1}\exp\left(-q(\lambda+k)\right)dq. \end{cases}$$

Параметры α и λ, входящие в эти выражения, получают из следующих соотношений. Для гамма-распределения известно, что

$$M[q]=\hat{q}_{\text{а}}=\alpha/\lambda. \qquad (8.37)$$

Запишем выражение для доверительной вероятности

$$P(q_{\text{в}}\geq q)=\int\limits_0^{q_{\text{в}}}\frac{\lambda(\lambda q)^{\alpha-1}}{\Gamma(\alpha)}\exp(-\lambda q)dq. \qquad (8.38)$$

Из (8.37) выразим α через λ и $\hat{q}_{\text{а}}$ и подставим в (8.38), получим

$$P(q_{\text{в}}\geq q)=\int\limits_0^{q_{\text{в}}}\frac{\lambda(\lambda q)^{\lambda\hat{q}_{\text{а}}-1}}{\Gamma(\lambda\hat{q}_{\text{а}})}\exp(-\lambda q)dq. \qquad (8.39)$$

Решая уравнение (8.39), определим значение параметра λ.

4. Проанализируем еще одну задачу байесовского оценивания ВБР, являющуюся частным случаем изложенной выше, в следующей постановке. Пусть имеется априорная информация в виде оценки ВБР $\hat{q}_{\text{а}}(T_{\text{р}})$ с известной точностью $\sigma_{\text{а}}$. В результате текущих наблюдений за группой однотипных объектов зафиксировано m отказов из k испытываемых образцов.

Предположим, что ВО в качестве априорной плотности распределения имеет гамма-распределение

$$h(q)=\frac{q^{\alpha-1}}{(\alpha-1)!}\lambda^{\alpha}\exp(-\lambda q).$$

В данном случае апостериорная плотность примет вид (8.34), а байесовская оценка и точность байесовской оценки будут определяться по (8.35), (8.36). В отличие от предыдущего случая оценивания параметры α и λ будем определять по следующим соотношениям:

$$\lambda=q_{\text{а}}/\sigma_{\text{а}}^2;\ \alpha=q_{\text{а}}^2/\sigma_{\text{а}}^2. \qquad (8.40)$$

Если априорная информация присутствует в виде испытаний $n_а$ образцов некоторого изделия, из которых за время $T_р$ $m_а$ образцов отказывает, то для учета такой информации необходимо сначала определить $q_а$ и $\sigma_а^2$ по формулам

$$q_а = 1 - \frac{m_а}{n_а}; \ \sigma_а^2 = \frac{m_а}{n_а^2}\left(1 - \frac{m_а}{n_а}\right)$$

и затем воспользоваться выражениями (8.34) – (8.36) и (8.40). Так, в частности из (8.40) получим, что $\lambda = n^2/m_а$, $\alpha = n^2/m_а - n$.

5. Известен интервал, в котором заключено значение оцениваемого параметра ВБР p. Результаты текущих исследований представлены в виде испытаний k объектов, из которых за время $T_р$ отказало m образцов.

В случае, когда по априорным данным определен интервал в виде нижней и верхней границ и отсутствуют данные, указывающие наиболее вероятное значение ВБР, можно считать в пределах этого интервала $(p_н, p_в)$ любое значение p возможным, т.е. до проведения текущих исследований все значения p из интервала $(p_н, p_в)$ равновероятны и априорная плотность распределения будет иметь вид

$$h(p) = \begin{cases} 0; \ 0 \le p < p_н, \\ \dfrac{1}{p_в - p_н}; \ p_н \le p \le p_в, \\ 0; \ p_в < p \le 1. \end{cases}$$

Результаты текущих исследований можно представить в виде испытаний по схеме Бернулли. В этом случае отношение правдоподобия запишем в виде

$$f(p, m, k) = C_k^m p^{k-m}(1-p)^m.$$

Тогда по формуле Байеса апостериорная плотность запишется следующим образом:

$$h_{апост}(p/\hat{p}) = \frac{p^{k-m}(1-p)^m}{\int\limits_{p_н}^{p_в} p^{k-m}(1-p)^m \, dp};$$

байесовская оценка ВБР будет иметь вид

$$\hat{p}_б = \frac{\int\limits_{p_н}^{p_в} p^{k-m+1}(1-p)^m \, dp}{\int\limits_{p_н}^{p_в} p^{k-m}(1-p)^m \, dp}; \qquad (8.41)$$

точность байесовской оценки

$$\sigma_{p_б}^2 = \frac{\int\limits_{p_н}^{p_в} p^{k-m+2}(1-p)^m \, dp}{\int\limits_{p_н}^{p_в} p^{k-m}(1-p)^m \, dp} - \hat{p}_б^2. \qquad (8.42)$$

Вычисление интегралов, входящих в (8.41) и (8.42), после подстановки численных значений k и m не вызывает особых затруднений.

Изложенные процедуры оценивания характеристик надежности объектов могут найти применение при расчетах показателей надежности по результатам эксплуатации, в первую очередь, электронных приборов и устройств, входящих в состав штатного оборудования систем управления и защиты, а также контрольно-измерительных приборов и автоматики объектов повышенного риска, например, таких как энергоблоки атомных станций. В качестве априорной информации необходимо использовать информацию, приведенную в паспортах или технических описаниях на соответствующие устройства. Анализ этих документов показывает, что в них в большинстве случаев приводится информация в виде либо доверительного интервала, либо доверительного интервала с указанием доверительной вероятности.

8.12. Оценивание вероятности отказа объектов при биномиальном распределении результатов испытаний

Рассмотрим в данном параграфе еще один подход к определению показателей надежности элементов с учетом априорной информации. Пусть проводятся испытания группы изделий объема k. В результате испытаний k образцов в течение времени $T_р$ зарегистрировано m отказавших изделий. Требуется определить вероятность отказа изделия при условии, что имеется априорная информация в виде результатов испытаний n изделий в течение времени $T_р$, в ходе которых $m_а$ образцов отказало.

Если размеры партии велики по отношению к размеру любой рассматриваемой выборки, то для расчета вероятности отказа изделия (вероятности того, что в результате испытания k образцов m из них откажет) можно использовать биномиальное распределение [43]

$$f(q,m,k)=\frac{k!}{m!(k-m)!}q^{m}(1-q)^{k-m}.$$

В качестве априорного распределения вероятности отказа в данном случае целесообразно использовать β-распределение

$$h(q)=\frac{(n-1)!}{(m_{а}-1)!\,(n-m_{а}-1)!}q^{m_{а}-1}(1-q)^{n-m_{а}-1}.$$

Это следует из результатов п. 8.6, где было показано, что биномиальное распределение и β-распределение являются сопряженными. Априорная оценка вероятности отказа (ВО) в этом случае может быть получена методом моментов и равна

$$\hat{q}_{а}=\int_{0}^{1}qh(q)\,dq.$$

Данное выражение можно переписать в виде

$$\hat{q}_{а}=\frac{m_{а}}{n}\int_{0}^{1}\frac{n!}{m_{а}!\,(n-m_{а}-1)!}q^{m_{а}}(1-q)^{n-m_{а}-1}\,dq.$$

Поскольку подынтегральное выражение само по себе является β-распределением, то этот интеграл равен единице, следовательно, получаем

$$\hat{q}_{а}=\frac{m_{а}}{n}.$$

Априорная дисперсия величины q оказывается равной

$$\sigma^{2}_{q_{а}}=\frac{m_{а}(n-m_{а})}{n^{2}(n+1)}.$$

Определим теперь апостериорную плотность распределения величины q, используя теорему Байеса:

$$h_{апост}(q/\hat{q})=\frac{q^{m_{а}+m-1}(1-q)^{n+k-m_{а}-m-1}}{\int_{0}^{1}q^{m_{а}+m-1}(1-q)^{n+k-m_{а}-m-1}\,dq}.$$

Значение интеграла в этом случае равно

$$I_{\beta}=\frac{(m_{а}+m-1)!\,(n+k-m_{а}-m-1)!}{(n+k-1)!}.$$

Отсюда видно, что апостериорное распределение ВО само является β-распределением с параметрами

$$m_{апост}=m_{а}+m\,;\quad n_{апост}=n+k.$$

Таким образом, можно определить байесовскую оценку ВО:

$$\hat{q}_{б}=(m_{а}+m)/(n+k).$$

Дисперсия байесовской оценки равна

$$\sigma^{2}_{q_{б}}=\frac{(m_{а}+m)\,(n+k-m_{а}-m)}{(n+k)^{2}(n+k+1)}.$$

Итак, получили байесовскую оценку ВО и оценили точность байесовской оценки ВО.

Подведем некоторые итоги. В последних двух главах рассмотрены параметрические методы, с помощью которых производится оценивание характеристик модели, описывающей объекты, имеющие случайную природу. Изложенные процедуры позволяют как получить оценки параметров модели, так и рассчитать точность произведенного оценивания. Изложены байесовские процедуры оценивания, которые позволяют повышать достоверность расчетов за счет использования дополнительных видов информации. Причем, если у исследователя имеется информация, полученная более чем на двух этапах наблюдения, то байесовские процедуры позволяют учитывать все виды наблюдения с использованием процедур последовательного учета накапливаемой информации. Оценивание точностных характеристик параметров модели позволяет в дальнейшем исследовать вопросы адекватности построения моделей, анализировать неопределенность моделей.

Глава 9

НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ АНАЛИЗА СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ

9.1. Общие замечания

Рассмотренные ранее параметрические методы анализа статистической информации в последнее время подвергаются критике специалистов по математической статистике. Дело в том, что применение параметрических методов требует предположений о виде закона распределения наблюдаемых случайных величин. Как правило, нельзя указать какие-либо веские причины, по которым конкретное распределение результатов наблюдений должно входить в то или иное параметрическое семейство. Известны результаты, согласно которым сумма независимых случайных величин описывается нормальным законом распределения; произведение таких величин приближается логарифмически нормальным распределением. В теории надежности доказано, что наработки до отказа однотипных элементов подчиняются экспоненциальному закону распределения, при этом характер отказов предполагается внезапным. Далее известно, что сумма случайных величин, подчиняющихся экспоненциальному закону распределения, распределена по гамма-закону. Пожалуй, этими фактами и исчерпываются все обоснования в поддержку того или иного семейства законов распределения. В подавляющем большинстве реальных ситуаций подобных обоснований для использования конкретного вида закона распределения нет, и приближение реального распределения с помощью параметрических семейств становится чисто формальной процедурой.

Первые публикации, в которых развивался непараметрический подход, относятся к первой половине 20-го столетия. В это время появились работы Кендалла и Спирмена, в которых исследовались критерии проверки гипотез, основанные на коэффициентах ранговой корреляции. В настоящее время эти критерии носят название авторов, разработавших и исследовавших их (см. [35, с. 77–82]). В 30-е годы появились работы А.Н. Колмогорова и Н.В. Смирнова, в которых были предложены и изучены статистические критерии, основанные на использовании эмпирического процесса. Но непараметрические методы, в которых не делается никаких нереалистических предположений о том, что функция распределения результатов наблюдений принадлежит тем или иным параметрическим семействам распределений, стали заметной частью статистического анализа только во второй трети XX в.

После Второй мировой войны развитие непараметрических статистических моделей пошло быстрыми темпами. Большую роль в развитии этих методов сыграли работы Вилкоксона и его школы. К настоящему времени с помощью непараметрических методов можно решать тот же круг задач, что и с помощью параметрических. Все большую роль играют непараметрические методы оценки плотности, непараметрические подходы в решении задач регрессионного анализа и теории распознавания образов. В настоящем разделе рассмотрим непараметрические методы оценки плотности распределения. Построив плотность распределения, можно далее переходить к определению широкого круга статистических показателей. Поэтому задача оценки плотности распределения наблюдаемой случайной величины является одной из ключевых задач статистического анализа.

9.2. Гистограммный метод восстановления плотности распределения

Для придания наглядности статистическому материалу его необходимо подвергнуть дополнительной обработке. С этой целью строится статистический ряд. Покажем, как осуществляется его построение.

Пусть имеются результаты наблюдения над непрерывной случайной величиной X, оформленные в виде простой статистической совокупности. Рассмотрим весь диапазон зафиксированных значений величины X и разделим его на интервалы. Диапазон зафиксированных значений случайной величины представляет собой область определения данной величины. Далее подсчитаем количество значений реализовавшейся случайной величины, попавших в каждый интервал, обозначим эти значения через m_i. Каждое из полученных значений разделим на общее число наблюдений n и определим частоту попадания случайной величины в i-й интервал наблюдения:

$$\hat{p}_i = \frac{m_i}{n}.$$

Сумма частот всех интервалов должна быть равна единице. Представив полученные результаты расчетов в виде таблицы, получим статистический ряд (см. табл. 9.1).

Таблица 9.1

I_i	$x_1; x_2$	$x_2; x_3$	...	$x_i; x_{i+1}$	...	$x_k; x_{k+1}$
$\hat{p}_i$	$\hat{p}_1$	$\hat{p}_2$	...	$\hat{p}_i$	...	$\hat{p}_k$

Здесь I_i – обозначение i-го интервала; x_i; x_{i+1} – границы данного интервала; $\hat{p}_i$ – соответствующая частота; k – количество интервалов.

При построении статистического ряда возникает вопрос о рекомендуемом количестве интервалов разбиения области определения наблюдаемой случайной величины. С одной стороны, количество интервалов не должно быть слишком большим, в этом случае ряд распределения становится невыразительным и частоты в нем обнаруживают незакономерные колебания; с другой стороны, оно не должно быть слишком малым, при малом числе интервалов свойства распределения описываются статистически слишком грубо. Чем богаче и однороднее статистический материал, тем большее число интервалов можно выбирать при составлении статистического ряда. В математической статистике известна формула Стаджесса, с помощью которой вычисляется количество интервалов разбиения области определения случайной величины. Согласно этой формуле количество интервалов определяется следующим образом:

$$k = 1 + 3{,}3 \lg n.$$

При построении статистического ряда возможны различные способы выбора длины интервалов; они могут быть как равными, так и различными. Однако следует отметить, что в практике построения статистического ряда наибольшее применение нашли два: метод равных интервалов и равночастотный метод. В первом методе, естественно, длины интервалов выбираются одинаковыми. Во втором методе длины интервалов различные. Они выбираются таким образом, чтобы количество попаданий случайной величины в каждый из интервалов было одним и тем же.

Графическое представление статистического ряда называется гистограммой. Гистограмма строится следующим образом. По оси абсцисс откладываются интервалы и на каждом из них строится прямоугольник, площадь которого равна частоте данного интервала. Для построения гистограммы необходимо частоту каждого интервала разделить на его длину и полученное значение взять в качестве высоты прямоугольника. В случае равных интервалов высоты прямоугольников пропорциональны соответствующим частотам. Из правила построения гистограммы следует, что полная площадь под гистограммой равна единице. Формулу построения гистограммы можно представить в следующем виде:

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{I\left[x_i \in A_{nj}\right]}{\lambda(A_{nj})},$$

где $I\left[x_i \in A_{nj}\right]$ – индикатор, равный единице, если условие в скобках выполняется, и нулю – в противном случае; A_{nj} – интервалы, на которые делится область определения наблюдаемой случайной величины; $\lambda(A_{nj})$ – ширина интервала A_{nj}.

Гистограмма является простейшей оценкой плотности распределения, не обладающей свойством несмещенности. Гистограммная оценка обладает рядом недостатков. В первую очередь, необходимо отметить значительную потерю информации, связанную с тем, что исследователю для построения гистограммы необходимо знать, сколько наблюдений попало в выбранный интервал разбиения и абсолютно не важны при этом значения наблюдений. Во-вторых, принцип разбиения (равных интервалов или равных частот), а также число разбиений, являются некоторыми «степенями свободы». Как исследователь выполнит разбиение, воспользовавшись такой свободой, зависит только от него. В-третьих, скорость сходимости гистограммной оценки к плотности крайне низкая.

Рассмотрим пример построения гистограммы по результатам наблюдения за случайной величиной, характеризующей наработки до отказа группы однотипных объектов. Пусть результаты функционирования группы однотипных объектов представлены в виде статистического ряда, приведенного в табл.9.2.

Таблица 9.2

I_i	$x_1; x_2$	$x_2; x_3$	...	$x_i; x_{i+1}$	...	$x_{k-1}; x_k$	$x_k; x_{k+1}$
m_i	2	1		N_i		0	1
$\hat{p}_i$	$2/n$	$1/n$	...	N_i/n		0	$1/n$

На основании данной таблицы построим гистограмму (рис.9.1).

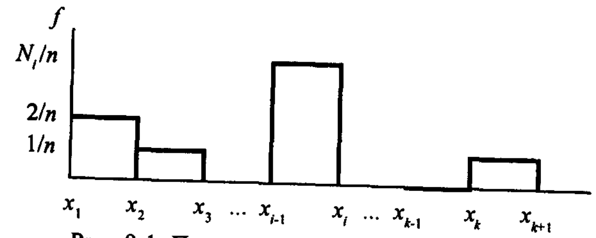

Рис. 9.1. Пример построения гистограммы

Пользуясь результатами построения статистического ряда, можно построить также эмпирическую функцию распределения наблюдаемой случайной величины. Для этого необходимо определить значения функции распределения на границах интервалов, по которым построена гистограмма, тогда получим

$$\hat{F}(0,x_1)=0;$$
$$\hat{F}(x_1,x_2)=\hat{p}_1;$$
$$\hat{F}(x_2,x_3)=\hat{p}_1+\hat{p}_2;$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\hat{F}(x_{k-1},x_k)=\sum_{i=1}^{k}\hat{p}_i;$$
$$\hat{F}(x_k,\infty)=\sum_{i=1}^{k}\hat{p}_i=1.$$

Эмпирическая функция распределения будет представлять собой ступенчатую функцию, изображенную на рис. 9.2. Построение эмпирической функции распределения решает задачу описания статистического материала. На основании данной функции можно производить оценивание вероятностных характеристик объектов, для которых ведется обработка статистического материала.

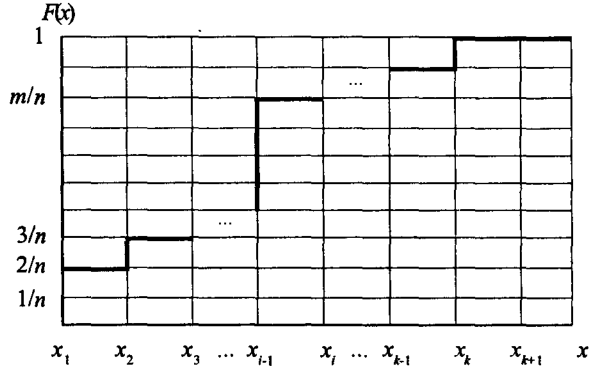

Рис. 9.2. Эмпирическая функция распределения

9.3. Построение эмпирической функции распределения по цензурированной выборке

Анализ работ по непараметрическим методам оценивания характеристик сложных систем на основании цензурированных данных показывает, что достаточно полно разработаны и исследованы методы получения точечных и интервальных оценок для случаев, когда у исследователя имеются большие объемы статистических данных.

В [38] изложены методы построения оценки функции $F(t)$, исследованы свойства этой оценки, рассмотрены методы оценивания некоторых показателей надежности на основании цензурированных данных незначительного объема. При оценивании характеристик объектов на основании информации, полученной на этапе их эксплуатации в составе штатного оборудования сложных систем, возникают ситуации, аналогичные рассмотренным в [38], а именно, распространенной является ситуация, когда информация представлена в виде цензурированных данных весьма ограниченного объема. Поэтому методы, изложенные в [38], можно эффективно применять при проведении экспресс-анализа характеристик объектов сложных систем на этапе проведения системных исследований. Рассмотрим метод построения эмпирической функции распределения, изложенный в [38].

Пусть имеется функция распределения $F(t)$, тогда вероятность попадания наблюдаемой случайной величины в интервал $(0, T)$ будет равна $F(T)$. Разобьем интервал наблюдения на k равных частей. Определим вероятность попадания наблюдаемой случайной величины в интервал $(0, \xi_j]$ для любого $j=\overline{1,k}$.

Попадание случайной величины на некоторый интервал $(0, \xi_j]$ есть событие, которое можно представить как сумму событий, состоящих в том, что случайная величина попадет либо в интервал $(0, \xi_{j-1}]$, либо в интервал $(\xi_{j-1}, \xi_j]$. Следовательно, для вероятности попадания случайной величины в интервал $(0, \xi_j]$ можно записать выражение

$$Q(\xi_j) = P\{T \le \xi_j\} = P\{T \le \xi_{j-1}\} + P(\xi_{j-1} \le T \le \xi_j\}, \qquad (9.1)$$

где T – значение наблюдаемой случайной величины. Вероятность попадания случайной величины на интервал $(\xi_{j-1}, \xi_j]$ можно определить в виде

$$P\{\xi_{j-1} < T \le \xi_j\} = (1 - F_{j-1})Q(\xi_{j-1}, \xi_j / \xi_{j-1}), \qquad (9.2)$$

где F_{j-1} – значение функции $F(t)$ в точке ξ_{j-1}, $F_{j-1} = Q(\xi_{j-1})$; $Q(\xi_{j-1}, \xi_j/\xi_{j-1})$ – условная вероятность попадания случайной величины в интервал $(\xi_{j-1}, \xi_j]$ при условии, что на интервале $(0, \xi_{j-1}]$ случайная величина T не реализовалась.

Подставляя выражение (9.2) в (9.1), получаем формулу для определения вероятности попадания случайной величины T в интервал $(0, \xi_j]$

$$Q(\xi_j) = Q(\xi_{j-1}) + (1 - Q(\xi_{j-1}))Q(\xi_{j-1}, \xi_j / \xi_{j-1}), \qquad (9.3)$$

где $Q(\xi_{j-1})$ – вероятность попадания случайной величины в интервал $(0, \xi_{j-1}]$. Расписывая аналогичным образом вероятность попадания случайной величины в интервал $(0, \xi_{j-1}]$ через вероятности попадания в интервалы $(0, \xi_{j-2}]$ и $(\xi_{j-2}, \xi_{j-1}]$, получаем

$$Q(\xi_{j-1}) = Q(\xi_{j-2}) + (1 - Q(\xi_{j-2}))Q(\xi_{j-2}, \xi_{j-1} / \xi_{j-2}).$$

Повторяя данную процедуру для интервалов $(0, \xi_i]$, где $1 < i < j$, можно в конечном счете получить выражение для вероятности попадания наблюдаемой случайной величины в интервал $(0, T)$ в виде

$$F(T) = \sum_{j=1}^{k}(1 - Q(\xi_{j-1}))Q(\xi_{j-1}, \xi_j / \xi_{j-1}). \qquad (9.4)$$

Оценим функцию распределения для многократно цензурированной справа выборки следующего вида: $T_1, T_2, ..., T_\mu, T'_{\mu+1}, ... T'_{\mu+\nu}$, где $T_i, i = \overline{1, \mu}$ – значения наблюдаемой случайной величины с реализовавшимся признаком; T'_j, $j = \overline{\mu+1, \mu+\nu}$ – значения наблюдаемой случайной величины с нереализовавшимся признаком, т.е. цензурированные данные.

Разобьем результаты наблюдения на k интервалов:

1-й интервал;	2-й интервал;	... ;	k-й интервал;
$T_1, ..., T_{\mu_1}$;	$T_{\mu_1+1}, ..., T_{\mu_1+\mu_2}$;	... ;	T_μ;
$T'_1, ..., T'_{\nu_1}$;	$T'_{\nu_1+1}, ..., T'_{\nu_1+\nu_2}$;	... ;	T'_ν;

где μ_j – количество наблюдений случайной величины с реализовавшимся признаком, попавших в j-й интервал; ν_j – количество наблюдений случайной величины с нереализовавшимся признаком, попавших в j-й интервал,

$$\sum_{j=1}^{k}\mu_j = \mu; \ \sum_{j=1}^{k}\nu_j = \nu.$$

Для определения оценки функции распределения в точке t необходимо вместо значений величин $Q(\xi_{j-1})$, $Q(\xi_{j-1}, \xi_j/\xi_{j-1})$ в (9.4) подставить их оценки. В качестве условных вероятностей используют отношение количества реализаций μ_j, попавших на интервал $(\xi_{j-1}, \xi_j]$, к общему числу объектов, находящихся под наблюдением в начале этого интервала при условии, что в начале интервала под наблюдением находится хотя бы один объект. Если в начале интервала под наблюдением отсутствует хотя бы один объект, то условная вероятность наблюдения случайной величины на этом интервале равна нулю. Поэтому оценку условной вероятности на интервале полагают также равной нулю. С учетом этого выражение для оценки функции распределения в точке t будет иметь вид

$$\begin{cases} \sum_{j=1}^{k}\left(1 - \hat{F}_{j-1}\right)\dfrac{\mu_j}{N_{yj}} \text{ при } N_{yj} > 0, \\ 0 \text{ при } N_{yj} \le 0, \end{cases} \qquad (9.5)$$

где $\hat{F}_{j-1}$ – оценка функции распределения в точке ξ_{j-1};

$$N_{yj} = N - \sum_{i=1}^{j-1}\mu_i + \nu_i. \qquad (9.6)$$

Если в первом интервале одна или несколько случайных величин с нереализовавшимся признаком меньше первой случайной величины с реализовавшимся признаком, то $\nu_0 \ne 0$, в противном случае $\nu_0 = 0$. Если в последнем интервале значение последней случайной величины с реализовавшимся признаком больше всех значений случайных величин с нереализовавшимся признаком, то $\nu_k = 0$, в противном случае $\nu_k \ne 0$.

Величина N_{yj}, входящая в выражение (9.5), называется условным объемом цензурированной выборки на интервале $(\xi_{j-1}, \xi_j]$. Она равна числу объектов, за которыми ведется наблюдение на указанном интервале.

Выражение (9.5) можно переписать, несколько детализировав представление функции распределения на каждом интервале:

$$\hat{F}(t)=\begin{cases} 0 \text{ при } t\le 0; \\ \dfrac{i_1}{N_{y1}} \text{ при } 0\le t\le \xi_1,\ i_1=0,1,\ldots,\mu_1; \\ \left.\begin{matrix} \hat{F}_1+(1-\hat{F}_1)\dfrac{i_2}{N_{y2}} \text{ при } N_{y2}>0 \\ \hat{F}_1 \text{ при } N_{y2}\le 0 \end{matrix}\right\} \text{для } \xi_1<t\le\xi_2;\ i_2=0,1,\ldots,\mu_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \left.\begin{matrix} \hat{F}_{k-1}+(1-\hat{F}_{k-1})\dfrac{i_k}{N_{yk}} \text{ при } N_{yk}>0 \\ \hat{F}_{k-1} \text{ при } N_{y2}\le 0 \end{matrix}\right\} \text{для } \xi_{k-1}<t\le\xi_k;\ i_k=0,1,\ldots,\mu_k. \end{cases} \tag{9.7}$$

В [38] даются рекомендации для разбиения области, на которой проводится оценивание функции распределения, на интервалы. В частности, указано, что для повышения точности оценки $F(t)$ необходимо весь отрезок наблюдения $[0, T]$, где $T=\xi_k=\max[T_\mu, T'_\nu]$, разбить на интервалы, границы которых совпадают со значениями случайных величин с нереализовавшимся признаком. Например, если выборку наблюдений, состоящую из случайных величин с нереализовавшимся признаком и случайных величин с реализовавшимся признаком записать в порядке возрастания ее членов в виде

$$T_1,\ T_2,\ T'_1,\ T_3,\ T'_2,\ldots,\ T'_\nu,\ T_{\mu-1},\ T_\mu\ , \tag{9.8}$$

то интервалы наблюдений будут равны

$$[0,\ T'_1],(T'_1,\ T'_2],\ldots,(T'_{\nu-1},\ T'_\nu],(T'_\nu,\ T_\mu],$$

а μ_j, ν_j соответственно будут равны

$$\mu_1=2,\ \nu_1=1;\ \mu_2=1,\ \nu_2=1;\ldots;\ \mu_k=2,\ \nu_k=0\,.$$

Если в выборке наблюдений окажется, что отдельные значения случайных величин с реализовавшимся признаком будут равны некоторым значениям случайных величин с нереализовавшимся признаком, то в вариационном ряду (9.8) сначала указывают значения случайных величин с реализовавшимся признаком, затем значения случайных величин с нереализовавшимся признаком.

Таким образом, используя правило (9.7), можно строить эмпирическую функцию распределения случайной величины при многократно цензурированных выборках.

9.4. Ядерная оценка плотности

Гистограммные оценки плотности распределения, рассмотренные в предыдущих параграфах, обладают существенным недостатком, а именно, плотность или функция распределения, полученные данным методом, являются ступенчатыми функциями. Реальные функции распределения являются непрерывными функциями. Следовательно, гистограммные оценки хорошо аппроксимируют функции распределения только в случае больших объемов наблюдений, когда $n\to\infty$. В ситуации, когда наблюдения производят за функциональными характеристиками сложных систем, хотелось бы иметь более гладкие оценки плотности или функции распределения. Шагом в получении такой оценки явилась модель построения так называемой ядерной оценки плотности.

Ядерные оценки впервые были введены в работах Парзена [44] и Розенблатта [45]. Рассмотрим методику построения ядерных оценок для плотности распределения непрерывной случайной величины. Пусть в результате наблюдения за объектом исследования получена выборка $T_1, T_2, \ldots, T_n$. На основании данных результатов построим гистограммную оценку функции распределения. Гистограммную оценку функции распределения будем строить несколько отличающимся способом по сравнению с тем, как это было сделано в п. 9.2. Будем полагать, что изменение функции распределения происходит в каждой точке наблюдения, причем величина такого изменения равна $1/n$. Построенная таким образом функция распределения изображена на рис. 9.3. Ее можно записать в виде

$$F_n(t)=P(T_i<t)=\frac{1}{n}\sum_{i=1}^{n}I(T_i\le t)=\frac{1}{n}\sum_{i=1}^{n}H(T_i)\,, \tag{9.9}$$

где $I(T_i\le t)$ – индикаторная функция,равная 1, когда условие в скобках выполняется, и 0 в противном случае; $H(T_i)$ – функция Хевисайта, равная 1 при $t\ge T_i$, и 0 при $t<T_i$.

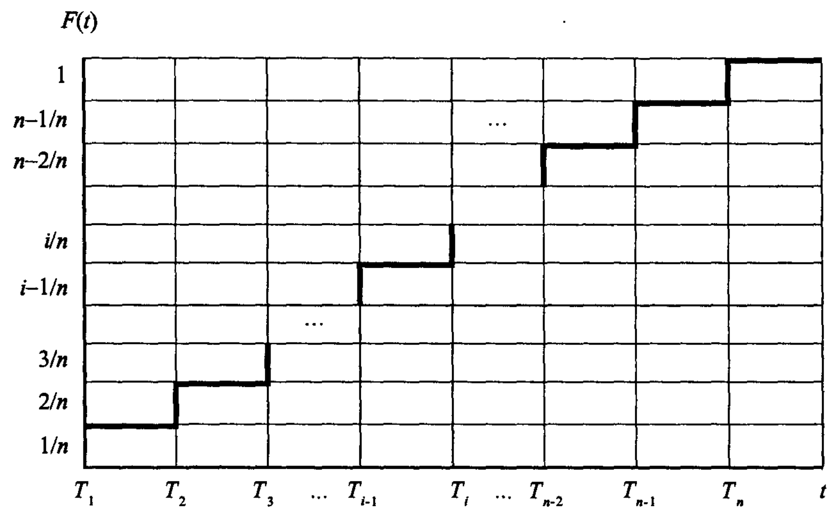

Рис. 9.3. Эмпирическая функция распределения

На основании выражения (9.9) определим плотность распределения как производную от функции распределения:

$$f_n(t) = F_n'(t) = \frac{1}{n}\sum_{i=1}^{n} H_i'(T_i) = \frac{1}{n}\sum_{i=1}^{n}\delta(T_i), \qquad (9.10)$$

где $\delta(T_i)$ – дельта-функция Дирака, обладающая свойствами

$$\delta(T_i) = \begin{cases} \infty \text{ при } t = T_i; \\ 0 \text{ при } t \neq T_i; \end{cases} \qquad (9.11)$$

$$\int_{-\infty}^{\infty}\delta(t)dt = 1, \quad \int_{-\infty}^{\infty}\delta(t-u)f(u)du = f(t).$$

Функция, определяемая выражением (9.10), называется ядерной оценкой плотности, а функция, стоящая под знаком суммы, называется ядром. Графическое изображение такой плотности распределения будет представлять собой набор вертикальных прямых, берущих начало в точках T_i и уходящих в бесконечность. Наглядность такого представления пока не очень высока и не отличается в лучшую сторону по сравнению с гистограммным представлением. Заменим в выражении (9.10) функцию Дирака некоторой конечной функцией, удовлетворяющей свойствам (9.11). Парзен и Розенблатт впервые предложили в качестве ядра вместо функции Дирака использовать функцию $K(t)$, определенную следующим образом:

$$K(t) = \begin{cases} \frac{1}{2} \text{ при } -1 \le t \le 1; \\ 0 \text{ при } t < -1,\ t > 1. \end{cases} \qquad (9.12)$$

Вид функции представлен на рис. 9.4. Если в качестве ядра взять данную функцию и сузить интервал определения, т.е. определить ее на интервале $[-h, h]$, $h < 1$, то получим оценку плотности в виде

$$f_n(t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-T_i}{h}\right),\ h < 1. \qquad (9.13)$$

Плотность распределения, построенная на основании формулы (9.13), будет представлять собой непрерывную функцию. В последующих за исследованиями Парзена и Розенблатта работах было предложено множество других функций, используемых в качестве ядра. Отметим некоторые из них. Так называемая треугольная (рис. 9.5) функция имеет вид

$$Tr = \begin{cases} 0 \text{ при } t < -1, \\ 1+t \text{ при } -1 \le t < 0, \\ 1 \text{ при } t = 0, \\ 1-t \text{ при } 0 < t \le 1, \\ 0 \text{ при } t > 1. \end{cases}$$

Широкое применение при построении ядерных оценок плотности распределения находит функция Гаусса

$$G(t) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right).$$

Изображение функции приведено на рис. 9.6.

Ядерная оценка плотности распределения будет получаться путем подстановки в выражение (9.13) соответствующего ядра, например, $G(t)$ вместо ядра $K(t)$. Так, для гауссовского ядра получим выражение плотности распределения в виде

$$f_n(t) = \frac{1}{nh}\sum_{i=1}^{n} G\left(\frac{t-T_i}{h}\right) = \frac{1}{nh\sqrt{2\pi}}\sum_{i=1}^{n}\exp\left[-\left(\frac{t-T_i}{\sqrt{2}h}\right)^2\right].$$

Качество восстановленной с помощью ядерного оценивания плотности зависит от выбора величины параметра h. Параметр локальнос-

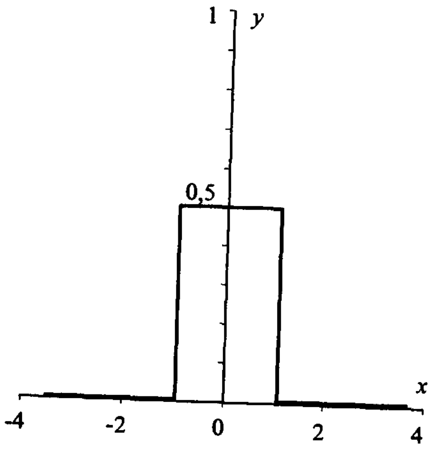

Рис. 9.4. Ядро Парзена-Розенблатта

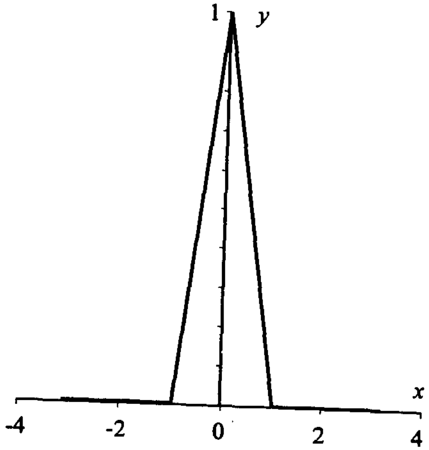

Рис. 9.5. Ядро – треугольная функция

ти h выступает в качестве основного управляющего параметра. Его значение оказывает существенное влияние на вид оценок плотностей распределения и их точность.

Можно показать, что дисперсия оценки $D\hat{f}_n \sim \frac{1}{nh}$. Следовательно, h нельзя брать бесконечно малым, так как при этом дисперсия оценки плотности распределения будет стремиться к бесконечности. С другой стороны, нельзя брать параметр h слишком большим, поскольку при этом увеличивается систематическая ошибка: $M\left(\hat{f}_n - f\right) \sim h$.

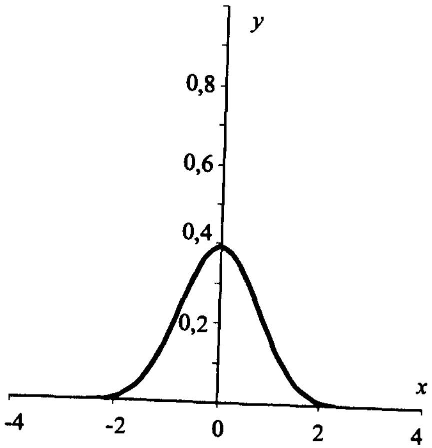

Рис. 9.6. Гауссовское ядро

Таким образом, возникает оптимизационная задача выбора параметра сглаживания. Приведем метод определения оптимального значения параметра h, основанный на вычислении функции правдоподобия. Суть метода состоит в следующем. Пусть имеется выборка $T_1, T_2, \ldots, T_n$. На первом шаге выбираем произвольное значение параметра h_1. Далее исключаем из выборки значение T_1 и на основании оставшихся значений $T_2, T_3, \ldots, T_n$ строим плотность $f_{n-1}(h_1,t)$. Затем определяем значение плотности в точке T_1. В результате получаем $f_{n-1}^1(h_1,T_1)$. На следующем шаге исключаем значение T_2. На основании оставшихся значений $T_1, T_3, \ldots, T_n$ строим плотность $f_{n-1}^2(h_1,t)$. Далее вычисляем значение ядерной оценки плотности в точке T_2. Получаем $f_{n-1}^2(h_1,T_2)$. Повторяем данную процедуру по всем T_i до T_n. Получаем массив $\left\{f_{n-1}^i(h_1,T_i)\right\}$, $i=\overline{1,n}$. На основании вычисленных значений оценок плотностей в точках строим функцию правдоподобия

$$L(h_1,T_i) = \prod_{i=1}^{n} f_{n-1}^i(h_1,T_i).$$

На втором этапе устанавливаем значение константы h, равное h_2, и повторяем описанную процедуру заново. Вычисляем функцию правдоподобия

$$L(h_2,T_i) = \prod_{i=1}^{n} f_{n-1}^i(h_2,T_i) \text{ и т.д.}$$

Оптимальное значение h выбираем как результат решения выражения

$$h_{\text{опт}} = \arg\max_{h_j} L(h_j,T).$$

Данная процедура сложна в реализации, но обеспечивает вычисление оптимального значения параметра h.

В работе [46] приводятся результаты исследования сходимости ядерных оценок. В частности отмечено, что оптимальная скорость сходимости ядерной оценки плотности обеспечивает выбор параметра h на уровне

$$h_n = (\alpha/\beta n)^{1/5}, \qquad (9.14)$$

где $\alpha = \int K^2(t)dt$, $\beta = \left(\int t^2 K(t)dt\right)^2 \int f''(t)^2 dt$. При этом также предполагается, что $h_n \to 0$, $nh_n \to \infty$, при $n \to \infty$ и $f(t)$ – ограниченная плотность, имеющая две непрерывные производные, и $\int (f''(t))^2 dt < \infty$. Данный результат также можно использовать для определения оптимального значения параметра h. В приведенной формуле (9.14) неизвестным показателем является плотность $f(t)$, используемая для определения коэффициента β. Заменим данную плотность ее ядерной оценкой. Поскольку ядерная оценка зависит от параметра h, то для поиска оптимального значения параметра необходимо организовать итеративную процедуру проведения расчетов. На первом этапе выбираем произвольное значение параметра $h_1<1$. Для данного значения строим ядерную оценку плотности $f_n(h_1, t)$ и на ее основании рассчитываем значения коэффициентов α_1, β_1. Далее подставляем полученные значения коэффициентов в формулу (9.14), вычисляем новое значение параметра h_2. На основании данного значения параметра строим новую ядерную оценку $f_n(h_2, t)$. Полученное значение плотности используем вновь для расчета коэффициентов α, β. Повторяем данную процедуру до тех пор, пока не будет выполняться условие сходимости результата вычисления оптимизируемого параметра h, а именно, $|h_m - h_{m-1}| < \varepsilon$, где ε – малое число, определяющее заданную точность проведения расчетов.

9.5. Проекционное оценивание плотности распределения

Следующим шагом в развитии непараметрических методов оценивания плотности распределения непрерывной случайной величины явились проекционные методы. Впервые метод оценивания плотности распределения, получивший название проекционный, был предложен в работе Н.Н. Ченцова [47]. Для построения оценки плотности распределения были использованы результаты теории ортогональных функций. Для получения проекционной оценки Ченцов использовал разложение функции в ряд Фурье. Итак, пусть $f(t)$ – функция, имеющая область определения [0,1]. Следовательно, ее можно разложить в ряд Фурье:

$$f(t) = \sum_{j=1}^{\infty} c_j \varphi_j(t),$$

где $\{\varphi_j(t)\}$ – ортонормированная тригонометрическая система на [0,1].

$$\varphi_{2l}(t) = \sqrt{2}\cos 2\pi lt;$$
$$\varphi_{2l+1}(t) = \sqrt{2}\sin 2\pi lt;$$
$$\varphi_1(t) = 1,\ l \ge 1.$$

Оценка функции $f(t)$ тогда будет определяться по формуле

$$f_{n,N}(t) = \sum_{j=1}^{N} \hat{c}_j \varphi_j(t), \qquad (9.15)$$

т.е. в разложении в ряд Фурье берется конечное число членов суммирования. Оценки коэффициентов в разложении определяются по формуле

$$\hat{c}_j = \frac{1}{n}\sum_{i=1}^{n} \varphi_j(T_i).$$

Оценка (9.15) называется проекционной оценкой Ченцова. Условие $t \in [0, 1]$ не ограничивает общности. Функция $f(t)$ может быть определена на любой ограниченной области $D \in R^1$. Отнормировав данную область, можно обеспечить условие $t \in [0, 1]$. После построения плотности распределения на интервале [0,1] необходимо выполнить обратный переход в область определения функции D.

В выражении (9.15) неопределенной осталась величина N, которая представляет собой число слагаемых в разложении функции в ряд Фурье. Данный параметр называется параметром сглаживания. В [46] отмечается, что тригонометрическая система недостаточно богата для оценивания любых плотностей. Однако этот недостаток уравновешивается рядом преимуществ проекционных оценок, в частности их превосходным поведением в случае, когда разложение плотности распределения в ряд содержит конечное число членов или является бесконечным с быстро убывающими коэффициентами. Известна оценка для определения оптимального числа гармоник N_n состоящая в следующем:

$$N_n^o = \arg\min_{N \subset [1,n]} A_{n,N}, \qquad (9.16)$$

где $A_{n,N} = \rho(N) + \dfrac{N}{2n}$, $\rho(N) = \sum_{j=N}^{\infty} c_j^2$.

В работе [48] показано, что для выбора порядка числа гармоник (9.16) можно пользоваться оценкой

$$N_n = \arg\min_{1<N<n} \tau_N,$$

где $\tau_N = \sum_{j=N+1}^{2N} \hat{c}_j^2$.

В литературе (например [48]) приводится значительное количество ортонормальных систем, которые можно использовать для построения проекционных оценок плотности распределения. Помимо тригонометрической системы это полиномы Лежандра, которые образуют ортонормальную систему на [–1, 1], оценки с рядом Эрмита. Функции с рядом Эрмита образуют ортонормальную систему, определенную на [–∞, ∞]. Оценка с рядом Лагерра образует также ортонормальную систему, определенную на [0, ∞]. Ортонормальная система Хаара отличается от всех предыдущих тем, что она является базисом в области определения [0, 1]. Формулы разложения для данных систем, а также свойства проекционных оценок с этими разложениями приведены, например, в [48].

Рассмотренные в последних трех главах методы обработки статистической информации используются для определения параметров элементов, составных частей и подсистем сложных систем на этапе построения моделей систем. Рассмотренные методы обработки информации претендуют на полноту охвата моделей, используемых при решении задачи статистического оценивания. Так, рассмотрены параметрические методы, которые представлены моделями максимального правдоподобия и байесовскими процедурами, а также непараметрические методы, включающие в себя гистограммные, ядерные и проекционные оценки. Представленный материал преследует цель получения оценок статистических показателей сложной системы с высокой степенью точности и соответственно высокой достоверностью. Высокая степень достоверности оценок достигается за счет использования цензурированных данных, а также за счет использования априорной информации.

Таким образом, в представленном материале рассмотрен комплекс вопросов, касающихся методологии системного анализа, процедуры проведения системных исследований, построения моделей систем. Вопросы построения моделей систем охватывают широкий комплекс проблем, начиная от классификации моделей, подходов к построению имитационных моделей, и заканчивая методами проверки адекватности моделей и оценки параметров систем.

Глава 10

МАТЕМАТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Методы математического программирования представляют собой класс моделей, применяемых для формализации задач планирования целенаправленной деятельности, предусматривающих распределение ограниченного количества ресурсов разных видов. Подобного рода задачи решаются в различных отраслях деятельности: в экономике, при разработке проектов, составлении расписаний, планировании военных операций и т.п. Модели математического программирования относятся к категории детерминированных моделей. Термин программирование в применении к рассматриваемому типу задач понимается как поиск наилучших планов (от английского слова programming – составление плана, программы действий). Когда говорят о задачах математического программирования, имеют в виду задачи, цель которых состоит в повышении эффективности промышленных, транспортных систем, систем управления деятельностью учебных, проектных, научных организаций.

Математическое программирование подразделяется на линейное, целочисленное, нелинейное, динамическое программирование. Рассмотрим некоторые постановки задач, методы и алгоритмы их решения.

10.1. Математические постановки задач, приводящие к моделям линейного программирования

Задачи линейного программирования относятся к категории оптимизационных. Они находят широкое применение в различных областях практической деятельности: при организации работы транспортных систем, в управлении промышленными предприятиями, при составлении проектов сложных систем. Многие распространенные классы задач системного анализа, в частности, задачи оптимального планирования, распределения различных ресурсов, управления запасами, календарного планирования, межотраслевого баланса укладываются в рамки моделей линейного программирования. Несмотря на различные области приложения данные задачи имеют единую постановку: найти значения переменных $x_1, x_2, \ldots, x_n$, доставляющие оптимум заданной линейной формы $z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$ при выполнении системы ограничений, представляющих собой также линейные формы.

Рассмотрим примеры конкретных постановок задач, формализация которых приводит к моделям линейного программирования. Вначале рассмотрим задачу определения оптимального ассортимента. Имеется p видов ресурсов в количествах $b_1, b_2, \ldots, b_i, \ldots, b_p$ и q видов изделий.

Задана матрица $\overline{A} = \|a_{ij}\|$, где a_{ij} характеризует нормы расхода i-го ресурса на единицу j-го изделия (j=1, 2,..., q). Эффективность выпуска единицы j-го изделия характеризуется показателем c_j, удовлетворяющим условию линейности. Требуется определить план выпуска изделий (оптимальный ассортимент), при котором суммарный показатель эффективности принимает наибольшее значение. Обозначим количество единиц j-го изделия, выпускаемых предприятием, через x_j, тогда математическая модель задачи будет иметь следующий вид:

определить максимум линейной формы $\max z = \sum\limits_j c_j x_j$

при ограничениях на ресурсы $\sum\limits_j a_{ij} \le b_i,\ i = 1,\ 2, \ldots,\ p$.

Кроме указанных ограничений по ресурсам в модель могут быть введены дополнительные ограничения на планируемый выпуск продукции $x_j \ge x_{j0}$, условия комплектности изделий и т.п.

При рассмотрении типовых задач системного анализа были рассмотрены задачи распределения ресурса, в числе которых описаны задачи, возникающие при проектировании систем, а именно, задача составления титульного списка и задача определения оптимальной очередности разработки. Приведем их формулировки в развернутой постановке.

Задача составления титульного списка. Сформулирован перечень задач, решаемых на первом этапе автоматизации. После составления перечня задач, включаемых в первый этап разработки, необходимо оценить требуемый состав ресурсов на их разработку и требуемое время для их внедрения. Пусть время, требуемое на разработку задач, превышает заданный срок ввода первой очереди в эксплуатацию, тогда возникает проблема составления титульного списка, т.е. возникает необходимость ограничения перечня задач, автоматизируемых на первом этапе. Проблема выбора комплекса задач из сформированного перечня в условиях дефицита времени и ресурсов на разработку всего

перечня задач, выполняемых на первом этапе автоматизации, называется задачей составления титульного списка. Таким образом, формулировка задачи будет выглядеть так: требуется сформировать перечень задач, подлежащих автоматизации (титульный список), с учетом имеющихся материальных, временных, трудовых и прочих ресурсов.

Формализуем данную задачу. Введем обозначения: пусть i – номер задачи; I – полное их число; T – заданное время разработки и внедрения первой очереди системы; c_i – ожидаемый эффект от внедрения i-й задачи; x_i – переменная, принимающая значение 1, если задача включается в титульный список, и 0 – в противном случае; j – индекс ресурсов, необходимых для разработки и внедрения системы; J – количество видов ресурсов. Далее пусть $r_{ij}(t)$ – потребляемый поток ресурса j-го вида для разработки и внедрения i-й задачи; тогда общая потребность j-го ресурса для i-й задачи будет определяться следующим образом:

$$W_{ij} = \int_0^T r_{ij}(t)dt.$$

Пусть B_j – суммарная величина наличия j-го ресурса в системе. Данный показатель можно определить через величину $R_j(t)$ потока j-го ресурса, поступающего для разработки и внедрения системы ($0 \le t \le T$). Суммарная величина j-го ресурса определяется по формуле

$$B_j = \int_0^T R_j(t)dt.$$

Определим еще одну характеристику $r_{ij}^{\text{в}}$ – максимально допустимое значение потребляемого потока ресурса j-го вида для разработки и внедрения i-й задачи. Данная величина представляет собой верхнюю границу величины $r_{ij}(t)$.

Тогда задачу составления титульного списка можно сформулировать в следующем виде:

максимизировать линейную форму $\sum\limits_{i=1}^{I} c_i x_i$

при ограничениях $\sum\limits_{i=1}^{I} W_{ij} x_i \le B_j,\ j = \overline{1,\ J};\quad \sum\limits_{i=1}^{I} r_{ij}^{\text{в}} x_i \le \max R_j(t),\ j = \overline{1,\ J};$

Первое ограничение требует, чтобы распределяемый ресурс j-го вида не превышал имеющегося количества данного ресурса, выделяемого для решения i-й задачи. Второе ограничение говорит о том, что максимально допустимое значение потребляемого ресурса не должно превышать максимального значения потока ресурса, выделяемого для решения i-й задачи. Таким образом, данная задача представляет собой задачу целочисленного линейного программирования, причем переменная x_i может принимать только два значения: 0 и 1.

Задача определения оптимальной очередности разработки встает перед проектировщиками на следующем этапе проектирования после составления титульного списка задач, подлежащих автоматизации. Суть задачи состоит в распределении ресурсов, выделяемых на разработку системы, между задачами и упорядочении процесса разработки задач во времени. Формализованная постановка данной задачи будет выглядеть следующим образом: необходимо оптимизировать некоторый функционал при выполнении ограничений на потребление ресурсов, выделяемых на разработку проекта, не больше заданного объема в заданном временно́м интервале.

Будем использовать условия и обозначения, введенные при формулировании предыдущей задачи, т.е. будем считать заданными требуемые и выделяемые потоки и суммарные величины ресурса. Кроме того, введем обозначения α_{ij} и β_{ij} – моменты начала и окончания использования j-го ресурса для i-й задачи. При заданных моментах начала и окончания использования j-го ресурса для i-й задачи связь между потоком требуемых ресурсов и его суммарным количеством выражается в виде

$$\int_{\alpha_{ij}}^{\beta_{ij}} r_{ij}(t)dt = W_{ij}.$$

Если моменты начала и окончания использования j-го ресурса для i-й задачи являются неопределенными, то поток ресурсов и его суммарное количество связаны соотношением

$$\int_{0}^{\beta_i} r_{ij}(t)dt = W_{ij},\ i=\overline{1,\ i},\ j=\overline{1,\ J},$$

где $\beta_i = \max\limits_{(i)} \beta_{ij}$ – момент окончания разработки i-й задачи.

Сформулируем ограничения, которые должны иметь место в данной задаче:

ограничение на потребление ресурсов всем проектом

$$\sum_{i=1}^{I} r_{ij}(t) \le R_j(T),\ j=\overline{1,\ J},$$

ограничение на время выполнения всех задач проекта к установленному сроку

$$\max_{(i)} \beta_i \le T.$$

При выполнении оговоренных условий и обозначений задача распределения ресурсов между операциями проекта обычно формулируется следующим образом: имеется комплекс операций, для которых определены отношения частичного порядка, задаваемые в виде графа G. Необходимо распределить ресурсы, заданные в количестве $R_j(t)$, между операциями комплекса таким образом, чтобы некоторая целевая функция достигла своего экстремального значения.

В качестве критерия оптимальности можно использовать

$$\max_{(i)} \beta_i \to \min, \tag{10.1}$$

что соответствует выполнению проекта за минимальное время.

Более общий критерий для данной задачи может быть записан в виде

$$\sum_{i=1}^{I} \varphi_i(\beta_i) \to \text{opt}, \tag{10.2}$$

где $\varphi_i(\beta_i)$ – неубывающая функция β_i, представляющая собой некоторый функционал, имеющий экономический смысл. Чаще всего данный функционал представляет собой расходы, связанные с проектированием. В этом случае функционал необходимо минимизировать.

Задача (10.1) характерна для проектирования технических систем (технологических линий, летательных аппаратов и т.п.). Для таких систем после завершения этапа разработки начинается этап функционирования. При разработке автоматизированных систем существенными становятся моменты окончания разработок отдельных функциональных задач. Сказывается этапность разработки и внедрения систем. Для этих задач целесообразно пользоваться критерием (10.2).

Рассмотрим еще одну постановку задачи, возникающей при организации деятельности транспортных предприятий, так называемую **транспортную задачу**. В некоторых пунктах $a_1, a_2, \ldots, a_n$ находятся склады, в которых хранятся товары в количествах $X_1, X_2, \ldots, X_n$ соответственно. В пунктах $b_1, b_2, \ldots, b_m$ находятся потребители, которым необходимо поставить эти товары в количествах, не меньших, чем $Y_1, Y_2, \ldots, Y_m$ соответственно. Обозначим через c_{ij} стоимость перевозки единицы груза из пункта a_i в пункт b_j, x_{ij} – количество товара, перевозимого из пункта a_i в пункт b_j. Для того, чтобы удовлетворить запросы потребителей, необходимо, чтобы выполнялась система неравенств

$$\sum_{i} x_{ij} \ge Y_j,\ j=\overline{1,\ m}.$$

С другой стороны, необходимо учитывать, что с i-го склада нельзя вывезти больше продукта, чем там имеется. Следовательно, должна выполняться еще одна система

$$\sum_j x_{ij} \geq X_i, \; i=\overline{1, n}.$$

Удовлетворить сформулированным условиям можно бесконечным числом способов. Для того, чтобы выбрать оптимальное правило перевозок, необходимо сформулировать критерий, который будет отражать представления о цели функционирования транспортного предприятия. В данной задаче одним из возможных критериев может выступать стоимость перевозок, тогда вид функционала будет определяться очевидным образом:

$$\sum_i \sum_j c_{ij} x_{ij} \to \min.$$

Таким образом, рассмотрены задачи, математическая формулировка которых описывается схожими моделями, а именно, и оптимизируемый функционал, и ограничения представляют собой линейные формы некоторых переменных. Рассмотрим подходы к решению такого типа задач.

10.2. Задача линейного программирования

Постановка задачи линейного программирования. Задачи линейного программирования (ЗЛП) – простейший тип оптимизационных задач. Постановка данной задачи выглядит следующим образом.

Имеется множество переменных $X = (x_1, x_2,\ldots, x_n)$. Целевая функция линейно зависит от управляемых параметров:

$$F = \sum_{i=1}^{n} c_i x_i.$$

Имеются ограничения, которые представляют собой линейные формы:

$$\sum_i a_{ji} x_i \leq b_j, \text{ где } j=\overline{1, m}.$$

Задача линейного программирования формулируется так: определить максимум линейной формы

$$F(x_1, x_2,\ldots, x_n) = \max(c_1 x_1 + c_2 x_2 + \ldots + c_n x_n) \tag{10.3}$$

при условии, что точка $(x_1, x_2,\ldots, x_n)$ принадлежит некоторому множеству D, которое определяется системой линейных неравенств

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1; \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leq b_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leq b_m. \end{cases} \quad x_i > 0, \; i=\overline{1, n}. \tag{10.4}$$

Любое множество значений $(x_1^*, x_2^*,\ldots, x_n^*)$, которое удовлетворяет системе неравенств (10.4) задачи линейного программирования, является допустимым решением данной задачи. Если при этом выполняется неравенство

$$c_1 x_1^0 + c_2 x_2^0 + \ldots + c_n x_n^0 \geq c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$$

для всего множества значений $x_1, x_2,\ldots, x_n$, то значение $x_1^0, x_2^0,\ldots, x_n^0$ является оптимальным решением задачи линейного программирования.

Задачу линейного программирования удобно представлять в векторной форме, тогда она будет выглядеть следующим образом:

найти $\max F(x) = \max (\mathbf{c}^{\text{т}} x)$

при условии $\mathbf{A}X \leq \mathbf{P}_0$; $X \geq 0$,

где $\mathbf{c} = (c_1, c_2,\ldots, c_n)$ представляет собой n-мерный вектор, составленный из коэффициентов целевой функции, причем $\mathbf{c}^{\text{т}}$ – транспонированная вектор-строка; $\mathbf{x} = (x_1, x_2,\ldots, x_n)$ – n-мерный вектор переменных решений;

$$\mathbf{P}_0 = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$ – m-мерный вектор свободных членов ограничений;

матрица $\mathbf{A}$ размером $(m \times n)$ – матрица, составленная из коэффициентов всех линейных ограничений:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{v2} & \ldots & a_{mn} \end{pmatrix}.$$

Простые ЗЛП допускают геометрическую интерпретацию, позволяющую непосредственно из графика получить решение и проиллюстрировать идею решения более сложных задач линейного программирования.

Каноническая форма задачи линейного программирования. Любую задачу линейного программирования можно свести к некоторой стандартной форме с ограничениями, записанными в виде уравнений. Это достигается путем введения свободных переменных во все ограничения. Свободная переменная учитывает разницу между правой и левой частями неравенства.

Пусть

x_{n+1} – дополнительная переменная, которая численно равна $b_1 - \sum_{i=1}^{n} a_{1i}x_i$;

x_{n+2} – дополнительная переменная, которая численно равна $b_2 - \sum_{i=1}^{n} a_{2i}x_i$, и т.д.;

x_{n+m} – дополнительная переменная, которая численно равна $b_m - \sum_{i=1}^{n} a_{mi}x_i$.

В результате получаем новую систему ограничений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + 1x_{n+1} + 0x_{n+2} + \ldots + 0x_{n+m} = b_1; \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + 0x_{n+1} + 1x_{n+2} + \ldots + 0x_{n+m} = b_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + 0x_{n+1} + 0x_{n+2} + \ldots + 1x_{n+m} = b_m; \\ x_i \geq 0,\ i = 1,\ 2, \ldots,\ n,\ n+1, \ldots,\ n+m. \end{cases}$$

Целевая функция будет иметь вид

$$\max F(x_1, x_2, \ldots, x_n) = \max(c_1x_1 + c_2x_2 + \ldots, c_nx_n + 0x_{n+1} + 0x_{n+2} + \ldots + 0x_{n+m}).$$

Или в матричной форме:

$$\max F(x) = \max(\mathbf{c}^{\mathrm{T}}x) \tag{10.5}$$

при условии $\mathbf{AX}_1 + \mathbf{BX}_2 = \mathbf{P}_0$, $\mathbf{X}_1 \geq 0$, $\mathbf{X}_1 \geq 0$,

где $\mathbf{X}_1$ – вектор первоначальных переменных; $\mathbf{X}_2$ – вектор свободных переменных; $\mathbf{B}$ – единичная матрица $m \times n$.

$$\mathbf{B} = \begin{pmatrix} 1\ 0\ \ldots\ 0 \\ 0\ 1\ \ldots\ 0 \\ \ldots\ldots\ldots \\ 0\ 0\ \ldots\ 1 \end{pmatrix}.$$

Такую форму записи называют канонической формой задачи линейного программирования. В канонической форме ограничения записываются в виде **равенств**.

При записи задачи линейного программирования в стандартной или канонической форме число линейно независимых уравнений, как правило, меньше числа переменных (на практике всегда $m < n$, где m – число уравнений). Отметим некоторые свойства, касающиеся системы ограничений:

1) уравнения линейно независимы, если ни одно из них не может быть получено из остальных путем алгебраических преобразований, т.е. никакие из них не являются следствием остальных;

2) если число независимых уравнений больше числа переменных, то такая система не имеет решения и называется несовместимой;

3) если число независимых уравнений равно числу переменных, то такая система имеет единственное решение, которое либо оптимально, если все компоненты положительны, либо недопустимо, если хотя бы одна из компонент отрицательна;

4) если удается найти множество неотрицательных значений $X = (x_1, x_2, \ldots, x_n)$, которое является решением системы m линейных уравнений с $n+m$ неизвестными, то такое решение называют базисным, а ненулевые переменные – базисными переменными.

Пример задачи линейного программирования. Рассмотрим двухмерную задачу линейного программирования.

Пусть требуется найти максимум линейной формы

$$\max F(x_1, x_2) = \max(x_1 + 2x_2)$$

при условии

$$\begin{aligned} &x_1 + x_2 \leq 120, \\ &0 \leq x_1 \leq 100, \\ &0 \leq x_2 \leq 75. \end{aligned}$$

Изобразим область, описываемую совокупностью ограничений на плоскости $x_1 \circ x_2$ (рис. 10.1). Переменные x_1 и x_2 неотрицательные, поэтому множество точек (x_1, x_2), являющихся возможными решениями задачи, находятся в I квадранте. Заменим знак в $x_1 + x_2 \leq 120$ на знак равенства, получим уравнение прямой: $x_1 + x_2 = 120$. Эта прямая делит плоскость на две полуплоскости. Все точки одной полуплоскости удовлетворяют неравенству $x_1 + x_2 > 120$, другой – неравенству $x_1 + x_2 < 120$.

Построив аналогичные прямые $x_1 = 100$ и $x_2 = 75$, получим многоугольник, множество точек которого (x_1, x_2) удовлетворяет всем неравенствам системы ограничений. Этот многоугольник и представляет собой область допустимых решений D.

Из множества точек (x_1, x_2) многоугольника необходимо выбрать такую, в которой функция $F(x_1, x_2) = (x_1 + 2x_2)$ принимает максимальное

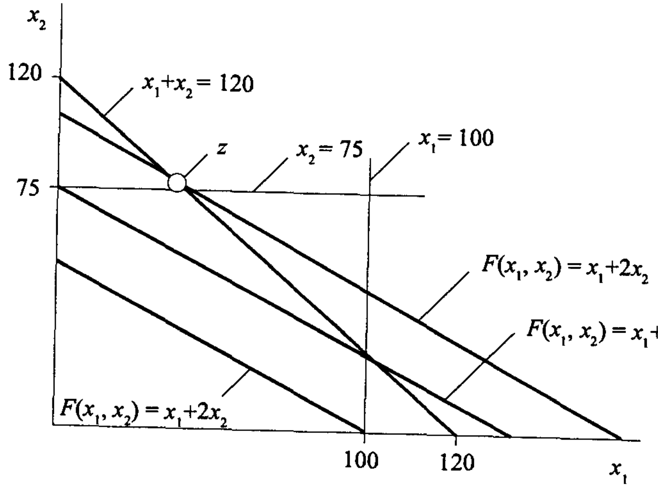

Рис. 10.1. Геометрическая интерпретация задачи лиейного программирования

значение. Для некоторого фиксированного значения F^* линейная функция $F^*(x_1, x_2) = (x_1 + 2x_2)$ представляет собой прямую линию. Задаваясь различными значениями F^*, получим семейство параллельных прямых. Увеличение значений линейной функции соответствует перемещению прямой параллельно самой себе вверх. Следовательно, как видно из рис. 10.1, максимальное значение целевой функции на допустимом множестве точек соответствует прямой, проходящей через точку z пересечения прямых $x_1 + x_2 = 120$ и $x_2 = 75$.

Решив эту систему, получим $x_1 = 45$. Тогда максимальное значение функции $F = \max(x_1 + 2x_2) = 195$.

10.3. Решение задач линейного программирования симплекс-методом

Метод полного исключения. Перейдем к изложению методов решения задач линейного программирования. Рассмотрим вначале один из вспомогательных методов, иллюстрирующий возможность преобразования системы линейных уравнений, а именно, приведения матрицы, составленной из параметров уравнений ограничений, к единичному виду. Переобозначим свободные коэффициенты ограничений $a_{j0}=b_j$.

Пусть имеется система ограничений в виде

$$\sum_{i=1}^{n} a_{ij}x_i = a_{j0},\ j=\overline{1,m}.$$

Перепишем данную систему уравнений в матричной форме: $\mathbf{AX}=\mathbf{A}_0$, и $\mathbf{A}_p = [\mathbf{A}, \mathbf{A}_0]$, где $\mathbf{A}_p$ – расширенная матрица.

Метод полного исключения состоит из конечного числа однотипных шагов и заключается в приведении матрицы **А** к единичному виду. Метод основан на следующих двух операциях.

Одну из строк расширенной матрицы умножают на число, отличное от нуля.

Из каждой строки расширенной матрицы ($\mathbf{A}_p$) вычитают данную строку. Каждое такое преобразование называется преобразованием Гаусса. В результате получаем новую систему линейных уравнений, эквивалентную исходной.

Рассмотрим более подробно метод полного исключения.

1. Среди элементов матрицы **А** выбирают произвольный элемент, отличный от нуля. Этот элемент называют направляющим элементом шага. Строку и столбец, содержащие направляющий элемент, называют направляющей строкой и направляющим столбцом данного преобразования.

2. Все элементы направляющей строки расширенной матрицы ($\mathbf{A}_p$) делят на направляющий элемент. В результате получают направляющую строку с направляющим элементом, равным единице. Далее из каждой строки матрицы **А** вычитают новую направляющую строку, умноженную на элемент, который расположен на пересечении строки и направляющего столбца. Итак, пусть исходная система уравнений имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = a_{10}; \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = a_{20}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = a_{m0}. \end{cases}$$

Возьмем в качестве направляющей строки вторую строку, в качестве направляющего столбца второй столбец, тогда направляющий элемент будет x_2 (коэффициент при нем равен a_{22}). Разделим направляющую строку на этот коэффициент и перепишем вновь полученную систему ограничений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = a_{10}; \\ \dfrac{a_{21}}{a_{22}}x_1 + x_2 + \ldots + \dfrac{a_{2n}}{a_{22}}x_n = \dfrac{a_{20}}{a_{22}}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = a_{m0}. \end{cases}$$

Далее умножаем направляющую строку поочередно на элементы, стоящие в направляющем столбце преобразуемого уравнения, и результат вычитаем из соответствующего уравнения. Иными словами, вначале умножим направляющую строку на a_{12} и вычтем полученный результат из первого уравнения. Далее направляющую строку умножим на a_{32} и вычтем из третьего уравнения и т.д. до последнего m-го уравнения. Получим новую систему:

$$\begin{cases} \left(a_{11} - \dfrac{a_{21}}{a_{22}}a_{12}\right)x_1 + 0x_2 + \left(a_{13} - \dfrac{a_{21}}{a_{22}}a_{12}\right)x_3 + \ldots + \left(a_{1n} - \dfrac{a_{21}}{a_{22}}a_{12}\right)x_n = \left(a_{10} - \dfrac{a_{21}}{a_{22}}a_{12}\right); \\ \dfrac{a_{21}}{a_{22}}x_1 + 1x_2 + \dfrac{a_{23}}{a_{22}}x_3 + \ldots + \dfrac{a_{2n}}{a_{22}}x_n = \dfrac{a_{20}}{a_{22}}; \\ \left(a_{31} - \dfrac{a_{21}}{a_{22}}a_{32}\right)x_1 + 0x_2 + \left(a_{33} - \dfrac{a_{21}}{a_{22}}a_{32}\right)x_3 + \ldots + \left(a_{3n} - \dfrac{a_{21}}{a_{22}}a_{32}\right)x_n = \left(a_{30} - \dfrac{a_{21}}{a_{22}}a_{32}\right); \\ \ldots \\ \left(a_{m1} - \dfrac{a_{21}}{a_{22}}a_{m2}\right)x_1 + 0x_2 + \left(a_{m3} - \dfrac{a_{21}}{a_{22}}a_{m2}\right)x_3 + \ldots + \left(a_{mn} - \dfrac{a_{21}}{a_{22}}a_{m2}\right)x_n = \left(a_{m0} - \dfrac{a_{21}}{a_{22}}a_{m2}\right). \end{cases}$$

Матрицу, в которую преобразовалась расширенная матрица $\mathbf{A}_p$ после первого шага, обозначим $\mathbf{A}_p^{(1)}$. В полученной матрице все элементы направляющего столбца, отличные от направляющего элемента, стали равными нулю. Совокупность элементов первых n столбцов матрицы $\mathbf{A}_p$, лежащих вне направляющей строки и столбца предыдущего шага, называют главной частью матрицы $\mathbf{A}_p^{(1)}$ данного преобразования. Направляющий элемент второго шага выбирают среди ненулевых элементов главной части матрицы $\mathbf{A}_p^{(1)}$, полученной после проведенного преобразования.

Второй и дальнейший шаги метода проводят аналогично первому шагу. Последовательность действий продолжается до тех пор, пока

имеется возможность выбора направляющего элемента. Если после k-го шага главная часть матрицы $\mathbf{A}_p^{(k)}$ не содержит ни одного элемента или содержит только нулевые элементы, то процесс исключения переменных заканчивается. После проведения преобразований в системе может оказаться l уравнений, $l \le m$. Пусть это первые по порядку l уравнений тогда система уравнений может быть записана в виде

$$\sum_{j=1}^{n} a_{ij}^{(l)} x_j = a_{i0}^{(l)}. \tag{10.6}$$

Примем, что i-й направляющей строке соответствует i-й направляющий столбец, тогда $a_{ij}^{(l)} = \begin{cases} 0; & i \ne j \\ 1; & i = j \end{cases}$; $j = 1, 2, \ldots, l$.

Следовательно, (10.6) можно записать в виде $x_i = a_{i0}^{(l)} - \sum_{j=l+1}^{n} a_{ij}^{(l)} x_j$, причем переменные x_i являются базисными, а переменные x_s ($s = l+1, \ldots, n$) – небазисными.

Пример применения метода полного исключения. Рассмотрим пример применения метода полного исключения Гаусса для исследования системы уравнений

$$\begin{cases} x_1 + 2x_2 + x_3 + x_4 = 3; \\ 2x_1 + x_2 + x_3 + 3x_4 = 3; \\ 4x_1 + 5x_2 + 3x_3 + 5x_4 = 9. \end{cases}$$

Расширенная матрица имеет вид

$$\mathbf{A}_p = \begin{vmatrix} 1 & 2 & 1 & 1 & 3 \\ 2 & 1 & 1 & 3 & 3 \\ 4 & 5 & 3 & 5 & 9 \\ & & & & \\ A_1 & A_2 & A_3 & A_4 & A_0 \end{vmatrix}$$

Первый шаг. В качестве первого направляющего элемента возьмем a_{11}=1. Умножив первую строку матрицы **A** на 2, вычтем результат из второго уравнения и, умножив на 4, вычтем результат из третьего уравнения; получим

$$\mathbf{A}_{\mathrm{p}}^{(1)} = \begin{array}{|ccccc|} A_1 & A_2 & A_3 & A_4 & A_0 \\ 1 & 2 & 1 & 1 & 3 \\ 0 & -3 & -1 & 1 & -3 \\ 0 & -3 & -1 & 1 & -3 \end{array}.$$

Второй шаг. Поскольку главная часть матрицы $\mathbf{A}_{\mathrm{p}}^{(1)}$ содержит отличные от нуля элементы, продолжим процесс исключения. Выберем элемент $a_{22}^{(1)} = -3$ в качестве направляющего элемента на втором шаге преобразования. Разделим вторую строку на –3, получим

$$\mathbf{A}_{\mathrm{p}} = \begin{vmatrix} 1 & 2 & 1 & 1 & 3 \\ 0 & 1 & 1/3 & -1/3 & 1 \\ 0 & -3 & -1 & 1 & -3 \end{vmatrix}$$

Далее умножим вторую строку на 2 и вычтем результат из первой строки, умножим вторую строку на –3, результат вычтем из третьей строки, в итоге $\mathbf{A}_{\mathrm{p}}^{(2)} = \begin{vmatrix} 1 & 0 & 1/3 & 1\frac{2}{3} & 1 \\ 0 & 1 & 1/3 & -1/3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{vmatrix}.$

Как видим, главная часть матрицы $\mathbf{A}_{\mathrm{p}}^{(2)}$, состоящая из элементов $a_{33}^{(2)}$ и $a_{34}^{(2)}$, содержит только нули. Следовательно, процесс исключения заканчивается. Исследуем матрицу $\mathbf{A}^{(2)}$. Поскольку третья строка содержит только нулевые элементы, это свидетельствует о том, что начальные уравнения были зависимыми, и, следовательно, она может быть отброшена. Тогда эквивалентная матрица системы уравнений будет выглядеть следующим образом:

$$\mathbf{A}_{\mathrm{p}}^{(3)} = \begin{vmatrix} 1 & 0 & 1/3 & 1\frac{2}{3} & 1 \\ 0 & 1 & 1/3 & -1/3 & 1 \end{vmatrix}.$$

Теперь можно записать базисное решение

$$x_1 = 1;\ x_2 = 1;\ x_3 = 0;\ x_4 = 0,$$

а также соответствующее общее решение

$$x_1 = 1 - \frac{1}{3}\alpha_3 - \frac{5}{3}\alpha_4;\ x_2 = 1 - \frac{1}{3}\alpha_3 + \frac{1}{3}\alpha_4;\ x_3 = \alpha_3;\ x_4 = \alpha_4,$$

где α_3, α_4 – произвольные скаляры.

Симплексные преобразования. В основе симплекс-метода лежит идея поиска базисного решения с последующим переходом от одного базиса к другому таким образом, чтобы целевая функция при этом все время увеличивалась, если речь идет о задаче максимизации. Пусть мы имеем задачу линейного программирования в канонической форме. Матрица ограничений имеет вид

$$A_1 x_1 + \ldots + A_n x_n + e_1 x_{n+1} + \ldots + e_m x_{n+m} = A_0,$$

где $e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix};\ e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix};\ e_m = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$ – единичный базис, элементы $a_{i0} \geq 0$ для всех $i = 1, 2, \ldots, m$.

Будем применять метод полного исключения к расширенной матрице ограничений. Выбираем направляющий элемент a_{ij} на данной итерации. В результате преобразования Гаусса, методика которого была только что описана, получим новые значения коэффициентов:

$$a_{l0}^{(k+1)} = a_{l0}^{(k)} - \frac{a_{i0}^{(k)} a_{lj}^{(k)}}{a_{ij}^{(k)}};\ a_{i0}^{(k+1)} = \frac{a_{i0}^{(k)}}{a_{ij}^{(k)}},$$

если $l \neq i$, $l = 1, 2, \ldots, m$.

Выясним условия, при которых новое базисное решение будет допустимым, т.е. $a_{l0}^{(k+1)} > 0$ для всех l. По предположению, $a_{l0} \geq 0$, тогда $a_{i0}^{(k+1)} = \dfrac{a_{i0}^{(k)}}{a_{ij}^{(k)}} \geq 0.$

Если $a_{lj}^{(k)} < 0$, то очевидно, $a_{l0}^{(k+1)} > 0$, так как $a_{i0}^{(k)} > 0$, $a_{ij}^{(k)} > 0$.

Если $a_{lj}^{(k)} > 0$, то $a_{l0}^{(k+1)} = a_{lj}^{(k)} \left(\dfrac{a_{l0}^{(k)}}{a_{lj}^{(k)}} - \dfrac{a_{i0}^{(k)}}{a_{ij}^{(k)}} \right)$ будет больше нуля при всех значениях $l = 1, 2, \ldots, m$ тогда и только тогда, когда $\dfrac{a_{i0}^{(k)}}{a_{ij}^{(k)}} = \min \left(\dfrac{a_{l0}^{(k)}}{a_{lj}^{(k)}} \right)$ при условии $a_{lj}^{(k)} > 0$.

Преобразование Гаусса называется симплексным преобразованием, когда направляющий элемент определяют по следующим правилам:

• направляющий столбец выбирают из условия, что в нем имеется хотя бы один положительный элемент;

• направляющую строку выбирают так, чтобы отношение $\frac{a_{i0}}{a_{ij}}$ было минимальным при условии, что $a_{ij} > 0$.

Задачи линейного программирования, решаемые с помощью симплекс-метода, основываются на представлении решения в табличной форме. Рассмотрим последовательность действий при решении задачи линейного программирования.

Пусть имеются линейная форма

$$c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max$$

и ограничения

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le a_{10}; \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le a_{20}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le a_{m0}. \end{cases}$$

Приведем матрицу ограничений к каноническому виду:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{1n}x_n + \ldots + 1x_{n+1} + \ldots + 0x_{n+m} = a_{10}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + 0x_{n+1} + \ldots + 1x_{n+m} = a_{m0}. \end{cases}$$

На следующем шаге составим таблицу (табл. 10.1).

Таблица 10.1

C			C_1	C_2	C_3	…	C_j	…	C_n	0	…	0
	B_x	A_0	A_1	A_2	A_3	…	A_j	…	A_n	A_{n+1}	…	A_{n+m}
c_{n+1}	x_{n+1}	a_{10}	a_{11}	a_{12}	a_{13}	…	a_{1j}	…	a_{1n}	1	…	0
c_{n+2}	x_{n+2}	a_{20}	a_{21}	a_{22}	a_{23}	…	a_{2j}	…	a_{2n}	0	…	0
…	…	…	…	…	…	…	…	…	…	…	…	…
c_{n+i}	x_{n+i}	a_{i0}	a_{i1}	a_{i2}	a_{i3}	…	a_{ij}	…	a_{in}	0	…	0
…	…	…	…	…	…	…	…	…	…	…	…	…
c_{n+m}	x_{n+m}	a_{m0}	a_{m1}	a_{m2}	a_{m3}	…	a_{mj}	…	a_{mn}	0	…	1
		a_{00}	a_{01}	a_{02}	a_{03}	…	a_{0j}	…	a_{0n}	$a_{0,m+1}$	…	a_{0m+n}

Нижняя строка элементов a_{0k}, $k = \overline{0, m+n}$ называется индексной. Элементы индексной строки вычисляются следующим образом:

$a_{0k} = \sum_{i=1}^{m} a_{ik}c_{n+i} - c_k$; $i = \overline{1, m}$ означает номер соответствующей строки. Поскольку на первом шаге заполнения таблицы все элементы c_{n+i} равны нулю, то элементам индексной строки присваиваются значения соответствующих элементов целевой функции данного столбца, взятые с обратным знаком, т.е. $a_{0k} = -C_k$. Последняя строка таблицы служит для определения направляющего столбца.

Элемент a_{00} равен значению целевой функции, которое вычисляется по формуле $a_{00} = \sum_{k=1}^{m} c_{n+k}a_{k0}$, k – номер базисной переменной (индексация идет по *строкам таблицы*).

В столбце B_x записываются базисные переменные, на первом шаге в качестве базисных выбирают фиктивные переменные $\{x_{n+k}\}, k = \overline{1, m}$. В дальнейшем фиктивные переменные необходимо вывести из базиса.

В столбец C записываются коэффициенты при x_{n+k}, на первом шаге значения этих коэффициентов равны нулю.

Для перехода от базиса фиктивных переменных к базису реальных переменных применяют следующие правила:

• в качестве направляющего столбца выбирают столбец A_j, для которого выполняется условие $a_{0j} = \min\{a_{0\ell}\}$, $l = \overline{1, n+m}$ при $a_{0\ell} < 0$, т.е. выбирается минимальный элемент, при условии, что этот элемент отрицательный;

• выбирают направляющую строку, для чего каждый элемент столбца свободных членов делится на соответствующий элемент направляющего столбца, элемент столбца свободных членов находится в одной строке с элементом направляющего столбца $\frac{a_{i0}}{a_{ij}}$. Из всех возможных соотношений выбирается минимальное $\frac{a_{i0}}{a_{ij}} = \min\left\{\frac{a_{r0}}{a_{rj}}\right\}$ при условии, что $a_{rj} > 0$, $1 \le r \le m$.

Применяя сформулированные правила, определяем направляющий элемент. Далее выполняется шаг симплексных преобразований.

Переменная, которая соответствует направляющему столбцу, вводится в базис, а переменная, соответствующая направляющей строке, выводится из базиса. При этом для переменной, вводимой в базис, изменяется соответствующее значение коэффициента целевой функции. Вместо коэффициента c_{n+i}, соответствующего старой базисной переменной, в таблице записывается значение коэффициента целевой функции при переменной, вводимой в базис.

Если направляющий элемент a_{ij}, то переход от данной таблицы к следующей осуществляется с использованием следующих правил.

1. Для всех элементов направляющей строки

$$a_{i\ell}^{(k+1)} = \frac{a_{i\ell}^{(k)}}{a_{ij}^{(k)}},$$

где k – номер шага ($k = 1, 2,\ldots$); i – номер направляющей строки; j – номер направляющего столбца; $\ell = \overline{0,\, m+n}$, т.е. все элементы направляющей строки делим на направляющий элемент, в итоге направляющий элемент стал равным единице; $a_{i,j}^{(k+1)} = 1$.

2. В направляющем столбце необходимо получить $a_{rj}^{(k+1)} = 0$, для всех $r = \overline{1,\, m}$, $r \neq i$, при $a_{ij}^{(k+1)} = 1$, т.е. в направляющем столбце должны быть все нули кроме направляющего элемента, который равен единице.

Для всех остальных элементов, включая индексную строку, производим вычисления

$$a_{r\ell}^{(k+1)} = a_{r\ell}^{(k)} - \frac{a_{i\ell}^{(k)} a_{rj}^{(k)}}{a_{ij}^{(k)}},\ l \neq j, r \neq i.$$

Симплексные преобразования повторяют до тех пор, пока не реализуется один из двух возможных исходов:

а) все $a_{0\ell} \geq 0$ – это условие оптимальности базиса последней таблицы;

б) найдется такой элемент $a_{0j} < 0$, при котором все элементы столбца $a_{rj} \leq 0$, – это признак неограниченности целевой функции на множестве допустимых решений.

Пример решения задачи линейного программирования. Рассмотрим задачу линейного программирования в следующем виде: найти максимум линейной формы $4x_1 + 3x_2$ при ограничениях

$$x_1 \leq 4000,\ x_2 \leq 6000,\ x_1 + \frac{2}{3}x_2 \leq 6000,\ x_1, x_2 \geq 0.$$

Каноническая форма задачи линейного программирования будет иметь вид

$$4x_1 + 3x_2 + 0x_3 + 0x_4 + 0x_5 \to \max;$$
$$1x_1 + 0x_2 + 1x_3 + 0x_4 + 0x_5 = 4000;$$
$$0x_1 + 1x_2 + 0x_3 + 1x_4 + 0x_5 = 6000;$$
$$1x_1 + \frac{2}{3}x_2 + 0x_3 + 0x_4 + 1x_5 = 6000.$$

Составим исходную симплекс-таблицу (табл. 10.2).

Таблица 10.2

C_i			4	3	0	0	0
	B_x	A_0	A_1	A_2	A_3	A_4	A_5
0	x_3	4000	1	0	1	0	0
0	x_4	6000	0	1	0	1	0
0	x_5	6000	1	$\frac{2}{3}$	0	0	1
		0	–4	–3	0	0	0

Поскольку $-4 < -3 < 0$, то в качестве направляющего выбираем первый столбец. Составив отношение вида $\left\{\frac{a_{i0}}{a_{i1}}\right\}$, определяем направляющую строку. Для этого находим минимальное отношение $\min\left\{\frac{4000}{1}, \frac{6000}{0}, \frac{6000}{1}\right\} = 4000$. Следовательно, направляющая строка – первая, направляющий элемент – $a_{11}=1$. Применив первый шаг симплексного преобразования, получим новую таблицу (табл. 10.3).

Таблица 10.3

C_i			4	3	0	0	0
	B_x	A_0	A_1	A_2	A_3	A_4	A_5
4	x_1	4000	1	0	1	0	0
0	x_4	6000	0	1	0	1	0
0	x_5	2000	0	$\frac{2}{3}$	–1	0	1
		16000	0	–3	4	0	0

На данном этапе в качестве направляющего столбца выбираем второй, направляющая строка – третья, т.к. $\frac{2000}{2/3} < \frac{6000}{1} < \frac{4000}{0}$. Применим следующий шаг симплексного преобразования. В результате получим табл. 10.4

Таблица 10.4

C_i			4	3	0	0	0
	B_x	A_0	A_1	A_2	A_3	A_4	A_5
4	x_1	4000	1	0	1	0	0
0	x_4	3000	0	0	$\frac{3}{2}$	1	$-\frac{3}{2}$
3	x_2	3000	0	1	$-\frac{3}{2}$	0	$\frac{3}{2}$
		25000	0	0	$-\frac{1}{2}$	0	$\frac{9}{2}$

Так как $a_{03} = -\frac{1}{2} < 0$, то направляющий столбец A_3, направляющая строка – вторая, направляющий элемент $a_{23} = \frac{3}{2}$. Выполним очередной шаг преобразования, получим еще одну таблицу (табл. 10.5).

Таблица 10.5

C_i			4	3	0	0	0
	B_x	A_0	A_1	A_2	A_3	A_4	A_5
4	x_1	2000	1	0	0	$-\frac{2}{3}$	1
0	x_3	2000	0	0	1	$\frac{2}{3}$	–1
3	x_2	6000	0	1	0	1	0
		26000	0	0	0	$\frac{1}{3}$	4

Поскольку в индексной строке все элементы положительны, это означает, что найдено оптимальное решение $x_{10} = 2000$, $x_{20} = 6000$, $x_{30} = 2000$. Искомое значение целевой функции равно $4x_1 + 3x_2 = 26000$.

10.4. Двойственная задача линейного программирования

Структура и свойства двойственной задачи. Почти всякую ЗЛП можно рассматривать как экономическую задачу о распределении ресурсов $b_1, b_2, \ldots, b_m$ между различными потребителями, например, между некоторыми технологическими процессами $\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_n$. Любое допустимое решение ЗЛП $x_1, x_2, \ldots, x_n$ дает конкретное распределение, указывающее ту долю каждого из ресурсов, которая должна быть использована при осуществлении соответствующего технологического процесса.

Рассмотрим пример. Завод производит три вида продукции x_1, x_2, x_3, каждый из которых требует затрат времени на обработку на токарном, фрезерном и сверлильном станках. Количество машинного времени для каждого из станков ограниченно. Пусть c_1, c_2, c_3 – прибыль от реализации единицы соответствующего вида продукции. Требуется определить, какое количество каждого вида продукции необходимо производить в течение заданного интервала времени, чтобы получить максимальную прибыль.

Формально задача записывается так: найти

$$\max_{x_1, x_2, x_3} (c_1 x_1 + c_2 x_2 + c_3 x_3)$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \le b_1; \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \le b_2; \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \le b_3, \end{cases}$$

где a_{1j}, a_{2j}, a_{3j} – время, необходимое для обработки единицы j-го вида продукции соответственно на токарном, фрезерном и сверлильном станках (j = 1, 2, 3); b_1, b_2, b_3 – ресурс машинного времени соответственно для токарного, фрезерного и сверлильного станков. Сформулированная задача является прямой ЗЛП.

Рассмотрим далее задачу в несколько иной постановке. Обозначим через U_1, U_2, U_3 цену единицы времени работы на токарном, фрезерном и сверлильном станках соответственно, тогда $a_{11}U_1 + a_{21}U_2 + a_{31}U_3$ можно трактовать как расходы на изготовление единицы продукции первого вида, $a_{12}U_1 + a_{22}U_2 + a_{32}U_3$ – расходы на изготовление единицы продукции второго вида, $a_{13}U_1 + a_{23}U_2 + a_{33}U_3$ – расходы на изготовление единицы продукции третьего вида.

Для величин расходов выполняются следующие соотношения:

$$\begin{cases} a_{11}U_1 + a_{21}U_2 + a_{31}U_3 \geq c_1; \\ a_{12}x_1 + a_{22}U_2 + a_{32}U_3 \geq c_2; \\ a_{13}U_1 + a_{23}U_2 + a_{33}x_3 \geq c_3. \end{cases} \quad (10.7)$$

Учитывая, что b_1, b_2, b_3 – использованный ресурс машинного времени для каждого из станков, тогда $b_1U_1+b_2U_2+b_3U_3$ – суммарные расходы на производство.

Требуется найти такие U_1, U_2, U_3, удовлетворяющие условиям (10.7), при которых минимизируются суммарные расходы на производство. Таким образом, целевую функцию можно записать в виде

$$\min_{U_1, U_2, U_3} g(U) = b_1U_1 + b_2U_2 + b_3U_3, \quad (10.8)$$

причем U_1, U_2, $U_3 \geq 0$.

Задачу (10.8) с ограничениями (10.7) называют двойственной задачей по отношению к прямой, сформулированной ранее.

Соотношение прямой и двойственной задач. Между прямой и двойственной задачами имеет место взаимно однозначное отношение. Имея прямую задачу, всегда можно перейти к двойственной и наоборот. Данное отношение находит выражение в виде следующих правил:

1) если прямая задача является задачей максимизации, то двойственная будет задачей минимизации и наоборот;

2) коэффициенты целевой функции прямой задачи $c_1, c_2, \ldots, c_n$ становятся свободными членами ограничений двойственной задачи;

3) свободные члены ограничений прямой задачи $b_1, b_2, \ldots, b_m$ становятся коэффициентами целевой функции двойственной задачи;

4) матрицу ограничений двойственной задачи получают транспонированием матрицы ограничений прямой задачи;

5) знаки неравенств в ограничениях изменяются на обратные;

6) число ограничений прямой задачи равно числу переменных двойственной задачи, а число ограничений двойственной задачи равно числу переменных прямой задачи.

Переменные $U_1, U_2, \ldots, U_m$ двойственной задачи называют «теневыми ценами». Двойственную задачу выгоднее решать, чем исходную прямую, если в прямой задаче при малом количестве переменных $(m > n)$ имеется большое количество ограничений.

Запишем обе задачи в матричном виде.

Прямая задача. Найти $\max f(\bar{X}) = \max \bar{C}^{\text{т}} \bar{X}$ при $\bar{A}\bar{X} \leq A_0$; $\bar{X} \geq 0$.

Двойственная задача. Найти $\min g(\bar{U}) = \bar{A}_0^{\text{т}} \bar{U}$ при $\bar{A}^{\text{т}}\bar{U} \geq \bar{C}$; $\bar{U} \geq 0$.

Связь между оптимальными решениями прямой и двойственной задач устанавливается посредством следующих теорем.

Теорема 1. Если $\bar{X}_0$ и $\bar{U}_0$ – допустимые решения прямой и двойственной задач, т.е. если $\bar{A}\bar{X} \leq A_0$ и $\bar{A}^{\text{т}}\bar{U} \geq \bar{C}$, то $\bar{C}^{\text{т}}\bar{X}_0 \leq \bar{A}_0^{\text{т}}\bar{U}_0$, т.е. значения целевой функции прямой задачи никогда не превышают значений целевой функции двойственной задачи.

Теорема 2. Если $\bar{X}_0$ и $\bar{U}_0$ – допустимые решения прямой и двойственной задач и если $\bar{C}^{\text{т}}\bar{X}_0 = \bar{U}^{\text{т}}\bar{A}_0$, то $\bar{X}_0$ и $\bar{U}_0$ – оптимальные решения этих задач.

Таким образом, получено, что если прямая задача есть задача максимизации, то двойственная задача – задача минимизации. Остановимся на особенностях решения задач минимизации. Первое правило гласит, что в задачах минимизации направляющий столбец определяют по наибольшему положительному элементу индексной строки. Правило остановки для задачи минимизации формулируется так: все элементы индексной строки должны быть неположительными, т.е. $a_{0j} \leq 0$. И, наконец, главная сложность, возникающая при решении задач минимизации, – это выбор первоначального базиса. Дело в том, что при решении задач минимизации область допустимых значений ограничивается снизу, т.е. ограничения имеют вид $\bar{A}^{\text{т}}\bar{U} \geq \bar{C}$; $\bar{U} \geq 0$. При приведении системы ограничений к каноническому виду дополнительные переменные вводятся в модель с отрицательными знаками. По этой причине они не могут быть введены в базис. Отметим также, что в общем случае постановки задач линейного программирования допускают в ограничениях знаки неравенств как «больше или равно», так и «меньше или равно». Неравенства, в которых имеет место знак «меньше или равно», приводятся к каноническому виду путем добавления в первоначальный базис дополнительной переменной, как это было продемонстрировано при решении задачи максимизации симплекс-методом. Для неравенств, имеющих знак «больше или равно», дополнительная переменная вводится с отрицательным знаком, и в базис введена быть не может. По этой причине определение начального допустимого базиса в общем случае представляет значительные трудности. Для нахождения допустимых базисных решений разработаны специальные методы. Рассмотрим один из них.

10.5. Метод искусственных переменных

Рассмотрим задачу линейного программирования, в которой ограничения имеют вид $\bar{A}\bar{X} \le \bar{P}_0$. Если все $b_i \ge 0$, $i = 1, 2, \ldots, m$, то свободные векторы, образующие единичную подматрицу, составляют начальный базис, а соответствующие им переменные – начальное базисное решение.

В более общем случае, когда ряд неравенств имеет знак «больше или равно», например, $a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \ge b_i$, $i = 1, 2, \ldots, m$, для приведения их к стандартной канонической форме свободные переменные надо вычесть. Тогда расширенная форма задачи имеет вид

$$\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + a_{1n}x_n - 1x_{n+1} + 0x_{n+2} + \ldots + 0x_{n+m} = b_1;\\
&a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + 0x_{n+1} - 1x_{n+2} + \ldots + 0x_{n+m} = b_2; \qquad (10.9)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + 0x_{n+1} + \ldots - 1x_{n+m} = b_m.
\end{aligned}$$

Свободные переменные $\{x_{n+1}, x_{n+2}, \ldots, x_{n+m}\}$ уже нельзя использовать в качестве начального базиса, так как $x_{n+1} < 0$, $x_{n+2} < 0, \ldots, x_{n+m} < 0$.

С целью формирования начального базиса в уравнения (10.9) дополнительно вводят искусственные переменные $x_{n+m+1}, x_{n+m+2}, \ldots, x_{n+m+k}$, которые не имеют ничего общего с реальной задачей, и должны быть выведены из базиса как можно скорее. Чтобы гарантировать их быстрое выведение после начала итераций для задач максимизации, искусственным переменным в целевой функции приписывают очень большие по величине отрицательные коэффициенты ($-M$), где $M \gg c_i$, (i=1, 2, …, n). В случае решения задач минимизации искусственные переменные вводят в целевую функцию с большими по величине положительными коэффициентами ($+M$).

Знаки вводимых в ограничения искусственных переменных $x_{n+m+1}, x_{n+m+2}, \ldots, x_{n+m+k}$ должны совпадать со знаками соответствующих свободных членов. Искусственные переменные образуют начальное базисное решение. Применив симплекс-метод, решают задачу по выведению из базиса всех искусственных переменных. Если доказано, что от искусственных переменных избавиться нельзя, то это означает, что задача не имеет решения, т.е. ограничения задачи противоречивы.

Рассмотрим пример решения задачи линейного программирования с использованием метода искусственных переменных.

Требуется минимизировать линейную форму вида

$\min f(\bar{X}) = \min(15x_1 + 33x_2)$ при ограничениях

$$\begin{cases}
3x_1 + 2x_2 \ge 6,\\
6x_1 + x_2 \ge 6,\\
x_2 \ge 1,\\
x_1, x_2 \ge 0.
\end{cases}$$

Вводя свободные переменные x_3, x_4, x_5 приходим к расширенной форме задачи

$$\begin{cases}
3x_1 + 2x_2 - x_3 = 6,\\
6x_1 + x_2 - x_4 = 6,\\
x_2 - x_5 = 1.
\end{cases}$$

Переменные x_3, x_4, x_5 образуют недопустимое базисное решение

$$x_{36} = -6 < 0,\ x_{46} = -6 < 0,\ x_{56} = -1 < 0,$$

поэтому вводим в ограничения и в целевую функцию искусственные переменные x_6, x_7, x_8. Получим следующую запись задачи:

$$\begin{aligned}
&\min\{15x_1 + 33x_2 + Mx_6 + Mx_7 + Mx_8\},\\
&3x_1 + 2x_2 - x_3 + x_6 = 6,\\
&6x_1 + x_2 - x_4 + x_7 = 6,\\
&x_2 - x_5 + x_8 = 1.
\end{aligned}$$

Очевидно, начальное базисное решение $x_6^* = 6$, $x_7^* = 6$, $x_8^* = 1$.

Поскольку $\bar{A}_6$, $\bar{A}_7$, $\bar{A}_8$ образуют единичный базис, а все $a_{i0} > 0$, то для решения применим метод симплекс-таблиц.

Исходная таблица имеет следующий вид (табл. 10.6).

Таблица 10.6

			15	33	0	0	0	*M*	*M*	*M*
	B_x	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
M	x_6	6	3	2	–1	0	0	1	0	0
M	x_7	6	6	1	0	–1	0	0	1	0
M	x_8	1	0	1	0	0	–1	0	0	1
		13*M*	9*M*–15	4*M*–33	–*M*	–*M*	–*M*	0	0	0

Первая итерация. Элементы индексной строки вычисляем следующим образом: $a_{0l} = \sum_{i=1}^{m} a_{il}c_i - c_l$; $i = 1, 2, 3$ означает номер соответствующей строки

$$a_{01} = \sum_{i \in J_6} c_i x_{i1} - c_1 = \sum c_i x_i - c_1 = 3M + 6M + 0M - 15 = 9M - 15$$

$$a_{02} = \sum_{i \in J_6} c_i x_{i2} - c_2 = 2M + 1M + 1M - 33 = 4M - 33;$$

$$a_{03} = \sum_{i \in J_6} c_i x_{i3} - c_3 = -M;\ a_{04} = -M;\ a_{05} = -M;\ a_{06} = a_{07} = a_{08} = 0;$$

$$a_{00} = \sum_{i \in J_6}^{m} c_i x_i = 6M + 6M + 1M = 13M.$$

Поскольку решается задача минимизации, то направляющий столбец определяют по наибольшему положительному элементу индексной строки: $a_{0j} = \max\limits_{1 \le k \le n} \{a_{0k} / a_{0k} > 0\}$. Направляющий столбец – A_1; направляющая строка – вторая. Выполним один шаг преобразования. Разделим направляющую строку на направляющий элемент $a_{71} = 6$. Умножив преобразованную направляющую строку на 3, вычтем ее из первой. Третья строка остается без изменений, поскольку элемент направляющего столбца для нее равен нулю. Затем, умножив преобразованную направляющую строку на $(9M–15)$, вычтем ее из индексной. В результате получим табл. 10.7.

Таблица 10.7

c_i			15	33	0	0	0	M	M	M
	B_x	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
M	x_6	3	0	$1\frac{1}{2}$	–1	$\frac{1}{2}$	0	1	$-\frac{1}{2}$	0
15	x_1	1	1	$\frac{1}{6}$	0	$-\frac{1}{6}$	0	0	$\frac{1}{6}$	0
M	x_8	1	0	1	0	0	–1	0	0	1
		$4M+15$	0	$\frac{5}{2}M - 30\frac{1}{2}$	$-M$	$\frac{M}{2} - \frac{15}{6}$	$-M$	0	$\frac{5}{2} - \frac{3}{2}M$	0

Вторая итерация. Так как $a_{02} = \frac{5}{2}M - 30\frac{1}{2} > a_{0j} (j \ne 2)$, т.е. является максимальным элементом из всех элементов индексной строки, то направляющий столбец – A_2; направляющая строка – третья. Направляющий элемент $a_{82} = 1$.

Выполним шаг симплекс-преобразования. Умножив направляющую строку на $1\frac{1}{2}$, вычтем ее из первой строки. Затем, умножив направляющую строку на $\frac{1}{6}$, вычтем ее из второй и, наконец, умножив направляющую строку на $\left(\frac{5}{2}M - 30\frac{1}{2}\right)$, вычтем ее из последней строки. Получим табл. 10.8

Таблица 10.8

c_i			15	33	0	0	0	M	M	M
	B_x	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
M	x_6	$1\frac{1}{2}$	0	0	–1	$\frac{1}{2}$	$1\frac{1}{2}$	1	$-\frac{1}{2}$	$-1\frac{1}{2}$
15	x_1	$\frac{5}{6}$	1	0	0	$-\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	$-\frac{1}{6}$
33	x_2	1	0	1	0	0	–1	0	0	1
		$\frac{3}{2}M + 45\frac{1}{2}$	0	0	$-M$	$\frac{M}{2} - \frac{15}{6}$	$\frac{3}{2}M - 30\frac{1}{2}$	0	$\frac{5}{2} - \frac{3}{2}M$	$-\frac{5}{2}M + 30\frac{1}{2}$

Третья итерация. Поскольку $a_{05} = \frac{3}{2}M - 30\frac{1}{2} > a_{0j} (j \ne 5)$, то направляющий столбец – A_5. Направляющая строка – первая, направляющий элемент $a_{65} = 1\frac{1}{2}$. Выполнив очередной шаг симплекс-преобразования, выведем из базиса последнюю искусственную переменную x_6 и введем x_5.

Таким образом, приходим к таблице следующего вида (табл. 10.9).

Таблица 10.9

c_i			15	33	0	0	0	M	M	M
	B_x	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
0	x_5	1	0	0	$-\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{2}{3}$	$-\frac{1}{3}$	–1
15	x_1	$\frac{4}{6}$	1	0	$\frac{1}{9}$	$-\frac{2}{9}$	0	$-\frac{1}{9}$	$\frac{2}{9}$	0
33	x_2	2	0	1	$-\frac{2}{3}$	$\frac{1}{3}$	0	$\frac{2}{3}$	$-\frac{1}{3}$	0
		76	0	0	$-20\frac{1}{3}$	$\frac{46}{6}$	0	$\frac{61}{3} - M$	$-\frac{46}{6} - \frac{M}{2}$	$-M$

Четвертая итерация. Обратим внимание, что в этой таблице все искусственные переменные выведены из базиса. Направляющий столбец – A_4, направляющая строка – первая, направляющий элемент $a_{14} = \frac{1}{3}$. Выполнив еще один шаг симплекс-преобразования, получим табл. 10.10.

Таблица 10.10

c_i			15	33	0	0	0	M	M	M
	B_x	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
0	x_4	3	0	0	–2	1	3	2	–1	–3
15	x_1	$\frac{4}{3}$	1	0	$-\frac{1}{3}$	0	$\frac{2}{3}$	$\frac{1}{3}$	0	$-\frac{2}{3}$
33	x_2	1	0	1	0	0	–1	0	0	1
		53	0	0	–5	0	–23	5–M	$-\frac{M}{2}$	23–M

Поскольку в индексной строке все оценки $a_{0j} \le 0$, то найдено оптимальное решение $x_{1\,\text{опт}} = \frac{4}{3}$, $x_{2\,\text{опт}} = 1$, $x_{4\,\text{опт}} = 3$.

Искомое значение целевой функции $a_{00} = \min z = 53$.

Проверим это: $\min(15x_1 + 33x_2) = 15x_{1\,\text{опт}} + 33\,x_{2\,\text{опт}} = 53$.

Таким образом, рассмотренный метод позволяет решать задачи линейного программирования в общем случае, когда имеются ограничения в виде неравенств с различными знаками: как «больше или равно», так и «меньше или равно».

10.6. Дискретное программирование

Постановка задачи дискретного программирования. Многие задачи системного анализа, такие как распределение ресурсов, задачи сетевого планирования и управления, календарного планирования, описываются математическими моделями дискретного программирования.

Рассмотрим общую задачу максимизации.

Найти $\max f(x_1, x_2, \ldots, x_n)$ (10.10)

при условиях

$$\begin{gathered} g_1(x_1, x_2, \ldots, x_n) \le 0; \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ g_m(x_1, x_2, \ldots, x_n) \le 0; \end{gathered} \qquad (10.11)$$

$$X = (x_1, x_2, \ldots, x_n) \in D, \qquad (10.12)$$

где D – некоторое множество $R^{(n)}$.

Если множество D является конечным или счетным, то условие (10.12) – это условие дискретности, и данная задача является задачей дискретного программирования (ЗДП). Чаще всего условие дискретности разделено по отдельным переменным следующим образом:

$$x_j \in D_j, j = 1, 2, \ldots, n,$$

где D_j – конечное (или счетное) множество.

Если вводится ограничение x_j – целые числа ($j = 1, 2, \ldots, n$), то приходят к задачам целочисленного программирования (ЦП), которое является частным случаем дискретного программирования.

В задачах дискретного программирования область допустимых решений является невыпуклой и несвязной. Поэтому отыскание решения таких задач сопряжено со значительными трудностями. В частности, невозможно применение стандартных приемов, используемых при замене дискретной задачи ее непрерывным аналогом, состоящих в дальнейшем округлении найденного решения до ближайшего целочисленного. Например, рассмотрим следующую ЗЛП:

найти $\max(x_1 - 3x_2 + 3x_3)$

при условиях
$$\begin{cases} 2x_1 + x_2 - x_3 \le 4; \\ 4x_1 - 3x_2 \le 2; \\ -3x_1 + 2x_2 + x_3 \le 3, \end{cases}$$

где $x_1, x_2, x_3 \ge 0$, x_j – целые числа ($j = 1, 2, 3$). Игнорируя условие целочисленности, находим оптимальный план симплекс-методом:

$$x_{1\,\text{опт}} = \frac{1}{2},\ x_{2\,\text{опт}} = 0,\ x_{3\,\text{опт}} = 4\frac{1}{2}.$$

Проверка показывает, что никакое округление компонент этого плана не дает допустимого решения, удовлетворяющего ограничениям этой задачи. Искомое целочисленное решение задачи $x_{1\,\text{опт}} = 2$, $x_{2\,\text{опт}} = 2$, $x_{3\,\text{опт}} = 5$.

Таким образом, для решения задачи дискретного программирования (ЗДП) необходимы специальные методы. Методы решения ЗДП по принципу подхода к проблеме делят на три группы: 1) методы отсечения или отсекающих плоскостей; 2) метод ветвей и границ; 3) методы случайного поиска и эвристические методы.

Математические модели задач дискретного программирования. По структуре математической модели задачи дискретного программирования разделяют на следующие классы:

1) задачи с неделимостями;
2) экстремальные комбинаторные задачи;
3) задачи на несвязных и на невыпуклых плоскостях;
4) задачи с разрывными целевыми функциями.

Рассмотрим существо некоторых из них.

Задачи с неделимостями. Математические модели задач с неделимостями основаны на требовании целочисленности переменных $\{x_i\}$, вытекающем из физических условий практических задач.

К таким задачам относится задача об определении оптимальной структуры производственной программы, где $\{x_1, x_2,\ldots, x_n\}$ – объемы выпуска продукции.

Эта задача заключается в отыскании

$$\max_{\{x_i\}} \sum_{j=1}^{n} c_j x_j \qquad (10.13)$$

при

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \ (i=1, 2,\ldots, m) \qquad (10.14)$$

$$x_1 \ge 0, x_2 \ge 0,\ldots, x_n \ge 0; x_j \text{ – целые при } j \in J. \qquad (10.15)$$

Если $J = N = (1, 2, \ldots, n)$, то задача называется полностью целочисленной, в противном случае, если $J \ne N$ – частично целочисленной.

Задача о ранце. Одной из наиболее распространенных задач целочисленного программирования является так называемая задача о ранце.

Рассмотрим постановку данной задачи. Турист готовится к длительному переходу в горах. В рюкзаке он может нести груз, масса которого не более W. Этот груз может включать в себя n видов предметов, каждый предмет типа j, массой ω_j, $j = 1, 2, \ldots, n$. Для каждого вида предмета турист определяет его ценность E_j во время перехода. Задача заключается в определении количества предметов каждого типа, которые он должен положить в рюкзак, чтобы суммарная ценность снаряжения была максимальной.

Обозначим через x_j количество предметов j-го типа в рюкзаке.

Тогда математическая модель задачи такова:

$$\max \sum_{j=1}^{n} E_j x_j, \qquad (10.16)$$

при ограничениях

$$\sum_{j=1}^{n} \omega_j x_j \le W, \quad x_j \ge 0, \ x_j \text{ – целое}, i = 1,2,\ldots, m. \qquad (10.17)$$

Экстремальные комбинаторные задачи. В данных задачах необходимо найти экстремум некоторой целевой функции, заданной на конечном множестве, элементами которого служат перестановки из n символов (объектов).

Одной из наиболее простых задач этого класса является *задача о назначениях*: найти такую перестановку $(p_1, p_2,\ldots, p_n)$ из чисел 1, 2, 3, …, n, при которой обеспечен $\min \sum_{i=1}^{n} c_{ip_i}$ по всем перестановкам $(p_1, p_2,\ldots, p_n)$.

Каждая такая перестановка может быть представлена точкой в n^2-мерном евклидовом пространстве или в виде матрицы $\bar{X}^{n\times n} = \|x_{ij}\|$.

Вводим переменные :

$x_{ij} = 1$, если i-й механизм предназначен для j-й работы;

$x_{ij} = 0$ – в противном случае.

Очевидно, что должно выполняться условие

$$\sum_{i=1}^{n} x_{ij} = 1; \ \sum_{j=1}^{n} x_{ij} = 1; \ i = \overline{1, n}; j = \overline{1, n}. \qquad (10.18)$$

Данные ограничения означают, что один механизм может быть предназначен для выполнения только одной работы. Тогда задача будет состоять в определении таких чисел $\{x_{ij}\}$, при которых достигается минимум функционала $\min \sum_i \sum_j c_{ij} x_{ij}$ при ограничениях (10.18).

Задача о коммивояжере. Имеется $(n + 1)$ город. Задана матрица $C = \|c_{ij}\|$ расстояний между городами. Выезжая из исходного города A_0, коммивояжер должен побывать во всех остальных городах по одному разу и вернуться в город A_0. Требуется определить, в каком порядке следует объезжать города, чтобы суммарное пройденное расстояние было минимально.

Введем переменные:

$x_{ij} = 1$, если коммивояжер переезжает из населенного пункта A_i в A_j;

$x_{ij} = 0$ – в противном случае.

Математическая модель задачи имеет следующий вид:

найти

$$\min \sum_{i=0}^{n} \sum_{j=0}^{n} c_{ij} x_{ij} \qquad (10.19)$$

при условиях

$$\sum_{i=0}^{n} x_{ij} = 1, \ (j = 1, 2,\ldots, n); \qquad (10.20)$$

$$\sum_{j=0}^{n} x_{ij} = 1,\ (i = 1,\ 2,\ldots,\ n); \tag{10.21}$$

$$u_i - u_j + n x_{ij} \le n - 1,\ (i, j = 1,\ 2, \ldots n,\ i \ne j), \tag{10.22}$$

где u_i, u_j – произвольные целые и неотрицательные числа.

Условие (10.20) означает, что коммивояжер выезжает из каждого города один раз, а условие (10.21) – что он въезжает один раз в каждый город.

Если ограничить задачу только условиями (10.20) и (10.21), она будет эквивалентна задаче о назначениях, план которой не обязан быть цикличным. Иначе говоря, путь коммивояжера при этом можно представить как ряд несвязанных подциклов, в то время как его путь в действительности состоит из одного цикла.

Покажем, что для любого цикла, начинающегося в A_0, можно найти u_i, удовлетворяющие условию (10.22). Пусть $u_i = p$, если коммивояжер посещает город A_i на p-м этапе. Отсюда следует, что $u_i - u_j \le n - 1$ для всех i и j, и, таким образом, условие (10.22) выполняется при $x_{ij} = 0$.

При $x_{ij} = 1$ условие (10.22) выполняется как строгое равенство:

$$u_i - u_j + n x_{ij} = p - (p+1) + n = n - 1.$$

Метод ветвей и границ для задачи целочисленного программирования. Рассмотрим частично целочисленную задачу ЛП: минимизировать

$$z = f(x) = \sum_{i=1}^{n} c_i x_i\,, \tag{10.23}$$

при условиях

$$\sum_{i=1}^{n} a_{ij} x_i \ge b_j,\ (j = 1,\ 2,\ldots,\ m); \tag{10.24}$$

$$0 \le x_i \le d_i, (i = 1, 2, \ldots, n); \tag{10.25}$$

$$x_i \text{ – целые числа.} \tag{10.26}$$

Процесс поиска оптимального решения начинают с решения непрерывной задачи ЛП. Если полученный при этом оптимальный план $\bar{X}_0$ не удовлетворяет условию (10.26), то значение целевой функции $\xi_0 = f(X_0)$ дает нижнюю оценку для искомого решения, т.е. $\min z = \xi_0$.

Пусть некоторая переменная x_{i0} ($1 \le i_0 \le m$) не получила в плане $\bar{X}_0$ целочисленного решения. В целочисленном плане значение x_{i0} следует либо уменьшить, по крайней мере до $[x_{i0}]$, либо увеличить, по крайней мере до $[x_{i0}] + 1$.

Если границы изменения x_{i0} заранее не заданы, то их можно вычислить, решив для этого две вспомогательные задачи ЛП. Эти задачи состоят в максимизации и минимизации x_{i0} при условиях (10.24) и (10.25).

Теперь для каждого фиксированного целочисленного значения x_{i0} в найденном отрезке ($x_{i0\min}$, $x_{i0\max}$) находят $\min z$, решая задачу ЛП с ограничениями (10.24), (10.25) и с дополнительным ограничением $x_{i0} \le k_{i0}$.

Таким образом, все указанные выше возможности можно представить в виде некоторого дерева, в котором вершина 0 отвечает плану $\bar{X}_0$, а каждая из соединенных с ней вершин отвечает оптимальному плану следующей задачи: минимизировать z при условиях (10.24), (10.25) и дополнительном условии, что переменной x_{i0} дано значение $x_{i0} \le k_{i0}$, где k_{i0} – целое число. Каждой из таких вершин приписывают оценку $\xi^{(k)} = \xi(i_0, k)$, которая равна $\min z$ при указанных выше ограничениях. Очевидно, $\xi_0 \le \xi(i_0, k)$, для всех k.

Если оптимальные планы полученных задач удовлетворяют условиям целочисленности, то план с минимальной оценкой $\xi_0 = \xi(i_0, k)$ и будет оптимальным планом исходной задачи. В противном случае возникает необходимость в продолжении ветвления. При этом каждый раз для очередного ветвления выбирают вершину с наименьшей оценкой.

Любой маршрут в дереве от начальной вершины 0 до некоторой вершины определяет допустимую последовательность выбора целочисленных решений для переменных. Процесс продолжают до тех пор, пока продолжение ветвления становится невозможным.

Каждая конечная вершина отвечает некоторому допустимому целочисленному плану. Вершина с минимальной оценкой дает оптимальный план.

Рассмотрим алгоритм решения задачи целочисленного программирования. На первом этапе необходимо задать множество $G^{(0)}$, определяемое условиями (10.24), (10.25). Далее формируются множества $G_\nu^{(k)} (\nu = 1,\ 2,\ldots,\ \rho_k;\ k = 1,\ 2,\ldots)$, задаваемые условиями (10.24), (10.25) и дополнительным условием

$$x_j \le [x_{j0}] \text{ или } x_j \le [x_{j0}], \tag{10.27}$$

где $[x_{j0}]$ – целая часть x_{j0}.

Далее осуществляется вычисление оценок. Для множества $G^{(0)}$ оценку $\xi(G^{(0)})$ определяют как $\xi\left(G^{(0)}\right) = f\left(\bar{X}_0\right)$, где $\bar{X}_0$ – оптимальный план непрерывной задачи ЛП.

Для множества $G_\nu^{(k)}$ оценку $\xi(G_\nu^{(k)})$ определяют аналогично:

$$\xi\left(G_\nu^{(k)}\right)=f\left(\bar{X}_\nu^{(k)}\right),$$

где $\bar{X}_\nu^{(k)}$ – оптимальный план задачи с условиями (10.24), (10.25) и с дополнительным условием (10.27).

Если множество $G_\nu^{(k)}$ оказывается пустым, ему приписывают оценку $\xi(G_\nu^{(k)})=\infty$.

На следующем этапе осуществляется нахождение планов. Если план $\bar{X}_0$ удовлетворяет условию целочисленности (10.26), $\bar{X}_0$ – оптимальный план задачи. Если $\bar{X}_\nu^{(k)}$ удовлетворяет условию целочисленности (10.26), он является оптимальным планом задачи с условиями (10.24), (10.25), (10.27) и некоторым планом исходной задачи (10.23) – (10.26).

Далее выполняют ветвление. Ветвление производят в том случае, когда план $\bar{X}_\nu^{(k)}$ не удовлетворяет условию целочисленности (10.26).

Пусть $x_{\rho,\nu}^{(k)}$ – одна из нецелочисленных компонент плана, где $1\le\rho\le n_1$, тогда множество $G_\nu^{(k)}$ разбивают на два подмножества: $G_\nu^{(k)}=G_{\nu,1}^{(k)}\bigcup G_{\nu,2}^{(k)}$, причем

$$G_{\nu,1}^{(k)}=\left\{\bar{X}/\bar{X}\in G_\nu^{(k)},\ x_\rho\le\left[x_{\rho,\nu}^{(k)}\right]\right\}; \qquad (10.28)$$

$$G_{\nu,2}^{(k)}=\left\{\bar{X}/\bar{X}\in G_\nu^{(k)},\ x_\rho\ge\left[x_{\rho,\nu}^{(k)}\right]+1\right\}. \qquad (10.29)$$

Укажем некоторые особенности метода ветвей и границ для задач ЦП.

1. Если все коэффициенты c_j целевой функции – целые при $1\le j\le n_1$ и равны нулю при $j>n_1$, то оценку $\xi\left(G_\nu^{(k)}\right)$ можно заменить на более сильную оценку $\xi^1\left(G_\nu^{(k)}\right)=\left]f\left(\bar{X}_\nu^{(k)}\right)\right[$, где $]a[$ обозначено наименьшее целое, но не меньшее, чем a, т.е. округленное до ближайшего целого с избытком.

2. Алгоритм метода в вычислительном отношении представляет собой последовательность задач ЛП, причем конечность алгоритма следует из предполагаемой ограниченности множества G.

3. Из описания алгоритма следует, что в применении метода ветвей и границ для полностью целочисленных и для частично целочисленных задач нет никакой разницы.

Геометрически этот метод можно интерпретировать таким образом. Гипер-плоскость, определяемая целевой функцией задачи, вдавливается внутрь многогранника планов соответствующей задачи ЛП до встречи с ближайшей целочисленной точкой этого многогранника.

■ **Пример применения метода ветвей и границ для решения задач дискретного программирования.** Постановка задачи следующая: минимизировать функцию

$\min(-x_1+x_2)$ при ограничениях

$$\begin{cases}2x_1+11x_2\le 38;\\ x_1+x_2\le 7;\\ 4x_1-5x_2\le 5;\end{cases}$$

$x_1, x_2\ge 0$; x_1, x_2 – целые числа.

Нулевой шаг. Оптимальный план задачи ЛП $\bar{X}_0=\left(\frac{40}{9},\frac{23}{9}\right)$, тогда имеем $\xi(G_0)=\left]f(\bar{X}_0)\right[=]-7[=-7$.

Поскольку план X_0 не удовлетворяет условию целочисленности, возьмем его нецелочисленную компоненту $\bar{x}_1$ и разобьем множество $G^{(0)}$ на $G_1^{(1)}$ и $G_2^{(1)}$:

$$G_1^{(1)}=\left\{\bar{X}/\bar{X}\in G^0,\ x_1\le 4\right\};$$

$$G_2^{(1)}=\left\{\bar{X}/\bar{X}\in G^0,\ x_1\ge 5\right\}.$$

Первый шаг. Решаем две задачи ЛП, заключающиеся в минимизации исходной задачи по множествам $G_1^{(1)}$ и $G_2^{(1)}$.

В первом случае минимум достигается при

$$x_1=4,\ x_2=2\frac{8}{11}\ \text{ и }\ \xi\left(G_1^{(1)}\right)=\left]-6\frac{8}{11}\right[.$$

Множество $G_2^{(1)}$ оказывается пустым, поэтому $\xi\left(G_2^{(1)}\right)=\infty$.

Разбиваем $G_1^{(1)}$ на $G_{1,1}^{(1)}$ и $G_{1,2}^{(1)}$, где $G_{1,1}^{(1)}=\left\{\bar{X}/\bar{X}\in G_1^{(1)},\ x_2\le 2\right\}$ и $G_{1,2}^{(1)}=\left\{\bar{X}/\bar{X}\in G_1^{(1)},\ x_2\ge 3\right\}$. Обозначим $G_{1,1}^{(1)}=G_1^{(2)}$, $G_{1,2}^{(1)}=G_2^{(2)}$, $G_2^{(1)}=G_3^{(2)}$.

Второй шаг. Решаем две задачи ЛП, заключающиеся в минимизации исходной задачи при дополнительном ограничении $x_2\le 2$ и $x_2\ge 3$, тогда

$$X_1^{(2)} = \left\{3\frac{3}{4},\ 2\right\} \text{ и } \xi^1\left(G_1^{(2)}\right) = \left]-5\frac{3}{4}\right[= -5;$$

$$X_2^{(2)} = \left\{2\frac{1}{2},\ 3\right\} \text{ и } \xi^1\left(G_2^{(2)}\right) = \left]-5\frac{1}{2}\right[= -5;$$

$$G_3^{(2)} = 0,\ \xi^1\left(G_3^{(2)}\right) = \infty.$$

Производим разветвление множества $G_1^{(2)}$:

$$G_1^{(2)} = G_{1,1}^{(2)} \bigcup G_{1,2}^{(2)},$$

где $G_{1,1}^{(2)} = \left\{\bar{X} / \bar{X} \in G_1^{(2)},\ x_1 \le 3\right\}, G_{1,2}^{(2)} = \left\{\bar{X} / \bar{X} \in G_1^{(2)},\ x_1 \ge 4\right\}$.

Обозначим $G_{1,1}^{(2)} = G_1^{(3)},\ G_{1,2}^{(2)} = G_2^{(3)},\ G_2^{(2)} = G_3^{(3)},\ G_3^{(2)} = G_4^{(3)}$.

Третий шаг. Решаем две задачи ЛП, заключающиеся в минимизации исходной задачи по множествам $G_1^{(3)}$ и $G_2^{(3)}$.

Находим $X_1^{(3)} = \{3,2\}$ и $\xi^1(G_1^{(3)}) =]-5[= -5;\ G_2^{(3)} = 0$ и $\xi^1(G_2^{(3)}) = \infty$,

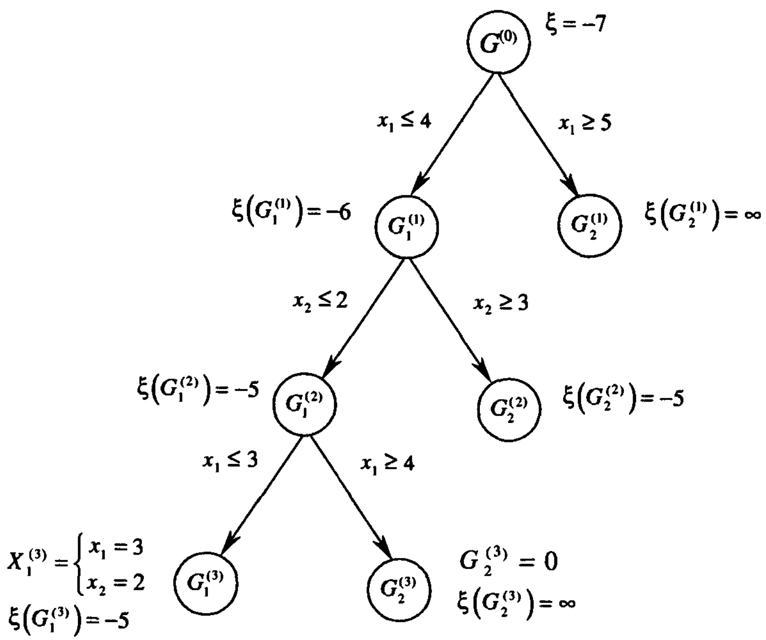

Рис. 10.2. Дерево решений задачи целочислеииого программирования

$X_3^{(3)} = X_2^{(2)} = \left\{2\frac{1}{2},\ 3\right\}$ и $\xi^1(G_3^{(3)}) = -5;\ G_4^3 = G_3^{(2)} = 0$ и $\xi^1(G_4^{(3)}) = \infty$.

Дерево решений приведено на рис. 10.2.

Итак получен целочисленный план $X_1^{(3)} = \{3,2\}$, причем

$$\xi^1 = \min\left\{\xi^1\left(G_1^{(3)}\right);\xi^1\left(G_2^{(3)}\right);\xi^1\left(G_3^{(3)}\right);\xi^1\left(G_4^{(3)}\right)\right\} = \min\{-5;\ \infty;\ -5;\ \infty\} = -5.$$

План $X_1^{(3)} = \{3,2\}$ – оптимальный, так как

$$f(X) = -5 \le \min\{\xi(G_1^{(3)}), \xi(G_1^{(3)}), \xi(G_1^{(3)})\}.$$

10.7. Нелинейное программирование

Постановки задач нелинейного программирования. Задачи нелинейного программирования на практике возникают довольно часто, например, когда затраты растут непропорционально количеству закупленных или произведенных товаров. Хорошо известно, что чем больше партия закупаемого товара, тем меньше стоимость единицы продукта. Любому покупателю знакомо понятие розничных и оптовых цен. Рассмотрим конкретный пример, иллюстрирующий данную ситуацию.

Планирование производства продукции. На действующем предприятии планируется организовать выпуск новых видов продукции. Для организации производства необходимо приобретать сырье, стоимость которого колеблется в зависимости от спроса. Цены на готовую продукцию предприятия предполагаются стабильными.

Итак, необходимо провести исследования о производстве двух видов изделий, для которых планируется приобретение сырья, цена которого зависит от объема закупаемой партии и спроса на рынке на сырье данного типа. Цены же на продукцию предприятия утверждены с учетом реальной обстановки и должны сохраняться неизменными. Объемы производства предстоит определить исходя из анализа сырьевой проблемы и ограниченности производственных ресурсов.

Пусть, x_1, x_2 – объемы производимой продукции 1-го и 2-го видов, c_1, c_2 – цена единицы продукции 1-го и 2-го видов соответственно. Затраты на приобретение и доставку сырья представляют собой нелинейную функцию, зависящую от объема закупаемого товара, $f_1(x_1), f_2(x_2)$. Таким образом, экономическая рентабельность планируемых мероприятий оценивается формулой

$$c_1x_1 + c_2x_2 - f_1(x_1) - f_2(x_2)\ .$$

Предприятие для производства новых видов продукции может выделить лишь часть своих мощностей, что накладывает дополнительные ограничения на максимальный объем выпуска новых видов изделий. Устанавливаются также лимиты на стоимость основных фондов (эксплуатация зданий, снабжение электроэнергией, амортизационные отчисления) в объеме b_1 и стоимость производственных процессов (вспомогательные материалы, заработная плата, накладные расходы и др.) в объеме b_2. Известно, что изготовление единицы продукции первого вида требует a_{11} затрат из основных фондов и a_{12} трудовых затрат, а единицы продукции второго вида затрат в размере a_{21} и a_{22} соответственно. Учет этих факторов приводит к условиям

$$a_{11}x_1 + a_{21}x_2 \le b_1;$$
$$a_{12}x_1 + a_{22}x_2 \le b_2.$$

Теперь можно сформулировать задачу: определить такие x_1, x_2, которые бы обеспечивали максимум функционала

$$\max z = c_1x_1 + c_2x_2 - f_1(x_1) - f_2(x_2)$$

при ограничениях

$$a_{11}x_1 + a_{21}x_2 \le b_1;$$
$$a_{12}x_1 + a_{22}x_2 \le b_2, x_1, x_2 \ge 0.$$

Таким образом сформулирована задача, в которой целевая функция является нелинейной.

Задача распределения удобрений. Будем рассматривать задачу распределения ограниченного количества удобрений между посевами n различных сельскохозяйственных культур. Предположим, что урожайность $f_i(x_i)$ культуры номера i является нелинейной вогнутой функцией от x_i – количества внесенных на единицу площади удобрений. Тогда урожай культуры номера i будет равен $s_i f_i(x_i)$, где s_i – площадь, занятая культурой номера i. Будем считать, что суммарная площадь фиксирована, т.е.

$$\sum_{i=1}^{n} s_i \le S, \tag{10.30}$$

где S – общая площадь, отводимая под посевы, заранее заданное число. Будем также считать, что продукция должна быть получена во вполне определенном ассортименте, т.е. должны иметь место равенства

$$\frac{s_i f_i(x_i)}{s_1 f_1(x_1)} = \lambda_i, \ i = 2, 3, \dots, n, \tag{10.31}$$

где λ_i – заданные числа. Введем ограничение

$$\sum_{i=1}^{n} s_i x_i \le X, \tag{10.32}$$

где X – суммарное количество удобрений.

Изменяя величины x_i, s_i так, чтобы не нарушить условий (10.30) – (10.32), мы будем получать различные варианты плана использования площади S. Эти планы необходимо научиться сравнивать между собой. Введем критерий следующим образом: обозначим через p_i цену единицы продукта номера i, через q цену единицы удобрений, тогда суммарный доход от продажи продукта за вычетом расходов на покупку удобрений равен

$$J(x,s) = \sum_{i=1}^{n} \left(p_i s_i f_i(x_i) - q s_i x_i\right); \tag{10.33}$$
$$x = (x_1, \dots, x_n), \ s = (s_1, \dots, s_n).$$

Требуется найти такой способ распределения земель, который максимизирует функционал (10.33) при ограничениях (10.30) – (10.32). Заметим, что величину X также можем считать искомой.

Формулировка задачи нелинейного программирования. Задача нелинейного программирования в общем виде формулируется следующим образом.

Найти

$$\max f(x_1, x_2, \dots, x_n) \tag{10.34}$$

при условиях

$$\begin{cases} g_1(x_1, x_2, \dots, x_n) \ge 0; \\ g_2(x_1, x_2, \dots, x_n) \ge 0; \\ \dots\dots\dots\dots\dots\dots\dots \\ g_m(x_1, x_2, \dots, x_n) \ge 0, \ x_1 \ge 0, \ x_2 \ge 0, \dots, \ x_n \ge 0, \end{cases} \tag{10.35}$$

где функции $f(x_1, x_2, \dots, x_n), g_i(x_1, x_2, \dots, x_n)$, $i = \overline{1, m}$ в общем случае нелинейны. Задачи условной оптимизации нелинейного программирования бывают двух типов: когда в ограничениях (10.35) имеют место 1) знаки равенства и 2) знаки неравенства. Рассмотрим вначале задачу условной оптимизации с ограничениями в виде равенств.

Ограничения в виде равенств. Начнем изложение метода условной оптимизации с рассмотрения примера максимизации функции двух переменных $z = f(x, y)$, где на x и y наложено ограничение, задаваемое уравнением $g(x, y) = 0$. Разрешим уравнение $g(x, y) = 0$ относительно y

22 — 4355

и решение представим в виде зависимости y от x, т.е. $y = h(x)$. Предположим, что функции $g(x, y)$ и $h(x)$ дифференцируемы, тогда можно определить производную функции $h(x)$. Она будет иметь внд:

$$\frac{dy}{dx} = \frac{dh(x)}{dx} = -\frac{\partial g}{\partial x} \Big/ \frac{\partial g}{\partial y}. \qquad (10.36)$$

Далее представим целевую функцию z как функцию одной независимой переменной $z = f(x, y) = f(x, h(x))$.

Необходимым условием максимума функции z будет соотношение

$$\frac{dz}{dx} = \frac{df}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0.$$

Подставляя в данное выражение формулу (10.36) и производя перегруппировку параметров, получаем

$$\frac{df}{dx} + \left(-\frac{\frac{\partial f}{\partial y}}{\frac{\partial g}{\partial y}} \right) \frac{\partial g}{\partial x} = 0. \qquad (10.37)$$

Выражение, стоящее в скобках, обозначим следующим образом:

$$\lambda = -\frac{\partial f(x,y)}{\partial y} \Big/ \frac{\partial g(x,y)}{\partial y}. \qquad (10.38)$$

Тогда можно отметить, что в точке максимума должны выполняться соотношения

$$g(x, y) = 0; \quad \frac{\partial f(x,y)}{\partial x} + \lambda \frac{\partial g(x,y)}{\partial x} = 0; \quad \frac{\partial f(x,y)}{\partial y} + \lambda \frac{\partial g(x,y)}{\partial y} = 0.$$

Первое выражение – это ограничение, которое указано в постановке задачи, второе следует непосредственно из (10.37) путем замены переменных (10.38), и наконец, третье выражение получается из (10.38) после умножения правой и левой частей на знаменатель выражения и переноса сомножителей в левую часть.

Получить эти три необходимых условия можно также, используя функцию Лагранжа $F(x, y, \lambda) = f(x, y) + \lambda g(x, y)$, которая представляет собой сумму целевой функции и произведения сомножителя λ на функцию ограничения. Коэффициент λ называется множителем Лагранжа. Тогда необходимые условия максимума функции $f(x,y)$ при наличии ограничений могут быть записаны в виде

$$\frac{\partial F(x,y,\lambda)}{\partial x} = \frac{\partial f(x,y)}{\partial x} + \lambda \frac{\partial g(x,y)}{\partial x} = 0;$$

$$\frac{\partial F(x,y,\lambda)}{\partial y} = \frac{\partial f(x,y)}{\partial y} + \lambda \frac{\partial g(x,y)}{\partial y} = 0;$$

$$\frac{\partial F(x,y,\lambda)}{\partial \lambda} = g(x,y) = 0.$$

Совместное решение данной системы уравнений относительно x, y и λ позволяет найти все точки, в которых имеет место условный экстремум.

Рассмотрим числовой пример.

Найти минимум функции $f(x, y) = x^2 + y^2$ при ограничении $x + y = 4$. Составим функцию Лагранжа; она будет иметь внд $F(x, y, \lambda) = x^2 + y^2 + \lambda(4 - x - y)$.

Соответствующие условия минимума можно записать следующим образом:

$$\frac{\partial F}{\partial x} = 2x - \lambda = 0;$$

$$\frac{\partial F}{\partial y} = 2y - \lambda = 0;$$

$$\frac{\partial F}{\partial \lambda} = 4 - x - y = 0.$$

Решением этой системы являются значения $x = y = 2$; $\lambda = 4$. Минимум функции равен 8. Таким образом, получено решение примера.

Необходимые условия задачи нелинейного программирования могут быть обобщены для функции n переменных при наличии m ограничений в виде равенств.

Рассмотрим задачу условной оптимизации функции

$$z = f(X) = f(x_1, x_2, \ldots, x_n),$$

где на вектор X наложены ограничения

$$g_1(X) = 0, g_2(X) = 0, \ldots, g_m(X) = 0.$$

Определим для этой задачи функцию Лагранжа в виде

$$F(X, \lambda) = f(X) + \sum_{i=1}^{m} \lambda_i g_i(X).$$

Необходимые условия экстремума функции $f(x)$ при наличии ограничений можно записать следующим образом:

$$\frac{\partial F}{\partial x_j}=\frac{\partial f}{\partial x_j}+\sum_{i=1}^{m}\lambda_i\frac{\partial g_i(X)}{\partial x_j}=0 \text{ при } j=1,\ldots,n;$$

$$\frac{\partial F}{\partial \lambda_i}=g_i(X)=0 \text{ при } i=1,\ldots,m.$$

Рассмотренный метод определения условного экстремума функции $z=f(X)$ называется методом множителей Лагранжа. Достоинство данного метода состоит в том, что он сводит задачу условной оптимизации к задаче безусловной оптимизации. Недостаток метода заключается в необходимости решения громоздкой системы уравнений, что далеко не всегда удается осуществить.

Ограничения в виде неравенств. Распространим метод множителей Лагранжа на задачи нелинейного программирования, в которых требуется осуществить условную оптимизацию функционала с заданными ограничениями в виде неравенств. Рассмотрим постановку задачи математического программирования в общем виде.

Оптимизировать функцию

$$z=f(X)=f(x_1, x_2,\ldots, x_n) \tag{10.39}$$

при наличии m ограничений

$$g_i(X)\le b_i \;\; (i=1, 2,\ldots, m),\;\; X=(x_1,x_2,\ldots x_n). \tag{10.40}$$

Такая запись задачи не ограничивает общности; если имеют место ограничения $\varphi(X)\ge c$, то их можно привести к виду $-\varphi(X)\le -c$. Экстремум целевой функции в этом случае может быть достигнут либо во внутренних точках области, заданной системой ограничений, либо на ее границах.

Рассмотрим метод решения задачи в указанной постановке.

Ограничения в виде неравенств (10.40) преобразуем в ограничения в виде равенств, добавляя к каждому из них неотрицательную ослабляющую переменную u_i^2:

$$g_i(X)+u_i^2=b_i \text{ или } g_i(X)+u_i^2-b_i=0. \tag{10.41}$$

Таким образом, свели задачу к минимизации функции $f(X)$ при наличии m ограничений в виде равенств. Сформируем функцию Лагранжа для задачи условной оптимизации с ограничениями в виде (10.41)

$$F(x,\lambda,u)=f(x)+\sum_{i=1}^{m}\lambda_i\left[g_i(x)+u_i^2-b_i\right].$$

Необходимые условия, которые должны выполняться в точке оптимума, записываются следующим образом:

$$\frac{\partial F}{\partial x_j}=\frac{\partial f}{\partial x_j}+\sum_{i=1}^{m}\lambda_i\frac{\partial g_i(X)}{\partial x_j}=0,\; j=\overline{1,\, n};$$

$$\frac{\partial F}{\partial \lambda_i}=g_i(X)+u_i^2-b_i=0,\; i=\overline{1,\, m};$$

$$\frac{\partial F}{\partial u_i}=2\lambda_i u_i=0,\; i=\overline{1,\, m}.$$

Умножим последнее уравнение на $\frac{u_i}{2}$, получим $\lambda_i u_i^2=0$ или $\lambda_i(b_i-g_i(X))=0$ для $i=\overline{1,\, m}$.

Данные уравнения являются необходимыми условиями оптимума в точке X^* при наличии ограничений в виде неравенств.

Есть также дополнительное условие, которое должно выполняться в зависимости от того, какая задача решается – минимизации или максимизации. Для задачи минимизации функции $f(X)$ при наличии ограничений в виде неравенств должно выполняться неравенство $\lambda_i\ge 0$. Для задачи максимизации требуется, чтобы $\lambda_i\le 0$, причем хотя бы для одного из множителей Лагранжа неравенство должно быть строгим. Окончательно можно записать необходимые условия минимума функции $f(X)$ при наличии ограничений $g_i(X)\le b_i$, $i=\overline{1,\, m}$:

$$\left.\begin{aligned}\frac{\partial f}{\partial x_j}+\sum\lambda_i\frac{\partial g_i}{\partial x_j}=0,\; j=\overline{1,\, n}\\ g_i(X)\le b_i,\; i=\overline{1,\, m}\\ \lambda_i(b_i-g_i(X))=0,\; i=\overline{1,\, m}\\ \lambda_i\ge 0,\; i=\overline{1,\, m}\end{aligned}\right\}.$$

Для задачи максимизации в последнем неравенстве знак меняется на противоположный.

Эти условия известны как условия Куна–Такера. Таким образом, проведено обобщение метода множителей Лагранжа на случай решения задачи условной оптимизации с ограничениями в виде неравенств.

Рассмотрим пример решения задачи нелинейного программирования с ограничениями в виде неравенств. Пусть требуется минимизировать функцию

$$f(x) = 3x_1^2 + 4x_1x_2 + 5x_2^2$$

при ограничениях $x_1 \geq 0;\ x_2 \geq 0;\ x_1 + x_2 \geq 4$.
Перепишем условие задачи следующим образом:

$$f(x) = 3x_1^2 + 4x_1x_2 + 5x_2^2,$$

$$-x_1 \leq 0;\ -x_2 \leq 0;\ -x_1 - x_2 \leq -4.$$

Преобразуем ограничения, добавив в левую часть ослабляющие переменные и изменив знак на равенство:

$$-x_1 + u_1^2 = 0;\ -x_2 + u_2^2 = 0;\ -x_1 - x_2 + u_3^2 = -4.$$

Далее запишем функцию Лагранжа

$$F(X,U,\Lambda) = 3x_1^2 + 4x_1x_2 + 5x_2^2 + \lambda_1(u_1^2 - x_1) + \lambda_2(u_2^2 - x_2) + \lambda_3(u_3^2 - x_1 - x_2 + 4).$$

Необходимые условия минимума для данной функции:

$$\begin{aligned}
&6x_1 + 4x_2 - \lambda_1 - \lambda_3 = 0;\\
&4x_1 + 10x_2 - \lambda_2 - \lambda_3 = 0;\\
&-x_1 \leq 0;\\
&-x_2 \leq 0;\\
&-x_1 - x_2 \leq -4;\\
&\lambda_1 x_1 = 0;\\
&\lambda_2 x_2 = 0;\\
&\lambda_3(4 - x_1 - x_2) = 0;\\
&\lambda_1, \lambda_2, \lambda_3 \geq 0.
\end{aligned}$$

Принимая во внимание три последних равенства, можно сделать вывод о том, что либо множители Лагранжа, либо искомые переменные должны быть равны нулю. При этом хотя бы один из множителей Лагранжа должен быть отличен от нуля. Положим $\lambda_1 = \lambda_2 = 0$, тогда (поскольку решается задача минимизации) должно выполняться неравенство $\lambda_3 > 0$, откуда следует $4 - x_1 - x_2 = 0$ или $4 - x_1 = x_2$. Подставляем данное выражение для параметра x_2 в два первых равенства и решаем их:

$$\begin{cases} 6x_1 + 4(4 - x_1) - \lambda_3 = 0;\\ 4x_1 + 10(4 - x_1) - \lambda_3 = 0. \end{cases}$$

$$\begin{cases} 6x_1 + 16 - 4x_1 - \lambda_3 = 0;\\ 4x_1 + 40 - 10x_1 - \lambda_3 = 0. \end{cases}$$

Приводя подобные члены, получим

$$\begin{cases} 2x_1 + 16 - \lambda_3 = 0;\\ -6x_1 + 40 - \lambda_3 = 0. \end{cases}$$

Перенесем множитель Лагранжа в правую часть:

$$\begin{cases} 2x_1 + 16 = \lambda_3;\\ -6x_1 + 40 = \lambda_3. \end{cases}$$

Решая данную систему, получим

$$x_1 = 3,\ x_2 = 4 - x_1 = 1,\ \lambda_3 = 22.$$

Нетрудно проверить, что эти решения являются условиями минимума и функция имеет минимальное значение, равное 44 в точке с координатами (3, 1).

Глава 11

СИСТЕМНЫЙ АНАЛИЗ И МОДЕЛИ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

11.1. Постановки задач, приводящие к моделям теории массового обслуживания

Модели теории массового обслуживания находят применение при решении задач системного анализа в случае, когда исследуемые величины имеют случайный характер. К числу таких задач относятся задачи управления запасами при случайном спросе, задачи организации работы предприятий торговли, связи, бытового и медицинского обслуживания, организации технического обслуживания предприятий, вопросы снабжения запасными частями и механизмами и т.п.

Рассмотрим основные понятия теории массового обслуживания. Каждая система массового обслуживания (СМО) состоит из некоторого числа обслуживающих объектов, называемых каналами обслуживания. Всякая СМО предназначена для обслуживания определенного потока заявок, поступающих в случайные моменты времени. Обслуживание заявки продолжается какое-то, в общем случае, случайное время, после чего канал освобождается и готов к приему следующей заявки. Случайный характер потока заявок и времен обслуживания приводит к тому, что в некоторые моменты времени на входе в СМО может образоваться очередь, в другие периоды времени каналы могут работать с недогрузкой или вообще простаивать.

Процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем. Состояния системы массового обслуживания меняются скачком в моменты реализации событий (поступление новой или окончание обслуживания заявки, момента, когда заявка, которой надоело ждать, покидает очередь).

Задача теории массового обслуживания состоит в построении моделей, связывающих заданные условия работы СМО с интересующими показателями эффективности системы, описывающими ее способность справляться с потоком заявок. В качестве показателей могут

применяться следующие: среднее число заявок, обслуживаемых в единицу времени; среднее число занятых каналов; среднее число заявок в очереди; среднее время ожидания обслуживания; вероятность того, что число заявок в очереди превысит некоторое число и т.д.

Во многих задачах теории массового обслуживания для определения необходимого показателя эффективности достаточно знать распределение входящего потока, дисциплину очереди (например, случайный выбор, обслуживание в порядке поступления или с приоритетом) и распределение времени обслуживания. В других задачах нужно иметь дополнительную информацию. Например, в случае отказов в обслуживании нужно определить вероятность того, что поступившее требование получит отказ сразу после прибытия или через некоторое время, т.е. покинет очередь до или после присоединения к ней.

С теоретической точки зрения очереди можно рассматривать как потоки, проходящие через систему пунктов обслуживания, соединенных последовательно или параллельно. На поток оказывают влияние различные факторы; они могут замедлять его, приводить к насыщению и т.д. При изучении систем массового обслуживания ключевыми характеристиками являются такие, как интенсивность входного потока требований и интенсивность обслуживания. В более общем случае могут использоваться распределение длительности интервалов времени между моментами поступления заявок на обслуживание и распределение длительности обслуживания.

Для всякой системы массового обслуживания (СМО) характерна структура, которая определяется составом элементов и связями между ними. Основными элементами системы являются

- входящий поток требований,
- каналы обслуживания,
- очередь требований,
- выходящий поток требований.

Схематично изображение системы массового обслуживания показано на рис. 11.1.

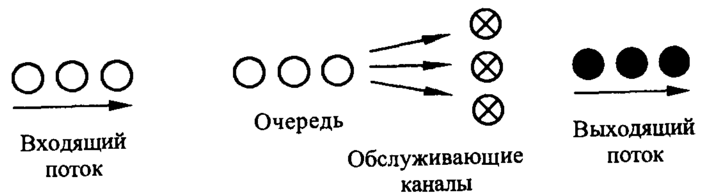

Рис. 11.1. Схема миогоканальной СМО с ожидаиием

Рассмотрим некоторые понятия теории массового обслуживания более подробно. Одно из исходных – понятие входящего потока требований, который представляет собой совокупность требований, поступающих в систему и нуждающихся в обслуживании. В начальный момент времени в системе может находиться некоторое число требований. Следующее требование поступает через случайное время, которое имеет определенный закон распределения. Случайны также длительности между моментами поступления последовательных требований. Промежутки между поступлениями требований во многих прикладных задачах можно считать взаимно независимыми. Однако существуют примеры, когда это предположение не выполняется. Скажем, при движении потока транспорта через перекресток интервалы между прохождением очередных машин будут зависимыми событиями.

В систему массового обслуживания требования могут поступать из конечной или бесконечной совокупности, которая может состоять из различных категорий требований. Требования каждой из категорий могут поступать с различным распределением, по одиночке или в составе группы, и занимать место в очереди в установившемся порядке. Распределение входящего потока требований может зависеть от распределения выходящего потока. Например, в больницу пациенты принимаются только при наличии освободившихся мест.

Каждой из СМО свойственна определенная организация. По составу СМО разделяют на одноканальные и многоканальные. Многоканальные системы могут состоять из каналов одинаковой и разной производительности (пропускной способности). СМО классифицируются по времени пребывания требования в системе до начала обслуживания. Выделяют каналы с неограниченным временем ожидания, каналы с отказами и каналы смешанного типа. В системах с неограниченным временем ожидания очередная заявка, застав все обслуживающие каналы занятыми, становится в очередь и дожидается до тех пор, пока один из обслуживающих каналов не освободится. В системах с отказами всякое вновь поступившее требование, застав все приборы занятыми, покидает систему. В системах смешанного типа поступившее требование, застав все приборы занятыми, становится в очередь. Но в ней оно может находиться ограниченное время, после истечения которого, не дождавшись обслуживания, требование покидает систему.

В системах с ожиданием используются следующие дисциплины обслуживания:

• первым пришел – первым обслужился (данная дисциплина находит наиболее широкое применение и обозначается как FIFO – используется аббревиатура английского названия first input – first output);

• дисциплина обслуживания по приоритетам;

• случайный выбор обслуживания.

При обслуживании по приоритетам в случае поступления в систему требования с более высоким приоритетом обслуживание остальных требований может прерываться либо оно может быть завершено. В редких случаях встречается дисциплина обслуживания «последним пришел – первым обслужился».

СМО классифицируют по порядку занятия свободных каналов вновь поступившими требованиями, а именно:

• каналы подключаются в строгом порядке, определяемом приоритетом использования;

• каналы обслуживают поступившие требования в порядке освобождения;

• каналы занимаются в случайном порядке.

В заключение параграфа еще раз подчеркнем, что общей особенностью всех задач теории массового обслуживания является случайный характер исследуемых явлений. Количество требований на обслуживание, временные интервалы между их поступлениями, длительность обслуживания – величины случайные. Основными задачами, решаемыми при анализе и исследовании систем массового обслуживания, являются определение характеристик системы, обеспечивающих заданное качество функционирования, минимизация времени ожидания клиентов в очереди, минимизация длины очереди и т.п.

11.2. Характеристика входящего потока требований

Процесс поступления в СМО потока требований является случайным. Будем рассматривать поток однородных событий, поступающих через случайные промежутки времени. Пусть E – однородные события, следующие одно за другим через случайные промежутки времени. Число n реализаций события E, происходящих в течение интервала времени τ, является случайной величиной, которую будем обозначать через $N(\tau)$. Вероятность того, что $N(\tau) = n$, будем обозначать $P_n(\tau)$.

Сделаем следующие предположения.

1. $P_n(\tau)$ зависит только от величины τ и не зависит от момента начала отсчета (условие стационарности потока).

2. $N(\tau)$ не зависит от числа реализаций события E, происшедших в любые предшествующие τ интервалы времени, т.е. случайная величина $N(\tau)$ не зависит от предыстории процесса (условие отсутствия последствий).

3. Вероятность того, что событие E произойдет более одного раза в интервале времени dt – бесконечно малая величина по сравнению с длительностью интервала dt. Вероятность того, что событие E произойдет один раз пропорциональна dt и равна λdt, λ – константа (условие ординарности потока).

Определим распределение вероятностей для простейшего потока (т.е. определим $P_n(\tau)$ для различных n, $n = 0, 1, 2,\ldots$). Для интервала любой длительности τ справедлива формула полной вероятности.

$$P_0(\tau)+P_1(\tau)+\ldots+P_k(\tau)+\ldots=1.$$

Рассмотрим в качестве временного интервала интервал Δt.

$$P_0(\Delta t)+P_1(\Delta t)+\ldots+P_k(\Delta t)+\ldots=1.$$

Применим условие ординарности при $\Delta t \to dt$: $P_k(dt)=0$ при $k > 1$, следовательно, $P_0(dt) + P_1(dt) = 1$.

Поскольку было отмечено, что $P_1(dt) = \lambda dt$, следовательно, $P_0=1 - \lambda dt$.

Произведем аналогичные вычисления для участка τ произвольной длительности. Рассмотрим интервал $[0, \tau+d\tau]$ и определим вероятность того, что в данном интервале произойдет ровно n событий E. Это возможно, когда

1) n событий реализовались в интервале $[0, \tau]$ и ни одного события в интервале времени $[\tau, \tau+d\tau]$;

2) $n - 1$ событие реализовано в интервале $[0, \tau]$ и одно событие в интервале времени $[\tau, \tau+d\tau]$.

Вероятность того, что в интервале времени $[\tau, \tau+d\tau]$ не произойдет ни одного события, посчитана и равна

$$P_0=1 - \lambda dt.$$

Вероятность того, что в интервале времени $[\tau, \tau+d\tau]$ произойдет одно событие, равна

$$P_1 = \lambda dt.$$

Запишем выражение для определения вероятности того, что на интервале $[0, \tau+d\tau]$ будет зафиксировано n событий

$$P_n(\tau+d\tau)=P_n(\tau)(1-\lambda d\tau)+P_{n-1}(\tau)\lambda d\tau.$$

Раскроем скобки:

$$P_n(\tau+d\tau)-P_n(\tau)=\lambda d\tau(P_{n-1}(\tau)-P_n(\tau)).$$

Произведем элементарные преобразования:

$$\frac{dP_n(\tau)}{d\tau}=\frac{P_n(\tau+d\tau)-P_n(\tau)}{d\tau}=\lambda(P_{n-1}(\tau)-P_n(\tau)), \qquad (11.1)$$

что справедливо для $n = 1, 2, \ldots$. При $n = 0$ имеем

$$\frac{dP_0(\tau)}{d\tau}=-\lambda P_0(\tau). \qquad (11.2)$$

Зададим начальные условия, состоящие в том, что априори постулируется отсутствие требований в начальный момент времени $P_0(0) = 1$, $P_n(0) = 0$, $n = 1, 2,\ldots$ С учетом начальных условий из последнего уравнения получаем $P_0(\tau) = e^{-\lambda\tau}$. Далее, подставляя выражение (11.2) в (11.1), получаем зависимость для вычисления вероятности нахождения в системе одного требования, и, продолжая процесс вычисления по индукции, приходим к следующим соотношениям:

$$P_1(\tau)=\frac{\lambda\tau e^{-\lambda\tau}}{1!};$$

$$\vdots$$

$$P_n(\tau)=\frac{(\lambda\tau)^n e^{-\lambda\tau}}{n!}.$$

Полученные выражения представляют собой распределение *Пуассона*. Отметим, что величина $\lambda\tau$ – среднее число событий E в интервале τ. Непосредственной подстановкой выражений полученной системы выводим следующую зависимость:

$$P_{n+1}(\tau)=\frac{\lambda\tau}{n+1}P_n(\tau).$$

Процесс многократного появления однородных событий через случайные интервалы времени при выполнении условий стационарности, ординарности и отсутствия последействия называется процессом Пуассона.

Распределение интервалов между двумя событиями. Определим закон распределения между двумя последовательными событиями.

Время поступления требований в систему подчиняется пуассоновскому закону распределения (процесс поступления требований – пуассоновский процесс), где θ – случайная величина, определяющая длину интервала между двумя соседними требованиями; $f(\theta)$ – плотность распределения случайной величины θ; $F(\theta)$ – функция распределения случайной величины θ $F(\theta) = P(\theta \le \Theta)$.

Вероятность того, что на рассматриваемом интервале времени не произойдет ни одного события, равна

$$P(\theta) = \frac{(\lambda\theta)^0}{0!} e^{-\lambda\theta} = e^{-\lambda\theta}.$$

Тогда плотность распределения случайной величины θ при пуассоновском процессе поступления требований будет иметь вид

$$f(\theta) = \lambda e^{-\lambda\theta}.$$

Моменты распределения определяются следующим образом:

$$M(\theta) = \frac{1}{\lambda}, \quad D(\theta) = \frac{1}{\lambda^2},$$

где λ – интенсивность потока требований.

Величина 1/λ представляет собой среднее значение длины интервала между двумя последовательными событиями, а величина λ – интенсивность потока требований или среднее число их за единицу времени.

Длительность интервала между (n, $n + m$) событиями. Рассмотрим процесс Пуассона с интенсивностью λ и определим закон распределения длительности интервала между n и $n + m$ событиями. Длительность этого интервала представляет собой сумму m подынтервалов, имеющих одно и то же распределение $f(\theta) = \lambda e^{-\lambda\theta}$. Распределение суммы одинаково распределенных подынтервалов определяется как свертка и имеет вид

$$f_m(\theta) = \frac{\lambda(\lambda\theta)^{m-1}}{(m-1)!} e^{-\lambda\theta}$$

– это распределение Эрланга. Его математическое ожидание и дисперсия выражаются соответственно следующим образом:

$$\bar{\theta} = M\theta = \frac{m}{\lambda}, \quad D\theta = \frac{m}{\lambda^2}.$$

Время обслуживания. Время обслуживания – это характеристика обслуживающего канала, которая определяет его пропускную способность. Предполагается, что время обслуживания – случайная величина $t_{обсл}$, которая полностью характеризуется законом распределения. В случае экспоненциального закона распределения времени обслуживания $G(t_{обсл}) = 1 - e^{-\mu t_{обсл}}$, где $\mu = \frac{1}{\bar{t}_{обсл}}$ – интенсивность обслуживания, т.е. предполагается, что время обслуживания как случайная величина подчиняется экспоненциальному закону распределения.

Показатели эффективности обслуживающих систем. Под эффективностью обслуживающей системы понимают характеристику уровня выполнения этой системой функций, для которых она предназначена.

Показатели эффективности зависят от трех факторов: 1) характеристик качества и надежности системы; 2) экономических показателей; 3) условий, в которых эксплуатируется система (комплекс внешних и внутренних факторов, параметры потока требований и пр.).

Эффективность систем обслуживания можно характеризовать большим числом всевозможных количественных показателей. Наиболее широкое применение находят следующие:

- вероятность потери требования в СМО; применительно к системам обслуживания эта величина равна вероятности того, что все обслуживающие каналы заняты;
- вероятность того, что занятыми являются k каналов обслуживания (P_k); P_0 – вероятность того, что все каналы свободны;
- среднее число занятых каналов $N_з = \sum_{k=1}^{m-1} kP_k$; этот показатель характеризует степень загрузки системы;
- среднее число каналов, свободных от обслуживания, $N_о = \sum_{k=0}^{m-1} (m-k)P_k$;
- коэффициент простоя каналов $K_п = \frac{N_о}{m}$;
- коэффициент загрузки каналов $K_з = \frac{N_з}{m}$.

Для систем с ожиданием используются дополнительные характеристики:

- среднее время ожидания требований (ждут сами требования);
- вероятность того, что время пребывания в очереди не превысит какой-либо определенной величины;
- средняя длина очереди;
- среднее число заявок в сфере обслуживания;
- вероятность того, что число требований в очереди, ожидающих начала обслуживания, больше некоторой заданной величины;

• стоимостные показатели: стоимость обслуживания каждого требования в системе; стоимость потерь в единицу времени, связанных с ожиданием заявки в очереди; стоимость потерь, связанных с уходом требования из системы; стоимость единицы времени простоя.

11.3. Система массового обслуживания с ожиданием

Система массового обслуживания с ожиданием – это такая система, которая имеет возможность ставить заявки в очередь, где эти заявки ожидают обслуживания. Системы массового обслуживания с ожиданием бывают разомкнутые и замкнутые. Разомкнутая система массового обслуживания – это система с неограниченным источником потока требований. Примерами разомкнутой системы массового обслуживания могут служить пассажиры, покупающие билеты в кассе, покупатели в магазине и т.п. Замкнутая система массового обслуживания – система, в которой поток требований ограничен. Примерами замкнутой системы может служить цех с ограниченным числом станков, которые требуется со временем ремонтировать (поток заявок – это станки, требующие ремонта), компьютерное оборудование некоторого предприятия, которое, выходя из строя, образует поток заявок на обслуживание. В замкнутых системах общее число циркулирующих требований конечно и чаще всего постоянно.

Разомкнутая система с одним каналом обслуживания. Рассмотрим следующую систему. Имеется один канал обслуживания. Время обслуживания распределено по экспоненциальному закону с параметром μ (μ – интенсивность обслуживания). Интенсивность входящего потока λ. Требования обслуживаются в порядке поступления в систему. Входящий поток требований пуассоновский. Процесс обслуживания не зависит от предыстории системы, зависит только от текущего состояния системы и потока поступающих требований. Процесс, описывающий состояние системы, будет марковским и полностью характеризуется матрицей переходных вероятностей. Описанный процесс представлен на рис. 11.2.

Введем обозначения: E_n – состояние, при котором в системе находится n требований; $P_n(t)$ – безусловная вероятность этого состояния в момент времени t.

По условию ординарности за время dt не может поступить больше одного требования. Вероятность поступления требования за отрезок времени длительностью dt равна λdt.

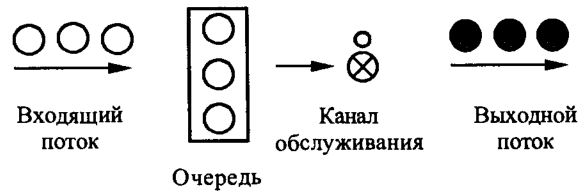

Рис. 11.2. Схема одноканальной разомкнутой СМО с ожиданием

Рассмотрим процесс функционирования такой системы. За отрезок времени длительностью dt возможны следующие переходы:

$$\begin{array}{l}
E_0 \to E_0;\quad E_0 \to E_1;\quad E_1 \to E_0;\\
E_1 \to E_1;\quad E_1 \to E_2;\quad E_2 \to E_1;\\
\vdots\\
E_i \to E_i;\quad E_i \to E_{i+1};\quad E_{i+1} \to E_i;\\
\vdots
\end{array}$$

Покажем, как вычисляются вероятности переходов. Пусть λdt – вероятность поступления нового требования (в интервале dt), μdt – вероятность окончания процесса обслуживания одного требования (в интервале dt).

Рассмотрим переход $E_n \to E_n$. Событие E_n при условии, что предыдущее событие было также E_n, может реализоваться двумя способами:

1) за время dt в систему поступило одно требование и одно требование системой обслужилось; вероятность такой реализации событий равна $\lambda dt \cdot \mu dt$;

2) в интервале времени dt ни одно требование не поступило и ни одно требование не обслужилось. Вероятность такой реализации соответственно равна $(1 - \lambda dt)(1 - \mu dt)$.

Данные события несовместны и поэтому можно записать

$$P(E_n \to E_n) = \lambda dt \cdot \mu dt + (1-\lambda dt)(1-\mu dt) = 1-\lambda dt - \mu dt.$$

В данном выражении мы пренебрегаем членом второго порядка малости $(dt)^2$.

По аналогии вычисляем вероятности для всех возможных переходов:

$$\begin{array}{l}
P(E_0 \to E_0) = 1-\lambda dt;\\
P(E_k \to E_{k+1}) = \lambda dt;\\
P(E_{k+1} \to E_k) = \mu dt;\\
P(E_k \to E_k) = 1-\lambda dt - \mu dt.
\end{array}$$

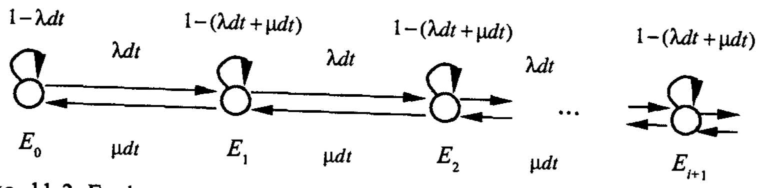

Рис. 11.3. Граф переходов одноканальной разомкнутой системы с ожиданием

Граф переходов одноканальной разомкнутой системы с ожиданием, иллюстрирующий функционирование системы массового обслуживания, изображен на рис. 11.3.

Вычислим теперь вероятности пребывания системы в некотором состоянии E_i в произвольный момент времени t. Вероятность нахождения процесса в некоторых состояниях в будущий момент времени исходя из состояния системы в настоящий момент будет вычисляться следующим образом.

Для состояния E_0

$$P_0(t+dt)=P_0(t)(1-\lambda dt)+P_1(t)\mu dt, \qquad (11.3)$$

где $P_0(t)$ – вероятность нахождения процесса в состоянии E_0 в момент времени t;$(1-\lambda dt)$ – вероятность остаться в состоянии E_0 за время dt (переход $E_0\to E_0$); $P_1(t)\mu dt$ – вероятность перехода $E_1\to E_0$. Приводя подобные члены в (11.3), получаем

$$\frac{dP_0(t)}{dt}=-\lambda P_0(t)+\mu P_1(t). \qquad (11.4)$$

Для состояния E_i

$$\begin{aligned}P_i(t+dt)&=P_i(t)(1-\lambda dt-\mu dt)+P_{i+1}(t)\mu dt+P_{i-1}(t)\lambda dt=\\&=P_i(t)(1-(\lambda+\mu)dt)+P_{i+1}(t)\mu dt+P_{i-1}(t)\lambda dt,\quad i\geq 1.\end{aligned} \qquad (11.5)$$

Производим с выражением (11.5) те же преобразования, что и с (11.3) и получаем

$$\frac{dP_i(t)}{dt}=\lambda P_{i-1}(t)-(\lambda+\mu)P_i(t)+\mu P_{i+1}(t),\ i\geq 1. \qquad (11.6)$$

Таким образом, получили систему дифференциальных уравнений Колмогорова – Чепмена для проведения расчетов вероятностей пребывания системы в некотором состоянии E_i в произвольный момент времени t.

Установившийся режим. Если при $t\to\infty$ для любого решения уравнений (11.4), (11.6) существует предел $\lim P_i(t)=P_i$ ($i=0,1,2,\ldots$), то в данной СМО имеет место установившийся режим. Такая система называется статистически устойчивой.

Величины P_i называются стационарным решением системы уравнений. Для стационарного решения производные от вероятностей $P_i(t)$, $i=0,1,2,\ldots$ равны нулю, следовательно,

$$\begin{aligned}&0=-\lambda P_0(t)+\mu P_1(t);\\&0=\lambda P_{i-1}(t)-(\lambda+\mu)P_i(t)+\mu P_{i+1}(t);\quad i\geq 1.\end{aligned} \qquad (11.7)$$

Решив систему (11.7), получим

$$\begin{aligned}&P_1=\frac{\lambda}{\mu}P_0;\\&P_k=\left(\frac{\lambda}{\mu}\right)^k P_0;\quad k\geq 1.\end{aligned} \qquad (11.8)$$

Отметим, что в любой момент времени система может находиться только в одном из состояний, следовательно, события несовместны:

$$\sum_{i=0}^{\infty}P_i=1$$

или

$$P_0+\sum_{i=1}^{\infty}P_i=1. \qquad (11.9)$$

Подставим из (11.8) в (11.9) выражения для соответствующих вероятностей, получим

$$1-P_0=\sum_{i=1}^{\infty}P_i=P_0\sum_{i=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^i. \qquad (11.10)$$

Обозначим $\frac{\lambda}{\mu}=\Psi$; в теории массового обслуживания данный показатель называется коэффициентом использования (загрузки) СМО. Если $\Psi>1$, установившегося режима не существует. Если $\Psi<1$, установившийся режим существует и возможно вычисление соответствующих вероятностей. Преобразуем выражение (11.10). Сгруппируем вероятности в левой части, а единицу перенесем в правую часть

$$P_0\sum_{k=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^k=1.$$

Данное выражение можно переписать как

$$P_0=\frac{1}{\sum_{k=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^k}.$$

Рассмотрим знаменатель последнего выражения. Применяя формулу для суммы бесконечно убывающей геометрической прогрессии и подставляя в полученное выражение обозначение для коэффициента загрузки СМО, получим

$$\sum_{k=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^k=\frac{1}{1-\frac{\lambda}{\mu}}=\frac{1}{1-\Psi};$$

$$P_0=1-\Psi; \tag{11.11}$$

$$P_k=\Psi^k P_0=\Psi^k(1-\Psi).$$

Таким образом, рассмотрены основные характеристики разомкнутой системы массового обслуживания с одним обслуживающим каналом. Подведем итог и перейдем к расчету основных числовых характеристик СМО.

Основные числовые характеристики системы

1. Среднее число требований в системе

$$E(N)=\sum_{n=1}^{\infty}nP_n=\sum_{n=1}^{\infty}n(1-\Psi)\Psi^n=\frac{\Psi}{1-\Psi}.$$

2. Средняя длина очереди

$$M=\sum_{n=1}^{\infty}(n-1)P_n=\frac{\Psi^2}{1-\Psi}.$$

3. Вероятность того, что в системе имеется хотя бы одно требование,

$$P(N>0)=1-P_0=\Psi.$$

4. Среднее время пребывания требования в системе вычислим, исходя из следующих соображений:

$$\frac{\lambda}{\mu}=1-P_0;$$

$$\lambda=\mu(1-P_0);$$

$$\bar{t}=\frac{E(N)}{\lambda}=\frac{E(N)}{\mu(1-P_0)}=\frac{\Psi}{1-\Psi}\frac{1}{\mu(1-P_0)}=\frac{1}{\mu(1-\Psi)}.$$

5. Среднее время ожидания в системе

$$\bar{t}_{\text{ож}}=\frac{M}{\lambda}=\frac{\Psi^2}{1-\Psi}\frac{1}{\lambda}=\frac{\Psi}{(1-\Psi)\mu}.$$

6. Среднее время обслуживания

$$\bar{t}_{\text{обсл}}=\frac{1}{\mu}=\bar{t}_{\text{преб}}-\bar{t}_{\text{ож}}.$$

Разомкнутая система с несколькими одинаковыми приборами. Рассмотрим обслуживающую систему, состоящую из m параллельно включенных каналов. Предположения относительно входящего потока и распределения времени обслуживания такие же, как в системе с одним каналом, т.е. интенсивность поступления требований λ. Исследуем случай, когда все каналы обслуживания идентичны и имеют производительность μ. Схема разомкнутой системы с μ каналами обслуживания приведена на рис. 11.4. Составим дифференциальные уравнения состояний системы. Обозначим через E_k состояние, при котором в системе обслуживания находится k требований. За отрезок времени длительностью dt в системе возможны следующие переходы:

$$E_0\to E_0;\ E_0\to E_1;\ E_1\to E_0,\ \ldots$$

$$E_n\to E_n;\ E_n\to E_{n+1};\ E_n\to E_{n-1},\ \ldots,\ n\geq 1.$$

Представим граф переходов в виде сетевого графика (см. рис. 11.5).

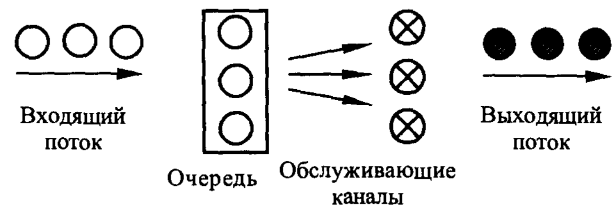

Рис. 11.4. Схема разомкнутой многоканальной СМО с ожиданием

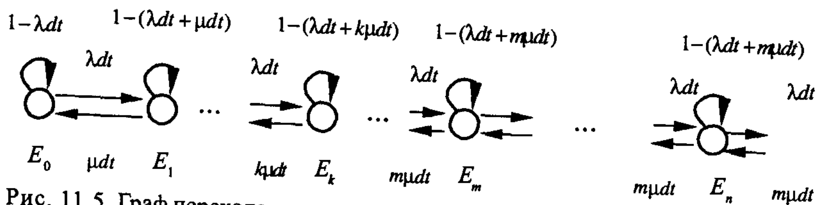

Рис. 11.5. Граф переходов одноканальной разомкнутой системы с ожиданием

Рассмотрим возможные переходы в системе и припишем каждому переходу соответствующие вероятности:

$$E_0 \to E_1 \qquad \lambda dt,$$
$$E_0 \to E_0 \qquad 1-\lambda dt,$$

$$\left.\begin{array}{ll} E_j \to E_{j+1} & \lambda dt, \\ E_j \to E_{j-1} & j\mu dt, \\ E_j \to E_j & 1-(\lambda + j\mu)dt, \end{array}\right. \qquad (j \le m)$$

$$\left.\begin{array}{ll} E_k \to E_{k+1} & \lambda dt, \\ E_{k+1} \to E_k & m\mu dt, \\ E_k \to E_k & 1-(\lambda + m\mu)dt. \end{array}\right. \qquad (k > m)$$

Запишем систему дифференциальных уравнений для указанных состояний в следующем виде :

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t) + \mu P_1(t);$$

$$\frac{d}{dt}P_k(t) = -(\lambda + k\mu)P_k(t) + \lambda P_{k-1}(t) + (k+1)\mu P_{k+1}(t), \text{ при } 1 \le k < m;$$

$$\frac{d}{dt}P_j(t) = -(\lambda + m\mu)P_j(t) + \lambda P_{j-1}(t) + m\mu P_{j+1}(t), \text{ при } j \ge m,$$

где $P_j(t)$ – вероятность состояния, когда в системе находится j требований.

Рассмотрим стационарный режим для СМО с несколькими одинаковыми каналами. Приравниваем производные нулю и получаем систему алгебраических уравнений:

$$-\lambda P_0 + \mu P_1 = 0;$$

$$-(\lambda + k\mu)P_k + \lambda P_{k-1} + (k+1)\mu P_{k+1} = 0, \text{ при } 1 \le k < m;$$

$$-(\lambda + m\mu)P_j + \lambda P_{j-1} + m\mu P_{j+1} = 0, \text{ при } j \ge n.$$

Решив данную систему уравнений, получим следующие основные показатели разомкнутой системы массового обслуживания.

1. Вероятность того, что в системе находится k требований, причем число требований меньше или равно числу обслуживающих каналов,

$$P_k = P_0 \frac{\psi^k}{k!}, \quad \psi = \frac{\lambda}{\mu}, \ k \le m.$$

2. Вероятность того, что в системе находится k требований, причем число требований больше числа обслуживающих каналов,

$$P_k = P_0 \frac{\psi^k}{m!m^{k-m}}, \ k > m.$$

3. Вероятность того, что в системе отсутствуют требования P_0, можно получить из условия $\sum_{k=0}^{\infty} P_k = 1$, тогда $P_0\left\{\sum_{k=0}^{m-1}\frac{\psi^k}{k!} + \sum_{k=m}^{\infty}\frac{\psi^k}{m!m^{k-m}}\right\} = 1$, откуда следует

$$P_0 = \frac{1}{\sum_{k=0}^{m-1}\frac{\psi^k}{k!} + \sum_{k=m}^{\infty}\frac{\psi^k}{m!m^{k-m}}} = \frac{1}{\sum_{k=0}^{n-1}\frac{\psi^k}{k!} + \frac{\psi^n}{m!\left(1-\frac{\psi}{m}\right)}}.$$

4. Вероятность того, что все каналы заняты обслуживанием и s требований находится в очереди

$$P_{m+s} = \frac{\psi^{m+s}}{m!m^s}P_0, \ s > 0.$$

5. Вероятность того, что время пребывания в очереди больше некоторой заданной величины t

$$P(\tau > t) = \pi e^{-\mu(m-\psi)t}, \text{ где } \pi = \sum_{k=m}^{\infty} P_k.$$

6. Среднее время ожидания

$$\bar{t}_{\text{ож}} = \frac{\pi \bar{t}_{\text{обсл}}}{(m-\psi)}, \text{ при } \frac{\psi}{m} < 1, \ \bar{t}_{\text{обсл}} = \frac{1}{\mu}.$$

7. Средняя длина очереди

$$M_о = \frac{\psi\pi}{m\left(1-\frac{\psi}{n}\right)^2}.$$

8. Среднее число свободных от обслуживания приборов

$$N_0 = \sum_{k=1}^{m-1} \frac{m-k}{k!} \psi^k P_0.$$

9. Коэффициент простоя $k_р = \frac{N_0}{m}$.

10. Среднее число занятых обслуживанием приборов $N_з = m - N_0$.

11. Коэффициент загрузки $k_з = \frac{N_з}{m}$.

12. Среднее число требований, находящихся в системе,

$$M = M_0 + P_0 \sum_{k=1}^{m} \frac{\psi^k}{(k-1)!}.$$

11.4. Замкнутые системы с ожиданием

При рассмотрении разомкнутых систем предполагалось, что источник обладает неограниченным числом требований. В настоящем параграфе рассмотрим случай, когда система предназначена для обслуживания конечного, постоянного числа требований. Будем предполагать, что как только требование обслужилось оно возвращается в источник. Схема такой системы изображена на рис. 11.6.

Основное отличие разомкнутой системы от замкнутой состоит в том, что в разомкнутой системе интенсивность поступления требований – характеристика источника требований. В замкнутой системе потенциальное число требований является величиной постоянной. После обслуживания требования оно возвращается в источник. В замкнутой системе интенсивность поступления требований λ – характеристика конкретного объекта, поступающего в систему.

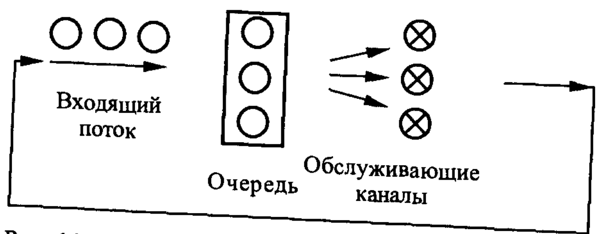

Рис. 11.6. Схема замкнутой многоканальной СМО с ожиданием

Рассмотрим пример замкнутой системы.

В организации имеется парк вычислительной техники в размере n штук и группа, обслуживающая вычислительную технику. В случае отказа компьютера он поступает в ремонт в указанную группу. Пусть λ – интенсивность отказа одной единицы техники (характеристика объекта). Данная величина характеризует интенсивность поступления на обслуживание данного объекта (только его одного). Потенциальное число заявок на обслуживание постоянное и равно n. Интенсивность входного потока требований зависит от числа исправно работающих объектов в источнике. В случае, когда все единицы вычислительной техники исправны, интенсивность потока требований равна $n\lambda$, после того как один компьютер откажет она станет $(n - 1)\lambda$ и т.д. За время $(t, t+dt)$ объект может потребовать обслуживания с вероятностью λdt. За время $(t, t+dt)$ объект, находящийся на обслуживании, может быть обслужен с вероятностью μdt. Если в некоторый момент времени число объектов, ожидающих обслуживания и обслуживаемых, равно k, то число объектов в источнике равно $n - k$. Вероятность поступления заявки на обслуживание хотя бы одного из данных объектов в интервале времени длительностью dt равна $(n - k)\lambda dt$. Таким образом, интенсивность потока требований изменяется скачкообразно всякий раз, когда компьютер выходит из строя, и возникает необходимость в его обслуживании.

Процесс размножения и гибели. Рассмотрим систему обслуживания, в которой возможны изменения состояний: $E_k \to E_{k+1}$; $E_k \to E_{k-1}$; $E_k \to E_k$; $k \geq 1$. Если в момент времени t система находится в состоянии E_k, то вероятность перехода $E_k \to E_{k+1}$ в интервале длительностью dt равна $\lambda_k dt$. Вероятность перехода $E_n \to E_{n-1}$ в интервале длительностью dt равна $\mu_n dt$. Вероятность перехода $E_n \to E_{n+k(-k)}$, $k \geq 2$ в интервале длительностью dt – бесконечно малая величина по сравнению с dt. Параметры λ_k и μ_k зависят только от k, где n – число требований в системе. Граф переходов для рассматриваемого случая приведен на рис. 11.7.

Число состояний графа конечно и определяется числом элементов в источнике. Для данного графа можно записать дифференциальные уравнения состояний, которые называются уравнениями размножения и гибели:

$$\frac{dP_0(t)}{dt} = -\lambda_0 P_0(t) + \mu_1 P_1(t);$$

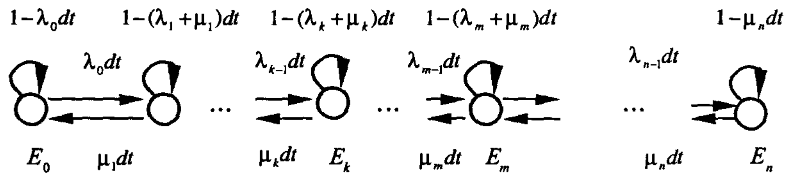

Рис. 11.7. Граф переходов замкнутой системы (процесс размножения и гибели)

$$\frac{d}{dt}P_k(t) = -(\lambda_k + \mu_k)P_k(t) + \lambda_{k-1}P_{k-1}(t) + \mu_{k+1}P_{k+1}(t);\ k \ge 1.$$

Используя эти уравнения, можно перейти к частным случаям исследования систем, если определить все λ_k и μ_k.

Замкнутые системы при $n > m$. Пусть n – потенциальное число требований, участвующих в процессе массового обслуживания; m – число каналов; μ – интенсивность обслуживания требования одним каналом. Будем считать, что все каналы идентичны. Интенсивность входящего потока зависит от числа поступивших требований. Если k – число поступивших требований, то $\lambda_k = (n-k)\lambda$.

Интенсивность обслуживания системы также зависит от числа требований и вычисляется как

$$\mu_k = k\mu,\quad 1 \le k \le m;$$
$$\mu_r = m\mu,\quad r \ge m.$$

Граф переходов, соответствующий этому случаю, идентичен изображенному на рис. 11.7; при этом интенсивности переходов будут иметь значения

$$\lambda_0 = n\lambda, \lambda_1 = (n-1)\lambda, \ldots, \lambda_{n-1} = \lambda;$$
$$\mu_1 = \mu, \mu_2 = 2\mu, \ldots, \mu_m = m\mu, \ldots, \mu_n = m\mu.$$

Дифференциальные уравнения для данного графа состояний:

$$\frac{d}{dt}P_0(t) = -n\lambda P_0(t) + \mu P_1(t);$$
$$\frac{d}{dt}P_k(t) = -\left[(n-k)\lambda + k\mu\right]P_k(t) + (n-k+1)\lambda P_{k-1}(t) + (k+1)\mu P_{k+1}(t);\quad k < m;$$
$$\frac{d}{dt}P_j(t) = -\left[(n-j)\lambda + m\mu\right]P_j(t) + (n+1-j)\lambda P_{j-1}(t) + m\mu P_{j+1}(t);\quad m \le j < n;$$
$$\frac{d}{dt}P_n(t) = -\lambda P_{n-1}(t) + m\mu P_n(t).$$

Для установившегося режима получим стационарное решение:

$$n\lambda P_0 = \mu P_1;$$
$$\left[(n-k)\lambda + k\mu\right]P_k = (n-k+1)\lambda P_{k-1} + (k+1)\mu P_{k+1};\quad k < m;$$
$$\left[(n-j)\lambda + m\mu\right]P_j = (n+1-j)\lambda P_{j-1} + m\mu P_{j+1};\quad m \le j < n;$$
$$\lambda P_{n-1} = m\mu P_n.$$

Используя обозначения ТМО

$$\omega_k = \frac{P_{k+1}}{P_k};\ \Psi = \frac{\lambda}{\mu},$$

после элементарных преобразований получим:

$$\omega_0 = \frac{P_1}{P_0} = n\Psi,$$

$$\omega_k = \frac{(n-k)\Psi + k}{k+1} - \frac{n-k+1}{k+1}\frac{\Psi}{\omega_{k-1}},\quad k < m;$$
$$\omega_j = \frac{(n-j)\Psi + m}{m} - \frac{n-j+1}{m}\frac{\Psi}{\omega_{j-1}},\quad m \le j \le n.$$

Последовательно решаем данную систему для $k = 1, 2, \ldots, m, \ldots, n$; получим

$$\omega_k = \frac{(n-k)\Psi}{k+1};\ k < m,\ \omega_j = \frac{(n-j)\Psi}{m};\ m \le j \le n.$$

Теперь можно выразить вероятности P_k через P_0:

$$\frac{P_k}{P_0} = \prod_{i=0}^{k-1}\omega_i = \prod_{i=0}^{k-1}\frac{n-i}{i+1}\Psi = \frac{n(n-1)\ldots(n-k+1)}{1\cdot 2\cdot\ldots\cdot k}\Psi^k = \frac{n!}{(n-k)!k!}\Psi^k,\ 1 \le k \le m;$$
$$\frac{P_k}{P_0} = \left[\prod_{i=0}^{m-1}\omega_i\prod_{i=m}^{k-1}\omega_i\right] = \prod_{i=0}^{m-1}\frac{n-i}{i+1}\Psi\prod_{i=m}^{k-1}\frac{n-i}{m}\Psi = \frac{n(n-1)\ldots(n-m+1)}{1\cdot 2\cdot\ldots\cdot m}\Psi^m \times \tag{11.12}$$
$$\times\frac{(n-m)(n-m-1)\ldots(n-k+1)}{m^{k-m}}\Psi^{k-m} = \frac{n!}{m!(n-k)!m^{k-m}}\Psi^k,\ m \le k \le n.$$

Принимая во внимание, что сумма всех вероятностей равна единице, можно записать

$$\left[\sum_{k=0}^{m}\frac{n!}{(n-k)!k!}\Psi^k + \sum_{k=m+1}^{n}\frac{n!}{m!(n-k)!m^{k-m}}\Psi^k\right]P_0 = 1,$$

откуда имеем

$$P_0 = \frac{1}{\left[\sum_{k=0}^{m} \frac{n!}{(n-k)!k!}\Psi^k + \sum_{k=m+1}^{n} \frac{n!}{m!(n-k)!m^{k-m}}\Psi^k\right]}.$$

Далее можно определить числовые характеристики системы.

1. Вероятность того, что в системе находится k требований, определяется из выражения (11.12).

2. Среднее число требований, ожидающих обслуживания,

$$\mathrm{M}_{\text{ож}} = \sum_{k=m+1}^{n} (k-m)P_k.$$

3. Среднее число требований, находящихся в системе обслуживания,

$$\mathrm{M} = \mathrm{M}_{\text{ож}} + \sum_{k=1}^{m} \frac{kn!}{(n-k)!k!}\Psi^k P_0.$$

4. Среднее число свободных каналов в установившемся режиме

$$N_0 = \sum_{k=0}^{m-1}(m-k)P_k.$$

5. Коэффициент простоя требований, ожидающих обслуживания,

$$k_{\text{р}} = \frac{\mathrm{M}_{\text{ож}}}{m}.$$

6. Коэффициент простоя каналов обслуживания

$$k_{\text{п}} = \frac{N_0}{m}.$$

11.5. Пример расчета надежности системы с ограниченным количеством запасных элементов

Рассмотрим применение методов теории массового обслуживания к задаче анализа показателей надежности систем, имеющих запасные элементы. Будем рассматривать объекты, работающие в составе промышленных установок, являющихся источниками повышенного риска. Характерным примером таких установок являются энергоблоки атомных электростанций.

Объекты ядерной энергетики имеют особенность, отличающую их от других технических объектов и состоящую в том, что к их показателям надежности предъявляются высокие требования. Так коэффициент неготовности (или вероятность невыполнения задачи) для каналов системы аварийной защиты должен быть не более чем 10^{-7}. Высокие требования предъявляются также к точности проведения расчетов.

Объекты систем ядерных энергетических установок относятся к классу высоконадежных объектов. Отказы их – события редкие. Наработки элементов до отказа сравнимы по порядку с общим временем эксплуатации системы. Высокие требования к точности результатов расчетов приводят к тому, что нельзя пренебрегать временем восстановления объектов после выявления факта отказа. В связи с вышеизложенным имеются особенности в решении задач анализа надежности указанных объектов.

Рассмотрим постановку задачи.

Требуется провести расчет характеристик надежности комплекта рабочий элемент – запасные элементы.

Стратегия функционирования элемента следующая. В начальный момент времени элемент находится в исправном состоянии. С интенсивностью $\lambda(t)$ элемент отказывает. В случае отказа элемент заменяется на резервный. Интенсивность замены элемента $\mu(t)$. Неисправный элемент отправляется в ремонт. После ремонта элемент считается восстановившим работоспособность и переходит в резерв. Интенсивность ремонта $\nu(t)$. Если исправных элементов в резерве не осталось, наступает отказ. Описанная стратегия функционирования может быть представлена с помощью графа, приведенного на рис. 11.8.

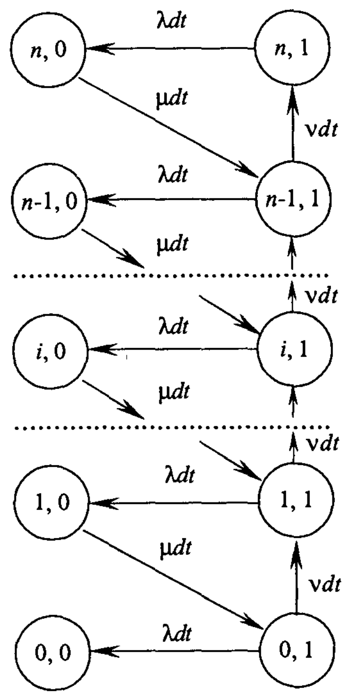

Рис. 11.8. Модель функционирования объекта с запасными элементами

Состояние объекта на графе обозначим двумя символами (k, i), где первый символ означает количество запасных элементов, $k = 0 \ldots n$, второй символ – состояние основного элемента, находящегося под нагрузкой, $i = 1$ элемент работоспособен, $i = 0$ элемент неработоспособен.

Рассмотрим функционирование объекта с запасными элементами более подробно. В начале работы элемент находится с вероятностью 1 в состоянии (n, 1) (в наличии имеется n запасных элементов, объект работоспособен). В случайный момент времени с интенсивностью отказа $\lambda(t)$ элемент переходит в состояние (n, 0) (n запасных элементов, объект в состоянии отказа, начинается замена элемента). С интенсивностью восстановления $\mu(t)$ объект переходит в состояние (n – 1, 1) ((n – 1) запасной элемент, объект работоспособен), из этого состояния возможны переходы в состояние $(n,1)$ с интенсивностью восстановления $\nu(t)$ (ремонт окончен, в резерве опять n элементов) или в состояние (n – 1, 0) с интенсивностью $\lambda(t)$ (ремонт не закончен до наступления следующего отказа) и так далее. Состояние (0, 0) является поглощающим и означает отказ объекта и отсутствие запасных элементов.

Рассмотренная стратегия функционирования может быть описана марковским процессом и представлена в виде системы дифференциальных уравнений

$$\begin{aligned}
&dP_{n,1}(t)/dt = -\lambda(t)P_{n,1}(t) + \nu(t)P_{n-1,1}(t);\\
&dP_{n,0}(t)/dt = -\mu(t)P_{n,0}(t) + \lambda(t)P_{n,1}(t);\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&dP_{i,1}(t)/dt = \mu(t)P_{i+1,0}(t) + \nu(t)P_{i-1,1}(t) - (\lambda(t)+\nu(t))P_{i,1}(t);\\
&dP_{i,0}(t)/dt = \lambda(t)P_{i,1}(t) - \mu(t)P_{i,0}(t);\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&dP_{0,1}(t)/dt = \mu(t)P_{1,0}(t) - (\lambda(t)+\nu(t))P_{0,1}(t);\\
&dP_{0,0}(t)/dt = \lambda(t)P_{0,1}(t).
\end{aligned} \qquad (11.13)$$

В большинстве случаев систему (11.13) можно упростить, если положить параметры модели $\lambda(t)$, $\mu(t)$, $\nu(t)$ постоянными величинами. Для электронных блоков и элементов после завершения периода приработки параметр потока отказов можно считать константой: $\lambda(t) = \lambda$. Аналогичные допущения можно сделать и для величин $\mu(t) = \mu$, $\nu(t) = \nu$. Тогда система (11.13) может быть записана в виде

$$\begin{aligned}
&dP_{n,1}(t)/dt = -\lambda P_{n,1}(t) + \nu P_{n-1,1}(t);\\
&dP_{n,0}(t)/dt = -\mu P_{n,0}(t) + \lambda P_{n,1}(t);\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&dP_{i,1}(t)/dt = \mu P_{i+1,0}(t) + \nu P_{i-1,1}(t) - (\lambda+\nu)P_{i,1}(t);\\
&dP_{i,0}(t)/dt = \lambda P_{i,1}(t) - \mu P_{i,0}(t);\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
&dP_{0,1}(t)/dt = \mu P_{1,0}(t) - (\lambda+\nu)P_{0,1}(t);\\
&dP_{0,0}(t)/dt = \lambda P_{0,1}(t);\\
&i = 1 \dots n-1.
\end{aligned} \qquad (11.14)$$

В общем случае при больших n решение системы вызывает значительные трудности. В частных случаях, задаваясь конкретным значением n – числа запасных элементов, решение системы можно получить аналитически. Покажем возможность аналитического решения для случая одного запасного элемента. Запишем систему дифференциальных уравнений:

$$\begin{aligned}
&dP_{1,1}(t)/dt = -\lambda P_{1,1}(t) + \nu P_{0,1}(t);\\
&dP_{1,0}(t)/dt = -\mu P_{1,0}(t) + \lambda P_{1,1}(t);\\
&dP_{0,1}(t)/dt = \mu P_{1,0}(t) - (\lambda+\nu)P_{0,1}(t);\\
&dP_{0,0}(t)/dt = \lambda P_{0,1}(t),
\end{aligned}$$

преобразуем эту систему с помощью преобразований Лапласа

$$\begin{aligned}
&(p+\lambda)R(p)_{1,1} - \nu R(p)_{0,1} = 1;\\
&(p+\mu)R(p)_{1,0} - \lambda R(p)_{1,1} = 0;\\
&(p+\lambda+\nu)R(p)_{0,1} - \mu R(p)_{1,0} = 0;\\
&pR(p)_{0,0} - \lambda R(p)_{0,1} = 0;
\end{aligned}$$

решим данную систему относительно $R_{i,j}(p)$, получим

$$R(p)_{1,1}=\frac{(p+\mu)(p+\lambda+\nu)}{p^3+(2\lambda+\nu+\mu)p^2+(\nu\mu+\lambda^2+\lambda\nu+2\lambda\mu)p+\lambda^2\mu};$$

$$R(p)_{1,0}=\frac{p+\lambda+\nu}{p^3+(2\lambda+\nu+\mu)p^2+(\nu\mu+\lambda^2+\lambda\nu+2\lambda\mu)p+\lambda^2\mu};$$

$$R(p)_{0,1}=\frac{\lambda\mu}{p^3+(2\lambda+\nu+\mu)p^2+(\nu\mu+\lambda^2+\lambda\nu+2\lambda\mu)p+\lambda^2\mu};$$

$$R(p)_{0,0}=\frac{\lambda^2\mu}{p(p^3+(2\lambda+\nu+\mu)p^2+(\nu\mu+\lambda^2+\lambda\nu+2\lambda\mu)p+\lambda^2\mu)}.$$

Выразим знаменатель данных соотношений в виде произведения

$$p^3+(2\lambda+\nu+\mu)p^2+(\nu\mu+\lambda^2+\lambda\nu+2\lambda\mu)p+\lambda^2\mu=(p-a)(p-b)(p-c),$$

где a, b и c являются корнями уравнения

$$p^3+(2\lambda+\nu+\mu)p^2+(\nu\mu+\lambda^2+\lambda\nu+2\lambda\mu)p+\lambda^2\mu=0.$$

(Аналитическое выражение корней через λ, μ и ν не приводится, в силу своей громоздкости.) После применения операции обратного преобразования Лапласа получим следующие результаты:

$$P_{1,1}=\exp(at)\left(\frac{\mu(\lambda+\nu)+a\mu+a(\lambda+\nu)+a^2}{bc-ac-ab+a^2}\right)-$$
$$-\exp(bt)\left(\frac{\mu(\lambda+\nu)+(\lambda+\nu)b+\mu b+b^2}{-b^2+ab+bc-ac}\right)+\exp(ct)\left(\frac{c^2+(\lambda+\nu)c+\mu c+\mu(\lambda+\nu)}{-bc+ab+c^2-ac}\right);$$

$$P_{1,0}=\exp(at)\left(\lambda\frac{-(\lambda+\nu)-a}{-bc+ab+ac-a^2}\right)+\exp(bt)\left(\lambda\frac{(\lambda+\nu)+b}{-bc-ab+ac+b^2}\right)+$$
$$+\exp(ct)\left(\lambda\frac{(\lambda+\nu)+c}{-ac-bc+ab+c^2}\right);$$

$$P_{0,1}=-\lambda\mu\frac{\exp(at)}{(-bc+ab+ac-a^2)}+\lambda\mu\frac{\exp(bt)}{(-bc-ab+ac+b^2)}+\lambda\mu\frac{\exp(ct)}{(-ac-bc+ab+c^2)};$$

$$P_{0,0}=-\lambda^2\mu\frac{\exp(at)}{a(-bc+ab+ac-a^2)}+\lambda^2\mu\frac{\exp(bt)}{b(-bc-ab+ac+b^2)}+$$
$$+\lambda^2\mu\frac{\exp(ct)}{c(-ac-bc+ab+c^2)}+\frac{\lambda^2\mu}{abc}.$$

Для системы дифференциальных уравнений типа (11.14) стационарного режима не существует, так как при времени работы, стремящемся к бесконечности, вероятность попасть в поглощающее состояние стремится к единице.

Однако систему (11.14) можно упростить, если записать условно-стационарное состояние. Для этого положим равным нулю все производные, стоящие в левых частях уравнений, кроме последнего. В результате такого допущения мы сознательно увеличиваем ошибку итогового результата, но такой подход, с одной стороны, позволяет существенно упростить решение, с другой стороны, сохраняет зависимость вероятностей от времени.

В итоге получим систему

$$\begin{aligned}
&0=-\lambda P_{n,1}(t)+\nu P_{n-1,1}(t);\\
&0=-\mu P_{n,0}(t)+\lambda P_{n,1}(t);\\
&\dots\dots\dots\dots\dots\dots\\
&0=\mu P_{i+1,0}(t)+\nu P_{i-1,1}(t)-(\lambda+\nu)P_{i,1}(t);\\
&0=\lambda P_{i,1}(t)-\mu P_{i,0}(t);\\
&\dots\dots\dots\dots\dots\dots\\
&0=\mu P_{1,0}(t)-(\lambda+\nu)P_{0,1}(t);\\
&dP_{0,0}(t)/dt=\lambda P_{0,1}(t);\\
&i=1\dots n-1.
\end{aligned}$$

Произведем элементарные преобразования, запишем итоговый результат

$$\begin{aligned}
&P_{1,0}=\frac{\lambda+\nu}{\mu}P_{0,1};\\
&\dots\dots\dots\dots\dots\dots\\
&P_{i,1}=\frac{\mu}{\lambda}P_{i,0};\\
&P_{i+1,0}=\frac{\lambda+\nu}{\lambda}P_{i,1}-\frac{\nu}{\mu}P_{i-1,1};\\
&\dots\dots\dots\dots\dots\dots
\end{aligned}$$

$$P_{n,1} = \frac{\mu}{\lambda} P_{n,0};$$

$$P_{n-1,1} = \frac{\lambda}{\nu} P_{n,0}.$$

Из последних соотношений видно, что каждая из вероятностей системы может быть выражена через $P_{0,1}(t)$, например

$$P_{1,1}(t) = \frac{\mu}{\lambda} P_{1,0}(t) = \frac{\lambda+\nu}{\lambda} P_{0,1}(t);$$

$$P_{2,0}(t) = \frac{\lambda+\nu}{\mu} P_{1,1}(t) - \frac{\nu}{\mu} P_{0,1}(t) = \frac{(\lambda+\nu)^2}{\lambda\mu} P_{0,1}(t) - \frac{\nu}{\mu} P_{0,1}(t);$$

$$P_{2,1}(t) = \frac{\mu}{\lambda} P_{2,0}(t) = \frac{(\lambda+\nu)^2}{\lambda^2} P_{0,1}(t) - \frac{\nu}{\lambda} P_{0,1}(t)$$

и так далее. Воспользовавшись условием нормировки можно записать

$$P_{0,0}(t) + P_{0,1}(t) + P_{1,1}(t) + P_{1,0}(t) + \ldots + P_{n,1}(t) = 1 . \qquad (11.15)$$

Поскольку все слагаемые кроме $P_{0,0}$ выражаются через вероятность $P_{0,1}$, то выражение (11.15) можно переписать в виде

$$P_{0,0}(t) + cP_{0,1}(t) = 1 \text{ или } P_{0,0}(t) = 1 - cP_{0,1}(t),$$

где c – некоторая константа, отражающая взаимосвязь вероятности $P_{0,1}$ с другими величинами.

Подставляя выражение для $P_{0,0}$ в последнее уравнение системы (11.14), получим

$$c\frac{dP_{0,1}(t)}{dt} = -\lambda P_{0,1}(t),$$

откуда $P_{0,1}(t) = \exp\left(-\frac{\lambda}{c}t\right)$.

Отсюда следует, что

$$P_{0,0}(t) = 1 - c\exp\left(-\frac{\lambda}{c}t\right).$$

Константа c довольно просто вычисляется, если известно количество запасных элементов. Так для одного запасного элемента

$$c = \frac{\lambda\mu + \lambda^2 + \lambda\nu + \mu\nu}{\lambda\mu},$$

для двух запасных элементов

$$c = \frac{(\nu+\lambda)^2}{\lambda^2} - \frac{\nu}{\lambda} + \frac{(\lambda+\nu)^2}{\mu\lambda} - \frac{\nu}{\mu} + \frac{\lambda+\nu}{\lambda} + \frac{\lambda+\nu}{\mu} + 1,$$

для трех запасных элементов

$$c = 1 + \frac{\lambda+\nu}{\mu} + \frac{\lambda+\nu}{\lambda} + \frac{(\lambda+\nu)^2}{\mu\lambda} - \frac{\nu}{\mu} + \frac{(\lambda+\nu)^2}{\lambda^2} - \frac{\nu}{\lambda} + \frac{(\lambda+\nu)^3}{\lambda^2\mu} - \frac{(\lambda+\nu)\nu}{\lambda\mu} -$$
$$- \frac{(\lambda+\nu)\nu}{\mu^2} + \frac{(\lambda+\nu)^3}{\lambda^3} - \frac{(\lambda+\nu)\nu}{\lambda^2} - \frac{(\lambda+\nu)\nu}{\mu\lambda}.$$

Таким образом, получено простое решение задачи расчета надежности объекта с ограниченным количеством запасных элементов, в случае возобновляемого ЗИП с заданными распределениями наработки до отказа, времени восстановления и ремонта. Рассмотрена стратегия функционирования системы, имеющей запасные элементы, в которой восстановление отказавшего элемента осуществляется путем его замены из числа запасных, отказавший элемент ремонтируется случайное время и после ремонта возвращается в ЗИП. Данная схема функционирования элемента максимально приближена к стратегии, имеющей место в практике функционирования систем, важных для безопасности атомных станций.

В заключение можно отметить, что предложенная стратегия является обобщением модели «размножения и гибели»; она более объективно отражает процесс функционирования объектов, так как учитывает пребывание системы в неработоспособном состоянии во время замены отказавшего элемента. Это обстоятельство является очень важным, когда речь идет об объектах повышенного риска.

Глава 12

ЧИСЛЕННЫЕ МЕТОДЫ В СИСТЕМНОМ АНАЛИЗЕ

12.1. Организация вычислительного процесса

При решении любой задачи системного анализа предполагается выполнение трех этапов: на первом этапе строится модель исследуемой системы, на втором – осуществляется формулирование цели исследования и постановка задачи и, наконец, на третьем – выполняется собственно решение поставленной математической задачи. Суть первых двух перечисленных этапов состоит в формализации объекта исследования и формализованном представлении цели исследования. В результате выполняемых действий разрабатываются модели: с одной стороны модель объекта исследования, с другой – модель целей исследования, выражением которой являются критерии и ограничения задачи. Третий этап заключается в решении сформулированных задач, получении числовых результатов и исследовании решений. Для грамотного проведения вычислительных процедур требуется глубокое знание математических методов, искусство организации вычислительных процедур, неформальное отношение к проводимым вычислениям, которое подразумевает привлечение исследовательского, творческого мышления. Важно осознавать, что главная цель расчетов не числа, а понимание сути исследуемых явлений и процессов. Иными словами за числами необходимо видеть существо проблемы.

Остановимся на некоторых вопросах, которые необходимо иметь в виду при выполнении третьего, заключительного этапа системных исследований. Одним из первых вопросов, на который требуется ответить, следующий: «Будут ли рассчитанные величины соответствовать требуемым результатам системного анализа?». В первую очередь следует заметить, что нельзя ожидать от заказчика работ по системному анализу точного ответа на вопрос, что он хочет получить в результате. Уже было отмечено ранее, что заказчик формулирует проблему в общем. С другой стороны вполне естественным является тот факт, что при проведении системных исследований на ряде стадий или даже в

целом, возможно не знать в точности, что ожидается получить в итоге. В некотором смысле, если достигается в точности ожидаемый результат, то это означает, что ничего нового о существе решаемой задачи получить не удалось. В процессе такого решения единственно чего добились – это повысили свою уверенность в поведении исследуемого процесса или явления. В самом деле, можно сказать: «Если исследователь знает, что он делает, то этого можно не делать». Таким образом, важно понимать, что исследователь хочет узнать в процессе выполняемых операций. Для этого работу надо специально планировать так, чтобы увеличить шансы заметить что-нибудь необычное. Если можно включить в процесс вычислений дополнительные побочные проверки исследуемой модели, то ради этого следует потратить немного машинного времени. Более того, необходимо обратить внимание на то, что нужно обдуманно выбирать данные, отображаемые в качестве выходных результатов. Вероятно, что кроме требуемого минимума надо вывести еще какой-то разумно выбранный набор чисел. Следует помнить, что многие великие открытия были сделаны в результате случайного наблюдения, важность которого понял подготовленный исследователь. Подводя итог сказанному, отметим еще раз, что, приступая к заключительному этапу системных исследований, необходимо ответить на вопрос: «Что мы собираемся делать с ответом?». Активность и воображение при ответе на данный вопрос могут дать многое для всего исследования, в то время как механическое проведение расчетов может помешать возникновению какого бы ни было понимания сути задачи, расходуя многие часы расчетного времени для получения очевидных числовых результатов.

Имеется также опасность допустить и другую ошибку – потребовать вывода слишком многих величин. Особенно такая ситуация характерна для исследования многопараметрических задач. Большое количество выводимых результатов может также скрыть понимание проблемы. В этом случае необходимо применять теорию планирования экспериментов, чтобы с ее помощью изменить постановку исследований и систематизировать обработку результатов.

Следующий вопрос, на который необходимо постараться дать ответ, это вопрос, касающийся всестороннего анализа исходной информации. Тщательный анализ изучаемой системы может дать о ней дополнительные сведения, использование которых может привести к уточнению модели или видоизменению постановки задачи. В первую очередь следует постараться предположить поведение исследуемой системы или ее частей в некоторых особых точках. Так, например, если при проведении исследования модели системы осуществляется вычисление некоторых показателей системы, то возможно предсказать значения этих показателей в моменты времени равные нулю и бесконечности. Учет такой информации может привести к уточнению модели или послужить для проверки правильности полученных результатов.

Иногда критический подход к анализу неизвестной ситуации может вызвать новые формулировки задачи, которые в свою очередь приведут к более глубокому пониманию исследуемых процессов. В процессе такого анализа может также быть обнаружено, что были сделаны излишне ограничительные предположения относительно модели и что их можно легко устранить. Во всяком случае, следует понять роль ограничений и включить в вычисления проверки, которые покажут ценность тех или иных предположений. Анализ входной информации, особых точек и специфических особенностей системы и ее частей может вызвать новые требования к содержательному оформлению выходной информации.

Только после того как проведен всесторонний и тщательный анализ исходной информации и продуманы требования к выходной информации, необходимо приступать к обдумыванию организации вычислительного процесса. На данном этапе в первую очередь необходимо выбрать метод решения поставленной задачи. Следует иметь в виду, что аналитическое решение часто гораздо лучше численного, а оценка ошибок может быть выполнена более точно.

Принятый план вычислений должен использовать как можно больше первоначальных данных. Математические приближения по формулам должны соответствовать характеру принятой модели. План вычислений должен включать как верификацию программ, так и проверку правильности итогового результата. Необходимо, чтобы была вычислена или получена из других источников некоторая избыточная информация, чтобы на ее основе можно было выполнить проверки результатов. И наконец, необходимо постараться объяснить любые полученные решения, верные или неверные, пусть это даже будет связано с затратами времени на выяснение факта их правильности.

Следует отметить еще одно обстоятельство, почти неизбежно в процессе вычислений появляется новая информация, которая может привести к необходимости внесения изменений в первоначальный план. Но прежде чем вносить изменения, требуется выяснить причины появления этой информации. Дает ли данное изменение что-то новое об используемой модели? Нет ли необходимости ввиду появления новых данных вновь подойти к вопросу о проверке построенной модели?

Изменения не должны вноситься поспешно, им следует посвятить столь же тщательное обсуждение, как и разработке первоначального

плана. Следует помнить, что если все идет как задумано, то ценность такого расчета не очень велика. Как раз из неожиданностей могут возникнуть новые идеи и решения. Таким образом, к ситуации, когда приходится вносить изменения в первоначальный план, следует относиться скорее как к счастливой возможности, чем как к неудаче. Естественно, что если такая ситуация возникла из-за недоработок на ранних этапах, то это будет лишним примером ценности предварительного обдумывания.

Всегда соблазнительно, взявшись за решение задачи, быстро внести мелкие изменения, не заботясь о последствиях и осложнениях, особенно если результат требуется получить к определенному сроку. И все-таки спешка в этот момент может свести на нет всю прежнюю тщательную работу.

Отметим еще раз, что цель расчетов не числа, а понимание, следовательно, специалист, который должен этого понимания достигнуть, обязан знать, как осуществляется процесс вычисления. Если он не понимает, что делается, то мало вероятно, чтобы он извлек из вычислений что-нибудь ценное. Он видит голые цифры, в то время как их истинное значение может оказаться скрытым в вычислениях. Результат, который получается в процессе вычислений, зависит от того, что поступает на вход модели, и от того, что с этими данными делают. Если не понимать промежуточные процессы, весьма легко перепутать эффекты использованной при вычислениях модели с эффектами, обусловленными всевозможными аппроксимациями, приближениями и т.п.

Часто процесс вычисления проливает свет на саму обрабатываемую модель. Вычисления являются средством получения числовых результатов, но они также представляют собой орудие разума для исследования мира. Если ставится задача понять суть происходящих явлений, автор модели должен следить за вычислениями. Это не означает выполнение всей мелкой работы, но если он не будет в достаточной степени понимать все, что делает машина, он вряд ли сумеет осмыслить даже правильно построенные вычисления.

Следует еще раз отметить, что объектом системных исследований являются сложные системы различной природы. Поэтому постановки задач системного анализа весьма сложны. Причем сложность обуславливается рядом факторов, это и большое количество параметров модели, и сложность самих моделей, многокритериальность задач при наличии многих ограничений и т.д. Ввиду этого при решении задач системных исследований большое применение находят численные методы. На рассмотрении некоторых наиболее важных и часто применяемых остановимся далее.

12.2. Метод последовательных приближений

Метод последовательных приближений применяется для решения уравнений или систем уравнений в случаях, когда искомые параметры не могут быть выражены в явном виде. В общем случае будем предполагать, что имеется некоторая функция $F(x)$ и необходимо найти такие значения аргумента x, для которых

$$F(x) = 0. \tag{12.1}$$

Функция $F(x)$ может иметь какой угодно вид, она может быть алгебраической или трансцендентной, единственное, что будем предполагать – это ее дифференцируемость.

В общем случае функции, которыми оперируют в задачах системных исследований, не имеют аналитических формул для своих корней. Поэтому приходится пользоваться приближенными методами нахождения корней, которые в основном состоят из двух этапов:

1) нахождение приближенного значения корня;

2) уточнение приближенного значения до некоторой заданной степени точности.

Приближенное значение корня уравнения $F(x) = 0$ часто бывает известно из физических соображений. Если это значение неизвестно, его можно найти с помощью грубого анализа функции. В качестве рекомендации можно предложить следующий метод. Определяются два такие значения x, для которых $F(x)$ имеет противоположные знаки, т.е. определяются такие x^* и x_*, для которых

$$F(x^*) > 0 \text{ и } F(x_*) < 0.$$

Тогда между x^* и x_* есть по крайней мере одна точка, где $F(x) = 0$. В качестве исходного приближения для нахождения корня $F(x)$ можно взять

$$x_0 = 0{,}5\,(x^* + x_*).$$

Рассмотрим теперь процедуры, относящиеся ко второму этапу – уточнению первоначального приближения. Численный метод, в котором производится последовательное уточнение первоначального грубого приближения, называется методом итераций. Каждый шаг в таком методе называется итерацией. Если при последовательных итерациях получаются значения, которые все ближе и ближе приближаются к истинному значению корня, то говорят, что метод итеративного решения сходится. Одним из методов получения решения с помощью итеративных процедур является метод последовательных приближений.

Метод последовательных приближений. Сформулируем постановку задачи. Имеется система уравнений вида (12.1). Необходимо определить вектор параметров $\mathbf{X} = (x_1, x_2, \ldots, x_k)$, при котором функция (12.1) достигает максимума.

Вначале рассмотрим применение метода последовательных приближений для случая, когда требуется определить один параметр. Преобразуем уравнение (12.1) к следующему виду: $G(x) = x$. Это преобразование можно получить, прибавив к правой и левой частям уравнения (12.1) искомую величину x, т.е.

$$F(x) + x = x. \tag{12.2}$$

Пусть x_0 будет исходным приближенным значением корня уравнения (12.2). Тогда в качестве следующего приближения принимается значение

$$x_1 = F(x_0) + x_0.$$

На втором шаге в качестве приближения возьмем

$$x_2 = F(x_1) + x_1.$$

Продолжая этот процесс дальше, в качестве n-го приближения принимаем значение

$$x_n = F(x_{n-1}) + x_{n-1}.$$

Процедура повторяется до тех пор, пока не будет достигнута заданная точность решения уравнения. Правило остановки можно задать следующим образом: вычислительный процесс заканчивается, когда выполняется соотношение $|x_n - x_{n-1}| \le \varepsilon$, где ε – заданная точность вычисления.

Представляет интерес рассмотрение метода последовательных приближений для двухпараметрического случая. Случай оценки двух параметров имеет прикладное значение в статистических задачах, когда исследуемые процессы описываются законом распределения с неизвестными параметрами. Для определения параметров приходится решать, например, уравнение правдоподобия. Часто систему уравнений максимального правдоподобия трудно решить в явном виде. Даже для экспоненциальных семейств система уравнений правдоподобия может быть нелинейна и трудна для решения, как это происходит при оценке параметров распределения Вейбулла. Двухпараметрические распределения, например нормальное, Вейбулла, гамма-распределение, находят широкое применение в теории надежности.

При оценивании двух параметров закона распределения необходимо решать систему

$$\begin{cases} x_1 = F(x_1, x_2) + x_1; \\ x_2 = F(x_1, x_2) + x_2. \end{cases}$$

Задавая вектор начального приближения (x_1^0, x_2^0), находим первые приближенные оценки:

$$x_1^1 = F(x_1^0, x_2^0) + x_1^0;$$
$$x_2^1 = F(x_1^0, x_2^0) + x_2^0.$$

Повторяем эту процедуру до тех пор, пока разность ($\left|x_1^n - x_1^{n-1}\right|, \left|x_2^n - x_2^{n-1}\right|$) не попадает в ε-окрестность. Иными словами, должны совместно выполняться соотношения:

$$\left|x_1^n - x_1^{n-1}\right| \le \varepsilon,$$

$$\left|x_2^n - x_2^{n-1}\right| \le \varepsilon.$$

Значения (x_1^n, x_2^n) принимаются за искомую оценку вектора вычисляемых параметров.

Усовершенствованный метод последовательных приближений. Этот метод разрабатывался в целях обеспечения более быстрой сходимости оценок. Более быстрая сходимость по сравнению с обычным методом последовательных приближений достигается за счет того, что при каждой итерации делается большая поправка к очередному значению параметра x_i. Иначе говоря, вместо того, чтобы полагать

$$x_n = F(x_{n-1}) + x_{n-1}$$

применяют следующую формулу:

$$x_n = \alpha\, F(x_{n-1}) + x_{n-1}\text{, где } \alpha > 1.$$

В [8] доказано, что наилучшая сходимость достигается в случае, когда параметр α вычисляется следующим образом:

$$\alpha = \frac{1}{1 - (F(\xi) + \xi)'_\xi},$$

где $x_n \le \xi \le a$, a – корень уравнения (12.1).

Значение ξ остается неизвестным, но для вычисления производной используется следующее приближение:

$$(F(\xi) + \xi)_\xi' = \frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}}.$$

Формула итеративного метода приобретает в этом случае следующий вид:

$$x_n = \frac{1}{1 - \dfrac{F(x_{n-1}) - F(x_{n-2})}{x_{n-1} - x_{n-2}}} F(x_{n-1}) + x_{n-1}.$$

Метод Ньютона–Рафсона. Метод основан на разложении функции $F(x)$ в окрестности a

$$F(x) = F(x_1) + (a - x_1)\, F'(x_1) = 0.$$

Произведем элементарные преобразования в данном уравнении, получим

$$a = x - \frac{F(x)}{F'(x)}.$$

Заменим искомый параметр его первым приближением, получим

$$x_1 = x_0 - \frac{F(x_0)}{F'(x_0)}.$$

Далее организуем итеративный процесс

$$x_i = x_{i-1} - \frac{F(x_{i-1})}{F'(x_{i-1})}. \qquad (12.3)$$

Если начальное приближение x_0 выбрано близко к корню уравнения (12.1) и если производная $F'(x_i)$ для i=1, 2,… не равна нулю, то последовательность, порожденная соотношением (12.3) сходится к a.

12.3. Численное интегрирование

Задачи, в которых требуется вычислить интегралы, возникают на разных этапах решения задач системного анализа. Иногда удается найти аналитическую формулу, т.е. выразить неопределенный интеграл в виде комбинаций алгебраических и трансцендентных функций, после чего остается вычислить значение определенного интеграла, подставив в формулу пределы интегрирования.

В большинстве случаев не удается найти никакой аналитической формулы или же она получается настолько сложной, что вычислять интеграл с ее помощью труднее, чем другими способами. В таких ситуациях приходится применять различные методы численного интегрирования, которые основаны на том, что интеграл представляется в виде предела суммы площадей. Далее вычисляют эту сумму с достаточно высокой степенью точности.

Рассмотрим постановку задачи. Пусть необходимо вычислить определенный интеграл

$$I = \int_a^b f(x)dx$$

при условии, что a и b конечны и $f(x)$ является непрерывной функцией x во всем интервале $a \le x \le b$. Представляют интерес также те случаи, когда один или оба предела интегрирования бесконечны, либо когда подынтегральная функция имеет особенности внутри интервала интегрирования или на его концах.

Общий подход к решению задач численного интегрирования будет следующим. Определенный интеграл I представляет собой площадь, ограниченную кривой $f(x)$, осью x и прямыми $x = a$ и $x = b$. Задача заключается в вычислении интеграла. При этом интервал интегрирования разбивается на множество более мелких подынтервалов, приближенно вычисляется площадь каждой полоски, получающейся при таком разбиении. Далее площади полосок суммируются.

Существует множество методов вычисления определенных интегралов, основанных на таком разбиении и последующем суммировании. Рассмотрим некоторые из них.

Метод прямоугольников. Пусть интервал $[a, b]$ разбит на множество подынтервалов $[x_i, x_{i+1})$. Будем считать, что на рассматриваемом подынтервале интегрируемая функция почти константа: $f(x) \approx \text{const}$. Тогда для данного подынтервала можно положить: $I_i \approx (x_{i+1} - x_i) f(\zeta)$, где ζ – произвольная точка на рассматриваемом подынтервале. Если в качестве такой точки взять среднюю точку подынтервала, получим формулу

$$I_i \approx (x_{i+1} - x_i) f\left(\frac{x_{i+1} - x_i}{2}\right).$$

Далее, производя суммирование по всем подынтервалам, получим формулу интегрирования по методу прямоугольников

$$I \approx \sum_{i=0}^{n-1} (x_{i+1} - x_i) f\left(\frac{x_{i+1} - x_i}{2}\right), \tag{12.4}$$

где n – количество подынтервалов на отрезке интегрирования.

Метод трапеций. В отличие от метода прямоугольников, где предполагалось, что интегрируемая функция на каждом подынтервале близка к константе, в методе трапеций принимается допущение, что функция на каждом подынтервале может быть приближена линейной функцией. При таком предположении интеграл заменяется площадью трапеции с высотой $(x_{i+1} - x_i)$ и основаниями $f(x_{i+1})$ и $f(x_i)$. В результате получим формулу трапеций

$$I \approx \sum_{i=0}^{n-1} (x_{i+1} - x_i) \frac{f(x_{i+1}) + f(x_i)}{2}. \tag{12.5}$$

На рис. 12.1 приведена иллюстрация рассмотренных методов интегрирования путем замены определенного интеграла конечной суммой, а именно показано увеличенное представление элемента площади, ограниченной функцией $f(x)$, осью x и прямыми $x = x_i$ и $x = x_{i+1}$. Пунктирной линией выделена трапеция, которой заменяется площадь под кривой интегрирования. В центре отрезка тонкой прямой линией отмечена высота прямоугольника, которая используется в качестве сомножителя в методе прямоугольников.

Ошибка интегрирования методом трапеций. При интегрировании с использованием формулы (12.5) возникает ошибка, равная сумме площадей между кривой $y = f(x)$ и хордами, соединяющими точки $y_i = f(x_i)$ и $y_{i+1} = f(x_{i+1})$. Оценим ошибку, разлагая функцию $y = f(x)$ в ряд Тейлора в точках x_i и x_{i+1}. Это разложение позволит получить уравнение исходной кривой в виде, удобном для сравнения точного значения интеграла с приближенным, вычисленным по формуле (12.5).

Рассмотрим разложение функции $y = f(x)$ в ряд Тейлора в окрестности точки $x = x_i$. Предположим, что интегрируемая функция имеет

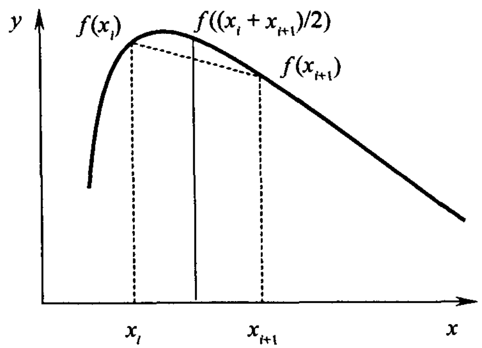

Рис. 12.1. Иллюстрация численного интегрирования методом трапеций

столько производных, сколько может потребоваться.

$$f(x)=f(x_i)+(x-x_i)f'(x_i)+\frac{(x-x_i)^2}{2}f''(x_i)+\ldots. \qquad (12.6)$$

Аналогично разложим данную функцию в ряд в окрестности точки x_{i+1}, получим

$$f(x)=f(x_{i+1})+(x-x_{i+1})f'(x_{i+1})+\frac{(x-x_{i+1})^2}{2}f''(x_{i+1})+\ldots. \qquad (12.7)$$

Обозначим длину подынтервала через h, т.е. $h=x_{i+1}-x_i$ и перепишем формулу (12.7)

$$f(x)=f(x_i)+(x-x_i-h)f'(x_{i+1})+\frac{(x-x_i-h)^2}{2}f''(x_{i+1})+\ldots. \qquad (12.8)$$

Найдем среднее из обеих представлений (12.6) и (12.8)

$$f(x)=\frac{f(x_i)+f(x_{i+1})}{2}+(x-x_i)\frac{f'(x_i)+f'(x_{i+1})}{2}-\frac{h}{2}f'(x_{i+1})+$$
$$+\frac{(x-x_i)^2}{4}(f''(x_i)+f''(x_{i+1}))-\frac{(x-x_i)h}{2}f''(x_{i+1})+\frac{h^2}{4}f''(x_{i+1})+\ldots.$$

Интегрируя $f(x)dx$ в пределах от x_i до x_{i+1}, получаем

$$\int_{x_i}^{x_{i+1}} f(x)dx=\frac{f(x_i)+f(x_{i+1})}{2}h-\frac{f'(x_{i+1})-f'(x_i)}{4}h^2-(f''(x_i)+f''(x_{i+1}))\frac{h^3}{12}+\ldots \qquad (12.9)$$

Это выражение представляет собой оценку значения интеграла. Оценка может быть сделана как угодно точной, потому что можно взять сколько угодно большое число членов в разложении функции в ряд Тейлора. Правило трапеций получается, если в формуле (12.9) отбросить все члены, содержащие h в степенях выше первой. Таким образом, ошибка при использовании метода трапеций равна

$$E=\frac{f'(x_{i+1})-f'(x_i)}{4}h^2-(f''(x_i)+f''(x_{i+1}))\frac{h^3}{12}+\ldots \qquad (12.10)$$

Для малых h первый член гораздо больше всех остальных, поэтому ошибка именно им и определяется. Более того, в [8] показано, что за счет наличия в выражении слагаемых с производными более высокого порядка ошибка определяется следующим образом

$$E\approx-\frac{f'(x_{i+1})-f'(x_i)}{12}h^2.$$

Полную ошибку интегрирования методом трапеций можно оценить следующим соотношением

$$e=\sum_{i=0}^{n-1}E_i\approx-\frac{f'(x_b)-f'(x_a)}{12}h^2. \qquad (12.11)$$

Усовершенствованный метод трапеций. Попытаемся найти более точное значение интеграла. Для этого произведем сравнительно простое усовершенствование метода трапеций. Рассмотрим ошибку (12.11) и преобразуем данную формулу. На основании теоремы о среднем значении можно написать

$$f'(x_b)-f'(x_a)=(b-a)f''(\xi),$$

где $a\le\xi\le b$, так что

$$e\approx-\frac{h^2}{12}(b-a)f''(\xi).$$

В данном выражении обозначим

$$C=-\frac{1}{12}(b-a)f''(\xi),$$

тогда ошибку можно записать

$$e=Ch^2.$$

Предположим теперь, что вторая производная от интегрируемой функции является константой, тогда и C будет константой.

Выберем другую величину шага разбиения отрезка интегрирования $k=(b-a)/m$, причем $m\neq n$. Тогда получим выражение для ошибки интегрирования в виде

$$e=Ck^2.$$

Запишем далее значение интеграла, вычисленное по правилу трапеций с шагом h – I_h и с шагом k – I_k. Тогда будут справедливы следующие выражения

$$I=I_h+Ch^2, \qquad (12.12)$$
$$I=I_k+Ck^2.$$

Приравняем эти два уравнения друг другу и решим их относительно C, получим

$$C = \frac{I_h - I_k}{k^2 - h^2} \ .$$

Подставляя это выражение в (12.12), получим

$$I = I_h + \frac{I_h - I_k}{k^2 / h^2 - 1} \ . \qquad (12.13)$$

Вычисленное таким образом значение интеграла является лучшим приближением, чем любое представление, полученное в (12.12). Если же вторая производная интегрируемой функции действительно постоянна на интервале интегрирования $[a, b]$, то ошибка в формуле (12.13) равна нулю.

Правило Симпсона. Способ интегрирования по правилу Симпсона наиболее широко известный и применяемый метод численного интегрирования. В данном методе, так же как и в рассмотренных ранее, интегрирование производится путем разбиения общего интервала интегрирования на множество более мелких отрезков. Однако в этом методе для вычисления площади над каждым из отрезков через три последовательных ординаты разбиения проводится квадратичная парабола. Рассмотрим процедуру получения формулы Симпсона. Пусть, как и ранее, n – количество отрезков разбиения интервала интегрирования; h – ширина интервала разбиения. Предположим, что число n является четным. Пусть k – ширина интервала при другом разбиении, такая, что $k = 2h$. Тогда можно записать

$$I_h = h/2(f(x_0) + 2f(x_1) + 2f(x_2) + \ldots + 2f(x_{n-1}) + f(x_n)) \ ; \qquad (12.14)$$

$$I_k = h(f(x_0) + 2f(x_2) + \ldots + 2f(x_{n-2}) + f(x_n)) \ . \qquad (12.15)$$

Преобразуем выражение (12.13) учитывая, что $k = 2h$, получим

$$I = I_h + \frac{I_h - I_k}{k^2 / h^2 - 1} = I_h + \frac{I_h - I_k}{h^2 / 4h^2 - 1} = I_h - \frac{4(I_h - I_k)}{3} \ .$$

Приведем подобные члены

$$I = \frac{4I_k - I_h}{3} .$$

Подставим данное выражение формулы (12.14) и (12.15), окончательно получим

$$I_h = (f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots +$$
$$+4f(x_{n-2}) + 2f(x_{n-1}) + f(x_n))h/3.$$

Данная формула называется формулой Симпсона. Ошибка при интегрировании с помощью формулы Симпсона равна

$$e \approx -\frac{h^4}{180}(b-a) f^{\mathrm{IV}}(\xi).$$

Как видно из данного выражения ошибка пропорциональна значению h^4, в то время как для метода трапеций она была пропорциональна h^2. Это означает, что формула Симпсона соответствует ряду Тейлора с точностью до членов третьего порядка, а метод трапеций соответствует этому ряду только с точностью до членов первого порядка. Поэтому при интегрировании многочленов степени не выше третьего порядка метод Симпсона дает точные значения интеграла.

Вычисление интегралов с бесконечными пределами. Рассмотрим еще одну важную задачу численного интегрирования – вычисление определенного интеграла для случая, когда один или оба предела интегрирования равны бесконечности. Таким образом, предметом рассмотрения будет вычисление одного из следующих выражений

$$\int_{-\infty}^{\infty} f(x)dx, \quad \int_{a}^{\infty} f(x)dx, \quad \int_{-\infty}^{b} f(x)dx.$$

Поскольку вычислительная техника не работает с такими понятиями как бесконечность, то задача аналитика состоит в том, чтобы заменить бесконечность реальным числом, обеспечив при этом приемлемую точность вычисления. Иными словами, необходимо найти такие числа A и B, для которых были бы справедливы выражения

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-A}^{B} f(x)dx + \varepsilon; \qquad (12.16)$$

$$\int_{a}^{\infty} f(x)dx = \int_{a}^{B} f(x)dx + \varepsilon_1; \qquad (12.17)$$

$$\int_{-\infty}^{b} f(x)dx = \int_{-A}^{b} f(x)dx + \varepsilon_2. \qquad (12.18)$$

Для того чтобы процедура вычисления была реализуема, необходимо, чтобы интегрируемая функция была ограничена и убывала на границах интегрирования. Изложение способа замены бесконечного предела конечным числом покажем на примере одного из пределов. Для других случаев процедура вычисления будет аналогичной. На рис. 12.2. видно, что функция стремится к нулю при стремлении аргумента к бес-

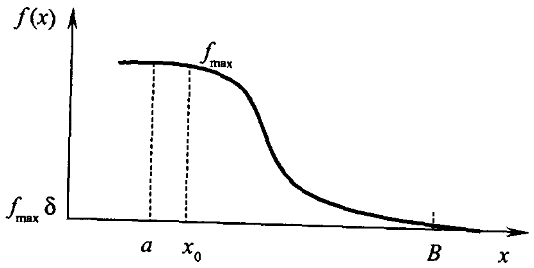

Рис. 12.2. Замена бесконечности в пределе интегрирования конечным числом

конечности. Из этого следует, что можно задаться некоторым значением ошибки и определить число B, которое будет удовлетворять уравнению (12.17).

Для определения числа B на первом этапе необходимо попытаться оценить максимальное значение интегрируемой функции, которое она принимает на интервале интегрирования. Пусть это значение равно f_{max}. Предположим, что данное значение функция приобретает в точке x_0. Далее зададимся числом δ, которое будет определять точность проведения вычислений, обычно это малое число. Например, $\delta = 10^{-4} - 10^{-6}$. Определим значение функции, ограничивающее процесс вычисления $f_{max}\delta$. Значениями функции, оказавшимися меньше данного значения, пренебрежем. На следующем этапе необходимо найти такое минимальное значение аргумента x^*, для которого будет выполняться условие

$$f(x^*) \le f_{max}\delta . \tag{12.19}$$

Для этого необходимо, задаваясь некоторым шагом Δx, двигаться в сторону уменьшения интегрируемой функции (в случае, представленном на рис.12.2 необходимо двигаться вправо), вычисляя в каждой точке значения интегрируемой функции и сравнивая результат вычисления $f(x_i)$ со значением, ограничивающим процесс интегрирования $f_{max}\delta$. Первое значение аргумента, для которого будет выполнено соотношение (12.19), принимается в качестве верхнего предела интегрирования условно заменяющего бесконечность, т.е. $x^*=B$. Определение нижнего предела интегрирования осуществляется аналогично, с той лишь разницей, что движение осуществляется влево от точки максимума функции. Если интегрируемая функция отрицательна, определяют ее минимум и осуществляют движение в сторону ее увеличения, соответственно вправо или влево, в зависимости от того, какой предел подлежит определению.

12.4. Методы поиска оптимального значения функции

С задачами оптимизации функции одной переменной впервые сталкиваются при изучении математического анализа. Решают такие задачи методами дифференциального исчисления. На первый взгляд может показаться, что эти задачи относятся к достаточно простым и методы их решения достаточно хорошо разработаны и изучены. Однако это не совсем так. Методы дифференциального исчисления находят ограниченное применение. Особенно актуальной проблема становится при организации вычислительного процесса на ЭВМ. В данном случае требуется разработка вычислительных процедур, реализующих численные методы поиска экстремального значения функции.

Рассмотрим постановку задачи. Пусть на числовой оси задано некоторое множество X, $F(x)$ – функция, определенная на множестве X и принимающая во всех точках $x \in X$ конечные значения. Для определенности будем рассматривать задачу минимизации функции $F(x)$ на множестве X. Перейдем к рассмотрению численных методов решения задачи поиска оптимального значения.

Метод деления отрезка пополам. Простейшим методом минимизации функции одной переменной, не требующим вычисления производной, является метод деления отрезка пополам. Рассмотрим данный метод. Будем предполагать, что минимизируемая функция $F(x)$ унимодальна на рассматриваемом отрезке $[a, b]$. Поиск минимума на $[a, b]$ начинается с выбора двух точек $x_1 = (a + b - \delta)/2$ и $x_2 = (a + b + \delta)/2$, где δ – постоянная, являющаяся параметром метода, $0 < \delta < b - a$. Величина δ является параметром, определяющим точность решения. Точки x_1 и x_2 располагаются симметрично на отрезке $[a, b]$ относительно его середины и при малых δ делят его почти пополам. После выбора точек x_1 и x_2 вычисляют значения $F(x_1)$ и $F(x_2)$ и сравнивают их между собой. Если $F(x_1) \le F(x_2)$, то полагают $a_1 = a$, $b_1 = x_2$; если $F(x_1) > F(x_2)$, то $a_1 = x_1$, $b_1 = b$. Поскольку $F(x)$ унимодальна на $[a, b]$, ясно, что точка минимума попадет на отрезок $[a_1, b_1]$. Длина этого отрезка будет равна

$$b_1 - a_1 = (b - a - \delta)/2 + \delta .$$

Данная процедура повторяется до тех пор, пока не будет достигнута заданная точность решения. При этом на k-м шаге точки для нового разбиения интервала выбираются по правилу $x_{2k-1} = (a_{k-1} + b_{k-1} - \delta)/2$, $x_{2k} = (a_{k-1} + b_{k-1} + \delta)/2$, $k \ge 2$. Далее вычисляются значения функции в этих точках $F(x_{2k-1})$ и $F(x_{2k})$. Затем определяются точки нового интервала. Если $F(x_{2k-1}) \le F(x_{2k})$, то полагаем $a_k = a_{k-1}$, $b_k = x_{2k}$, если же

$F(x_{2k-1}) > F(x_{2k})$, то полагаем $a_k = x_{2k-1}$, $b_k = b_{k-1}$. Полученный отрезок сравнивается с параметром, задающим точность вычисления $b_k - a_k < \varepsilon$. Если данное неравенство удовлетворяется, процесс вычисления останавливают. Следует отметить, что $\varepsilon > \delta$; можно привести также следующие соотношения: количество разбиений определяется числом $k > \log_2((b - a - \delta)/(\varepsilon - \delta))$. Поскольку каждое деление пополам требует двух вычислений значений функции, то для достижения заданной точности требуется $n = 2k > 2\log_2((b - a - \delta)/(\varepsilon - \delta))$ таких вычислений.

После определения отрезка $[a_k, b_k]$ в качестве приближения к значению аргумента, в котором функция принимает минимальное значение, можно взять точку $x_n = x_{2k-1}$ при $F(x_{2k-1}) \le F(x_{2k})$ и $x_n = x_{2k}$ при $F(x_{2k-1}) > F(x_{2k})$, а значение $F(x_n)$ может служить приближением для искомого минимального значения. Точность приближения в данном случае может быть оценена выражением $(b - a)2^{-n/2-1}$. Однако существуют методы, которые за такое же количество итераций позволяют получить решение с большей точностью. Одним из таких методов является метод золотого сечения.

Метод золотого сечения. Данный метод столь же прост как и метод деления отрезка пополам, но он позволяет решить задачу с требуемой точностью при меньшем количестве вычислений значений функции. Золотым сечением отрезка является деление на две неравные части так, чтобы отношение длины всего отрезка к длине большей части равнялось отношению длины большей части к длине меньшей части отрезка.

Золотое сечение отрезка $[a, b]$ производится двумя точками

$$x_1 = a + (3 - \sqrt{5})(b - a)/2$$

и

$$x_2 = a + (\sqrt{5} - 1)(b - a)/2,$$

расположенными симметрично относительно середины отрезка, причем

$$a < x_1 < x_2 < b.$$

Следует обратить внимание на одно обстоятельство, а именно то, что точка x_1, в свою очередь производит золотое сечение отрезка $[a, x_2]$, аналогично точка x_2 производит золотое сечение отрезка $[x_1, b]$. Используя данное свойство золотого сечения, разработан следующий метод минимизации унимодальной функции на отрезке $[a, b]$. На первом этапе производят золотое сечение указанного отрезка и вычисляют значения минимизируемой функции в точках x_1, x_2: $F(x_1)$ и $F(x_2)$. Далее, если $F(x_1) \le F(x_2)$, то за новый отрезок принимается $[a, x_2]$, т.е. $a_1 = a$, $b_1 = x_2$, если $F(x_1) > F(x_2)$, то $a_1 = x_1$, $b_1 = b$. Затем процесс деления отрезка методом золотого сечения продолжается и вычисляются значения минимизируемой функции в каждой новой точке. Данная процедура выполняется до тех пор, пока не получим на некотором n-м шаге отрезок $b_n - a_n < \varepsilon$, где ε – заданная точность. В качестве решения выбирают значение функции, удовлетворяющее условию

$$F(x^*) = \min(F(a_n), F(b_n)).$$

На этом процесс минимизации унимодальной функции заканчивается.

Метод прямого поиска. Рассмотрим еще один метод определения оптимального значения унимодальной функции на отрезке $[a, b]$ и нахождения значения аргумента, соответствующего оптимальному значению. В отличие от предыдущих случаев рассмотрим поиск максимального значения. Суть метода заключается в следующем. На первом шаге необходимо определить направление поиска оптимального значения функции. Для этого произвольным образом выбираем две точки x_1 и x_2 (рис. 12.3), в этих точках вычислим значения оптимизируемой функции $F(x_1)$ и $F(x_2)$. Затем сравним вычисленные значения друг с другом. Если $F(x_1) < F(x_2)$ и при этом $x_1 > x_2$, то двигаться в поиске максимума необходимо влево, если $F(x_1) > F(x_2)$ при том же соотношении между аргументами, то двигаться необходимо вправо. Положим, что $x_1 - x_2 = \Delta x$. Пусть $F(x_1) < F(x_2)$ так, как показано на рис. 12.3. Тогда максимум функции находится слева от точки x_2. Находим точку x_3 следующим образом $x_3 = x_2 - \Delta x$ и вычисляем значение функции в данной точке $F(x_3)$. Сравним значения функций между собой. Если выполняется неравенство $F(x_2) < F(x_3)$, то продолжаем двигаться в том же направлении. Пусть путем описанных вычислений достигнута точ-

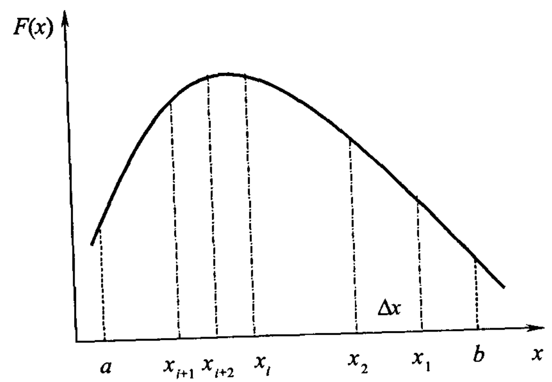

Рис. 12.3. Метод прямого поиска оптимального значения функции

ка x_{i+1}. Вычисляем значение функции в данной точке и проводим сравнение со значением в предыдущей точке, получаем $F(x_i) > F(x_{i+1})$. Результат сравнения показывает, что в неравенстве знак поменялся на обратный. Это говорит о том, что в процессе вычислений значение максимума функции осталось справа, т.е. его проскочили. В этом случае необходимо изменить величину шага и поменять направление движения. Положим $\Delta x = -\Delta x/2$ и определим новую точку $x_{i+2} = x_{i+1} + \Delta x$. Поскольку новое значение шага берется с обратным знаком, движение теперь осуществляется в другую сторону, в данном случае вправо. Теперь движение в этом направлении продолжаем до тех пор, пока выполняется соотношение $F(x_j) < F(x_{j+1})$. Как только в данном неравенстве знак меняется на обратный, необходимо опять уменьшить шаг и изменить его знак $\Delta x = -\Delta x/2$. Путем таких итераций процесс будет все ближе и ближе сходиться к точке, в которой функция имеет максимальное значение. Условием остановки процесса вычислений будет достижение заданной точности, которое можно сформулировать путем задания соотношения $|x_{k+1} - x_k| \le \varepsilon$. Полученное значение x_{k+1} принимается за аргумент, в котором функция достигает максимума, а значение функции в данной точке за оценку максимального значения.

12.5. Методы прямого поиска решений уравнений

Рассмотрим задачу вычисления корней уравнения в случае, когда исследуемая функция в общем случае имеет трансцендентный вид. Относительно вида функции сделаем одно предположение: будем считать, что функция монотонна на рассматриваемом участке. Пусть имеем уравнение

$$F(x) = 0. \tag{12.20}$$

Такая запись имеет достаточно общий вид, например, если имеется уравнение в виде $f(x) = g(x)$, то перенося правую часть влево, приходим к выражению (12.20). Один из методов решения уравнений вида (12.20) – метод последовательных приближений – был рассмотрен в п. 12.2. В методе последовательных приближений делалось предположение относительно дифференцируемости функции. В излагаемом методе такого предположения делать не будем.

Итак, изложим метод поиска решения уравнения, аналогичного методу поиска оптимального значения функции, рассмотренного в предыдущем параграфе. Пояснение метода поиска решения показано на рис. 12.4.

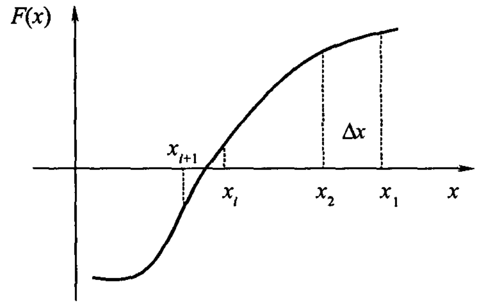

Рис. 12.4. Метод прямого поиска решений уравнений

На первом шаге произвольно выбираем точку x_1 и вычисляем значение функции $F(x_1)$. Вначале определяем квадрант, в который попало вычисленное значение функции. Пусть получили, как изображено на рис. 12.4, $x_1 > 0$, значение функции также больше нуля. Далее выбираем шаг, с которым будем осуществлять поиск решения, и вычисляем точку x_2: $x_2 = x_1 - \Delta x$. В этой точке также вычисляем значение функции $F(x_2)$. Далее определяем направление поиска решения. Если значение функции в точке x_2 больше чем в точке x_1, то пересечение функции с осью ординат будет находиться справа от первоначально выбранной точки. Соответственно двигаться в поиске решения необходимо вправо. Если же значение функции в точке x_2 меньше чем в точке x_1, то пересечение функции с осью ординат будет находиться слева от первоначально выбранной точки. И, следовательно, двигаться в поиске решения необходимо влево, так, как это изображено на рис. 12.4. Определив направление поиска, осуществляем движение с шагом Δx до тех пор, пока значение функции не станет меньше нуля, т.е. пока функция не изменит свой знак. Как только произошло изменение знака функции на противоположный (точка x_{i+1}) необходимо изменить направление движения, при этом необходимо также изменить шаг, с которым осуществляется поиск решения: $\Delta x = -\Delta x/2$. Теперь начинаем двигаться в противоположную сторону с меньшим шагом. Движение в эту сторону также идет до тех пор, пока значение функции не изменило свой знак. Как только произошло изменение знака функции, необходимо снова изменить направление поиска решения и уменьшить шаг. Ясно, что в процессе такого поиска решение все ближе и ближе сходится к нулю. Условием остановки процесса вычислений будет достижение заданной точности, которое можно сформулировать путем задания соотношения $|x_{k+1} - x_k| \le \varepsilon$. Осуществляя данную процедуру, находим решение уравнения (12.20).

Подведем итоги. В данном разделе представлены некоторые типы задач и изложены основные подходы к их решению. Рассмотренные методы не претендуют на полноту охвата всей теории численных методов. Реально теория численных методов является самостоятельной развивающейся дисциплиной, с широким спектром разделов. Например, имеются хорошо разработанные и представленные в литературе методы решения дифференциальных уравнений, интегральных уравнений Вольтерра и Фредгольма, всевозможных систем уравнений и т.п. разделы. Данные процедуры представлены в специальной литературе. Здесь ставилась цель изложить подходы к решению задач и отдельных процедур, широко встречающихся в системных исследованиях.

В заключение остановимся на некоторых рекомендациях по применению численных методов. Приступая к решению задачи системных исследований целесообразнее начинать решение с расчета простейших моделей, изучать модели с использованием простейших проверенных методов. Лишь после всестороннего анализа и уяснения всех аспектов задачи имеет смысл заниматься расчетом сложной модели с применением громоздких по своей структуре процедур.

Обратим внимание на некоторые опасности, с которыми может столкнуться системный аналитик при применении численных методов для решения задач. При алгоритмизации математической задачи с целью ее решения численными методами могут возникнуть ситуации, препятствующие успешному получению результата, а именно:

а) имеется опасность замены исходной задачи задачей, не имеющей отношения к исходной постановке;

б) алгоритм, применяемый для решения задачи, при возможностях современной вычислительной техники не поддается расчету.

Таким образом, необходимо тщательно и всесторонне анализировать постановку задачи. Исследование постановки задачи дает возможность начать изучение ее с простейшей модели с последующим усложнением метода и программы решения задачи. Важно отметить, что при решении задач численными методами большое внимание следует уделять разработке и применению специальных тестов и отладочных программ. Применение процедур контроля и тестовых проверок дает исследователю уверенность в успешном решении задачи.

Глава 13

ВЫБОР ИЛИ ПРИНЯТИЕ РЕШЕНИЙ

13.1. Характеристика задач принятия решений

Теория принятия решений представляет собой набор понятий и систематических методов, позволяющих всесторонне анализировать проблемы выбора и принятия решений, в том числе в условиях неопределенности. Цель теории – совершенствование процесса принятия решений.

Реализация выбора и принятия решений является одной из завершающих процедур системного анализа. За данной процедурой следует внедрение результатов системных исследований. В процессе принятия решения реализуется цель системного анализа, которая была сформулирована на соответствующих этапах проведения исследований: «Выбор является действием, придающим всей деятельности целенаправленность. Именно выбор реализует подчиненность всей деятельности определенной цели или совокупности целей» [1]. Таким образом, принятие решений является обязательным, центральным этапом системных исследований. Перед реализацией процесса выбора считаются выполненными два крайне важных этапа или процедуры системных исследований – определение целей системного анализа, для достижения которых собственно и осуществляется выбор, и генерирование альтернатив, т.е. порождение множества альтернатив, на котором предстоит осуществить выбор. В этом случае принятие решений можно рассматривать как операцию над множеством альтернатив, в результате выполнения которой множество сужается до подмножества выбранных альтернатив. В идеальной ситуации подмножество должно состоять из одной альтернативы. Эта ситуация желательна, но не всегда реализуема. Итак, при выполнении процесса принятия решений осуществляется анализ альтернативных способов действия, приводящих к достижению заданных целей. Важным обстоятельством является также наличие критериев, с помощью которых осуществляется сравнение вариантов достижения целей. Задание критериев или мер эффективности позволяет определить степень, с которой заданные цели могут быть достигнуты с помощью различных способов действия. Для каждого способа действия возможные исходы описываются в единицах принятых мер эффективности.

Таким образом, для того чтобы приступить к решению задач выбора, необходимо сформировать и детально проанализировать преследуемые системными аналитиками цели, задать исходное множество альтернатив, из которых требуется выбрать наиболее предпочтительные, а также определить критерии оценки и сравнения любых альтернатив. Но даже при выполнении всех перечисленных условий проблема выбора не тривиальна и допускает существенно различающиеся подходы к своему решению. Сложность решения задач выбора обусловлена их особенностями. Рассмотрим их.

1. Многоцелевой характер задач системного анализа. При проведении системных исследований приходится стремиться к достижению различных частных целей. В общем случае цели могут быть противоречивы. Продвижение по пути достижения одной цели может приводить к ухудшению результатов по другим. В такой ситуации лицо, принимающее решение, неизбежно оказывается перед необходимостью выбора между противоречивыми целями.

2. Воздействие фактора времени. Все важные последствия принятого решения сразу не проявляются и нельзя указать конкретный момент времени, когда можно наблюдать то или иное последствие.

3. Наличие неформализуемых понятий. Такие понятия как добрая воля, престиж, политические действия и тому подобные часто имеют место при проведении системных исследований, в особенности при анализе социотехнических систем. Эти понятия являются неформализуемыми понятиями, которые существенно усложняют задачу принятия решений.

4. Неопределенность. Маловероятно, что в момент принятия решения, т.е. выбора альтернативного действия, будут досконально известны последствия каждой из альтернатив.

5. Возможность получения информации. Часто можно организовать работу по сбору информации об объекте исследования, которая может помочь в решении задачи выбора альтернативного способа действия. Однако следует учитывать, что собираемая информация всегда требует затрат средств и времени и к тому же она практически всегда обладает ограниченной достоверностью.

6. Динамические аспекты процесса принятия решений. После того, как некоторое решение выработано и выбрана альтернатива, может оказаться, что задача не исчерпана до конца, и потребуется принять очередное решение через определенное время. Принятое решение может способствовать одним и препятствовать другим решениям, которые будут приниматься в будущем. Важно распознать заранее такие динамические проблемы и увидеть, какие возможности могут открыться в будущем благодаря тому или иному решению.

7. Влияние решений на структурные звенья объекта системных исследований. Некоторая выбранная альтернатива может повлиять на работу различных структурных элементов объекта системных исследований, например коллектива исполнителей того или иного подразделения. Принятое решение может затруднять работу одних и способствовать работе других коллективов исполнителей.

8. Коллективное принятие решений. Часто ответственность за выбор альтернативы несет не отдельное лицо, а группа исполнителей. Фактически для определенного круга задач нельзя четко разграничить функции и ответственность лиц, принимающих решение по некоторому кругу вопросов.

Сформулированные особенности задач принятия решений позволяют подойти к их классификации. В основу классификации задач принятия решений положены следующие принципы: число лиц, принимающих решение; вид показателя эффективности; уровень определенности информации, на основании которой принимается решение; зависимость элементов модели проблемной ситуации от времени.

По признаку числа лиц, принимающих решение, различают задачи индивидуального, принимаемого одним лицом, и группового принятия решений, когда решения принимаются коллективным органом. При групповом выборе решений определяющую роль играет проблема согласования индивидуальных предпочтений членов группы. Главной задачей на этом этапе является объединение предпочтений отдельных лиц в единое мнение. Степень согласованности целей при многостороннем выборе может варьироваться от полного совпадения интересов сторон (*кооперативный выбор*) до их противоположности (*выбор в конфликтной ситуации*). Возможны также промежуточные случаи, например *компромиссный выбор, коалиционный выбор, выбор в условиях нарастающего конфликта* и т.д. Решение задачи объединения предпочтений отдельных лиц в единое мнение можно достичь в рамках теории экспертного оценивания.

В зависимости от вида используемого показателя эффективности задачи принятия решений подразделяются на однокритериальные и многокритериальные. Задачи с одним критерием называются скалярными, многокритериальные задачи – задачами с векторным критерием эффективности. Следует иметь в виду, что в задачах с несколькими критериями эффективности различные цели могут противоречить друг другу. Также необходимо учитывать, что критерии могут иметь как количественный, так и качественный характер.

По признаку степени определенности информации о проблемной ситуации различают задачи принятия решений с полностью заданной исходной информацией, требуемой для решения проблемы, и задачи в условиях наличия неопределенности. Задачи принятия решений в условиях определенности называются детерминированными задачами. Они характеризуются наличием полной и достоверной информации о проблемной ситуации, критериях эффективности, ограничениях и последствиях, реализующихся в процессе принятия того или иного решения.

Задачи принятия решений в условиях неопределенности характеризуются наличием стохастической и нестохастической неопределенности.

Характерная особенность задач принятия решений в условиях неопределенности заключается в том, что исход операции зависит как от стратегии, выбираемой лицом, принимающим решения, так и от неопределенных факторов, не известных лицу, принимающему решения в момент его выработки.

По признаку зависимости элементов модели проблемной ситуации от времени различают статические и динамические задачи. Элементы модели в динамических задачах зависят от функций времени, описывающих поведение динамических объектов, и тем самым значительно усложняют процедуры выработки решений.

Режим выбора может быть однократным (разовым) или повторяющимся, допускающим обучение на опыте.

Различные сочетания перечисленных вариантов и приводят к многообразным задачам выбора, которые изучены не в одинаковой степени. В зависимости от типа задач принятия решений используют различные методы выработки решения. Рассмотрим некоторые методы и подходы к решению задач выбора и принятия решений.

13.2. Критериальный способ описания выбора

Наиболее развитым и чаще всего применимым при решении задач выбора является критериальный подход. При применении данного подхода предполагается, что каждую отдельно взятую альтернативу можно оценить численно, значением критерия. Тогда сравнение альтернатив сводится к сопоставлению соответствующих им чисел.

В зависимости от объекта и цели исследования критерии или параметры оптимизации могут быть весьма разнообразны. Чтобы ориен-

тироваться в этом многообразии, введем некоторую классификацию. Итак, среди множества параметров оптимизации выделяют: экономические, технико-экономические, технологические, статистические, психологические, эстетические и т.п. Примерами экономических параметров могут служить прибыль, рентабельность, себестоимость; технико-экономических – производительность, надежность, долговечность; технологических – выход продукта, характеристики качества и пр.

Критерий или параметр оптимизации – это признак, по которому оптимизируют процесс принятия решений. Критерий должен быть количественным, задаваться числом. Необходимо уметь измерять его при любой возможной комбинации воздействующих на исследуемую систему факторов. Множество значений, которые может принимать критерий (параметр оптимизации), называют областью его определения. Области определения могут быть непрерывными и дискретными, ограниченными и неограниченными. Например, выход реакции – это параметр оптимизации с непрерывной, ограниченной областью определения. Он может изменяться от 0 до 100%. Число бракованных изделий, число кровяных телец в пробе крови – примеры параметров с дискретной областью определения, ограниченной снизу. Если нет способа количественного измерения результата, то приходится пользоваться подходом, называемым ранжированием (ранговым подходом). В этом случае параметрам оптимизации присваивается оценка – ранг по заранее выбранной шкале. Ранг – это количественная оценка параметра оптимизации. Она носит условный субъективный характер. Чаще всего ранжирование используется в тех случаях, когда требуется оценить качественный признак. Ранговый параметр имеет дискретную область определения. В простейшем случае область содержит два значения (да – нет, хорошо – плохо). Это может соответствовать способу оценки качества продукции, является она годной или бракованной. Другим примером субъективной оценки может служить оценка знаний на экзамене. Знание – абстрактный параметр. В этом случае оценка знания осуществляется путем проставления ранга, определенного, скажем, по пятибалльной шкале.

Для каждого физически измеряемого параметра оптимизации можно построить ранговый аналог. Потребность в построении такого аналога возникает, если имеющиеся в распоряжении исследователя численные характеристики неточны или неизвестен способ построения удовлетворительных численных оценок. При прочих равных условиях всегда нужно отдавать предпочтение физическому измерению, так как ранговый подход менее чувствителен и с его помощью трудно изучать тонкие эффекты.

Желательно, чтобы параметр оптимизации выражался одним числом. Иногда это получается естественным образом, когда речь идет, скажем, о результатах измерения некоторым прибором. В ряде случаев приходится для определения параметра оптимизации производить вычисления.

Еще одно требование, связанное с количественной природой параметра оптимизации – однозначность в статистическом смысле. Заданному набору значений, воздействующих на исследуемую систему факторов, должно соответствовать одно определенное с точностью до ошибки эксперимента значение параметра оптимизации. Однако обратное утверждение неверно: одному и тому же значению параметра могут соответствовать разные наборы значений факторов.

Для успешного достижения цели исследования необходимо, чтобы параметр оптимизации действительно оценивал эффективность функционирования системы в заранее выбранном смысле. Это требование является главным, определяющим корректность постановки задачи.

Следующее требование к параметру оптимизации – требование универсальности и полноты. Под универсальностью параметра оптимизации понимается его способность всесторонне характеризовать объект. Универсальностью обладают обобщенные параметры оптимизации, которые строятся как функции от нескольких частных параметров. Желательно, чтобы параметр оптимизации имел физический смысл, был простым и легко вычисляемым. Требование физического смысла связано с необходимостью последующей интерпретации результатов процедуры принятия решений.

Рассмотрим постановку задачи критериального выбора. Пусть x – некоторая альтернатива из множества X. Считается, что для всех $x \in X$ может быть задана функция $q(x)$, которая называется *критерием или параметром оптимизации* (*критерием качества, целевой функцией, функцией предпочтения, функцией полезности* и т.д.) и обладает тем свойством, что если альтернатива x_1 предпочтительнее альтернативы x_2 (будем обозначать это $x_1 > x_2$), то $q(x_1) > q(x_2)$ и обратно.

Выбор как максимизация критерия

Если теперь сделать еще одно важное предположение, что выбор любой альтернативы приводит к однозначно известным последствиям (т.е. считать, что выбор осуществляется в условиях определенности) и заданный критерий $q(x)$ численно выражает оценку этих последствий, то наилучшей альтернативой x^* является, естественно, та, которая обладает наибольшим значением критерия:

$$x^* = \arg\max q(x), \text{ при условии, что } x \in X.$$

Задача определения оптимального решения x^*, простая по постановке, часто оказывается сложной для решения, поскольку метод ее решения (да и сама возможность решения) определяется как характером множества X, так и видом критерия. На возможность решения задачи оптимизации критерия оказывает влияние размерность вектора **x** и тип множества X – является ли оно конечным, счетным или континуальным. В свою очередь критерий может быть сформулирован в виде функции или функционала.

Однако сложность определения наилучшей альтернативы на практике существенно возрастает, так как оценивание любого варианта единственным числом обычно оказывается неприемлемым упрощением. Более полное рассмотрение альтернатив приводит к необходимости оценивать их не по одному, а по нескольким критериям, качественно различающимся между собой. При решении конкретных задач системным аналитикам следует учитывать множество критериев: технических, технологических, экономических, социальных, эргономических и пр.

Итак, пусть для оценивания альтернатив используется несколько критериев $q_i(x)$, $i = 1,..., p$. Теоретически можно представить себе случай, когда в множестве X окажется одна альтернатива, обладающая наибольшими значениями всех p критериев; она и является наилучшей. Однако на практике такие случаи почти не встречаются, и возникает вопрос, как же тогда осуществлять выбор.

Путь к единому параметру оптимизации часто лежит через обобщение. Из многих критериальных функций, определяющих альтернативу, трудно выбрать один, самый важный. Будем рассматривать ситуацию, когда необходимо множество критериальных функций свернуть в единый количественный признак. Каждый критерий в общем случае имеет свой физический смысл и свою размерность. Чтобы объединить различные критерии, прежде всего, приходится вводить для каждого из них некоторую безразмерную шкалу. Шкала должна быть однотипной для всех объединяемых критериев – это делает их сравнимыми.

Сведение многокритериальной задачи к однокритериальной

Рассмотрим наиболее употребительные способы решения многокритериальных задач. Первый способ состоит в том, чтобы многокритериальную задачу свести к однокритериальной. Это означает введение суперкритерия, т.е. скалярной функции векторного аргумента:

$$q_0(x) = q_0\ (q_1(x), q_2(x),..., q_p(x)).$$

Суперкритерий позволяет упорядочить альтернативы по величине q_o, выделив тем самым наилучшую (в смысле этого критерия). Вид функции q_o определяется тем, как мы представляем себе вклад каждого критерия в суперкритерий. Обычно для реализации данной процедуры используют аддитивные

$$q_0 = \sum_{i=1}^{p} \frac{\alpha_i q_i}{s_i}.$$

или мультипликативные функции

$$1 - q_0 = \prod_{i=1}^{p} \left(1 - \frac{\beta_i q_i}{s_i}\right).$$

Коэффициенты s_i обеспечивают безразмерность критериального значения (частные критерии могут иметь разную размерность, и тогда некоторые арифметические операции над ними, например сложение, не имеют смысла). Коэффициенты α_i, β_i отражают относительный вклад частных критериев в суперкритерий.

Итак, при данном способе задача сводится к максимизации суперкритерия:

$$x^* = \arg\max q_0 \,(q_i\,(x),..., q_p\,(x)), \text{ при } x \in X.$$

Очевидные достоинства объединения нескольких критериев в один суперкритерий сопровождаются рядом трудностей и недостатков, которые необходимо учитывать при использовании этого метода.

Рассмотрим примеры построения обобщенных критериальных показателей. Пусть рассматриваемая альтернатива характеризуется n частными критериальными функциями $q_i(i=1, 2,.., p)$. Каждая из функций q_i имеет свой физический смысл и, чаще всего, свою размерность. Введем простейшее преобразование: набор данных для каждого q_i поставим в соответствие с самым простым стандартным аналогом – шкалой, на которой имеется только два значения: 0 – брак, неудовлетворительное качество, 1 – годный продукт, удовлетворительное качество. В ситуации, когда каждый преобразованный критерий принимает только два значения 0 и 1, естественно желать, чтобы и обобщенный критерий принимал одно из двух возможных значений, причем так, чтобы значение 1 имело место тогда и только тогда, когда все частные критериальные показатели приняли бы значение равное 1. Если же, хотя бы один из показателей принял значение, равное 0, то и обобщенный критерий будет равным нулю. В этом случае для построения обобщенного критериального показателя естественно воспользоваться формулой

$$q_0 = \prod_{i=1}^{p} q_i.$$

Иногда применяют запись

$$q_0 = \sqrt[p]{\prod_{i=1}^{p} q_i},$$

где q_0 – обобщенный критериальный показатель; q_i – частные критериальные функции.

Если для каждого из частных критериев известен «идеал», к которому нужно стремиться, то можно предложить следующий метод построения обобщенного параметра оптимизации (критериального показателя). Пусть q_{i0} – наилучшее (идеальное) значение i-го критерия. Тогда $(q_i - q_{i0})$ – мера близости к идеалу. Поскольку при построении обобщенного критериального показателя необходимо, чтобы различные показатели были сопоставимы, надо привести их к безразмерному значению. Это можно осуществить, отнормировав полученное отклонение следующим образом

$$(q_i - q_{i0}) / q_{i0}.$$

Чтобы исключить влияние знаков, возведем последнее выражение в квадрат

$$((q_i - q_{i0}) / q_{i0})^2.$$

Тогда обобщенный критериальный показатель можно записать

$$q_0 = \sum_{i=1}^{p} ((q_{i0} - q_i) / q_{i0})^2.$$

Если все частные критерии совпадают с идеалом, то q_0 станет равным нулю. В таком правиле определения обобщенного критериального показателя каждый частный критерий входит в формулу на равных правах. На практике показатели бывают далеко не равноправны. Введем некоторые весовые коэффициенты α_i, тогда правило определения обобщенного параметра можно записать в виде:

$$q_0 = \sum_{i=1}^{p} \alpha_i ((q_{i0} - q_i) / q_{i0})^2, \text{ причем } \sum_{i=1}^{p} \alpha_i = 1,\ \alpha_i > 0.$$

Задача определения значений весовых коэффициентов – это отдельная задача, она не входит в рамки обсуждения.

Если удается построить обобщенный критериальный показатель, то метод поиска оптимального решения будет аналогичен методу оптимизации в случае единственного критерия. В зависимости от вида критериального показателя в качестве метода решения могут быть использованы прямые оптимизационные процедуры, в случае невозможности аналитического решения используются численные методы.

Условная максимизация

Вторым способом решения задач выбора в условиях наличия нескольких критериальных показателей является сведение задачи к задаче условной максимизации. Данный метод решения задач выбора целесообразно применять в тех случаях, когда заведомо известно, что частные критерии неравнозначны между собой, одни из них более важны, чем другие. В этом случае происходит выделение основного, главного критерия, остальные рассматриваются как вспомогательные, дополнительные к выделенному. Такое разделение критериев позволяет сформулировать задачу принятия решений как задачу определения условного экстремума:

$$x^* = \arg\left(\max f(x)/q_i(x) \le c_i,\ i = 1,\ 2,\ldots,\ p\right),$$

где через $f(x)$ обозначен основной критерий; $q_i(x)$ – вспомогательные или второстепенные критериальные функции. В ограничениях могут иметь место различные сочетания знаков: от строгого равенства до строгого неравенства. Например, если вспомогательный критерий характеризует стоимость затрат, то разумнее задавать их верхний уровень и формулировать задачу с ограничениями в виде неравенств. Для решения задач в такой постановке разработаны специальные методы математического программирования, рассмотренные в гл. 2 работы [57].

Нахождение паретовского множества

Следующий способ многокритериального выбора состоит в сравнении альтернатив между собой по всем сформированным критериям и выделении подмножества наилучших альтернатив. В данном подходе отказываются от поиска одной единственной наилучшей альтернативы. Решающее правило в этом случае строится на основе аксиомы В. Парето, которая формулируется следующим образом: «Если в задаче принятия решений частные критерии независимы по предпочтению и значение каждого из них желательно увеличивать, то из двух альтернатив, характеризуемых набором частных критериев, предпочтительнее та, для которой выполняются соотношения $q_{1i}(x) \ge q_{2i}(x)$ по всем i, где первый индекс характеризует номер стратегии, второй индекс – номер частного критерия. То есть первая альтернатива предпочтительнее второй только в том случае, когда значения ее частных критериев не меньше значений частных критериев второй альтернативы. Если все значения частных критериев одной альтернативы равны значениям критериев другой, то альтернативы равнозначны». Таким образом, предпочтение одной альтернативе перед другой можно отдавать только если первая по всем критериям лучше второй. Если же предпочтение хотя бы по одному критерию расходится с предпочтением по другому, то такие альтернативы признаются несравнимыми. В результате попарного сравнения альтернатив все худшие по всем критериям альтернативы отбрасываются, а все оставшиеся несравнимые между собой принимаются. Если все максимально достижимые значения частных критериев не относятся к одной и той же альтернативе, то принятые альтернативы образуют множество Парето и выбор на этом заканчивается. При необходимости выбора единственной альтернативы следует привлекать дополнительные соображения: либо корректировать систему предпочтений, либо обращаться к услугам экспертов, либо воспользоваться методами, рассмотренными ранее (построение обобщенного критерия или сведение задачи к задаче поиска условного экстремума).

13.3. Выбор в условиях неопределенности

Рассмотренные до настоящего времени задачи осуществления выбора формулировались таким образом, что последствия сделанного выбора предполагались однозначно определенными. Выбор одной из альтернатив был связан с известным однозначным следствием. В этом случае проблема выбора состояла в сравнении разных вариантов, т.е. альтернатив.

В реальной практике в большинстве случаев приходится иметь дело с более сложной ситуацией, когда выбор альтернативы неоднозначно определяет последствия сделанного выбора. Адекватное реальности описание проблемы практически всегда содержит различного рода неопределенности, отражающие то естественное положение, в котором находится исследователь: любое его знание относительно и неточно. Принято различать три типа неопределенностей. С одной стороны это неопределенности природы. К данному виду неопределенностей относят факторы, неизвестные исследователю. Далее неопределенности противника. Нередки ситуации, когда исследователь принимает решения в условиях, при которых результаты его решений не строго однозначны. Они зависят от действий других лиц (партнеров, противников и т.п.), которые он не может учесть или предсказать. И, наконец, существуют так называемые неопределенности целей. Такая ситуация возникает в случае, когда при принятии решений формулируется несколько целей, которые в общем случае могут противоречить друг другу. В этом случае мы приходим к многокритериальной задаче выбора. Подходы к принятию решения в условиях многокритериальной задачи рассмотрены в предыдущем параграфе.

Существует также классификация неопределенностей по соотношению альтернатив и исходов. Различают неопределенности дискретного и непрерывного типа, стохастические и расплывчатые неопределенности.

Рассмотрим следующую ситуацию: имеется набор возможных исходов $y \in Y$, из которых один окажется совмещенным с выбранной альтернативой, но какой именно – в момент выбора неизвестно, а станет ясным позже, когда выбор уже сделан и изменить ничего нельзя. Будем предполагать, что с каждой альтернативой x связано одно и то же множество исходов Y, для разных альтернатив одинаковые исходы имеют разное значение. В случае дискретного набора альтернатив и исходов такую ситуацию можно изобразить с помощью матрицы, представленной в табл. 13.1:

Таблица 13.1

X\Y	y_1	…	y_j	…	y_m
x_1	q_{11}	…	q_{1j}	…	q_{1m}
…	…	…	…	…	…
x_i	q_{i1}	…	q_{ij}	…	q_{im}
…	…	…	…	…	…
x_n	q_{n1}	…	q_{nj}	…	q_{nm}

В этой матрице все возможные исходы образуют вектор $y = (y_1,\ldots, y_m)$, числа q_{ij} выражают оценку ситуации, когда сделан выбор альтернативы x_i и реализовался исход y_j. В конкретных случаях числа q_{ij} могут иметь различный смысл: это может быть «выигрыш», «потери», «платежи» и т.п. Если все строки $q_i = (q_{i1},\ldots, q_{im})$ при любых i одинаковы, то проблемы выбора нет. Если же строки матрицы различны, возникает вопрос, какую альтернативу предпочесть, не зная заранее, какой из исходов реализуется.

Аналогичная ситуация возникает в случае, когда множества X и Y непрерывны. В этом случае зависимость между альтернативами и исходами задается в виде функции $q(x, y)$, $x \in X$, $y \in Y$ с соответствующей постановкой вопроса о выборе x.

Введенных до настоящего времени параметров недостаточно для формальной постановки задачи выбора. При различной конкретизации этой задачи она приобретает различный смысл и требует различных методов решения. Методологической базой для решения такого рода задач является теория игр. Метод решения конкретной задачи будет зависеть от характера воздействующих на ситуацию факторов, не зависящих от лица, принимающего решения. Здесь необходимо различать уже отмеченные ранее неопределенности природы и неопределенности противника. В задачах выбора с природной неопределенностью считается, что исходы $y = (y_1,\ldots, y_m)$ есть возможные состояния природы. Желательность каждой альтернативы x_i зависит от того, каково состояние природы, но узнать это состояние исследователь сможет лишь после того, как сделан выбор.

В задачах выбора с неопределенностью противника предполагается, что исходы Y – это множество альтернатив, на котором выбор осуществляет второй игрок. В отличие от природной неопределенности игрок преследует свои интересы, отличные от интересов исследователя (первого игрока). При этом матрица $Q = \|q_{ij}\|$, характеризующая оценки ситуаций с точки зрения системного исследователя или лица, принимающего решения, выбирающего x, уже недостаточна для описания всей ситуации. Необходимо задать вторую матрицу $U = \|u_{ji}\|$, описывающую систему предпочтений с позиции противника. Задание X, Y, Q и U называется нормальной формой игры. Расхождение между матрицами Q и U определяет степень антагонизма лица, принимающего решения, и его противника. Остановимся на анализе природных неопределенностей и неопределенностей противника.

Природные неопределенности

Рассмотрение природных неопределенностей начнем с примера. Пусть перед системным аналитиком стоит задача проложить маршрут океанского лайнера и распорядиться запасом горючего так, чтобы судно как можно быстрее дошло до пункта назначения. При этом известно, что время нахождения в пути будет существенно зависеть от погодных условий на трассе в момент ее прохождения. В данном случае погодные условия представляют собой природную неопределенность. Ситуация типична для ряда задач принятия решений. Рассмотрим ее формализацию. Запишем целевую функцию, например время нахождения судна в пути, в следующем виде

$$T = f(\vec{x}, \alpha),$$

где $\alpha \in G_\alpha$ – некоторый параметр, который заранее неизвестен. Выбор альтернативы x, которая бы минимизировала значение целевой функции будет существенно зависеть от того, какое значение параметра α реализуется в момент прохождения судна по трассе. Таким образом, говоря о природной неопределенности, имеем в виду, что выбор осуществляется в условиях, когда целевая функция задана, но не совсем точно, а именно, она содержит неопределенный параметр. Постановка задачи выбора будет выглядеть следующим образом

$$T = f(\vec{x}, \alpha) \to \min_x.$$

При этом решение будет представлять собой функцию от параметра α: $x = x(\alpha)$. Если никакой дополнительной информацией о факторе неопределенности α лицо, принимающее решение, не располагает, то результат оптимизации произволен. В реальных ситуациях информация о параметре α обычно имеет вид $\alpha \in G_\alpha$, где G_α – некоторое множество. Но подобной информации также недостаточно для однозначного решения задачи. Решение $x = x(\alpha)$ определяет лишь некоторое отображение множества неопределенности природных факторов G_α на множество альтернатив X, которое естественно назвать множеством неопределенности результата. Множество неопределенности результата X – важная характеристика процедуры принятия решений, но его построение сопряжено с большим объемом сложных вычислений.

Рассмотрим другой подход, который дает строгую оценку. Данный подход основан на применении принципа наилучшего гарантированного результата. Суть его состоит в следующем. Так как для любой альтернативы x справедливо неравенство

$$\max_\alpha f(x,\alpha) \ge f(x,\alpha),$$

то и для любого $\alpha \in G_\alpha$ будет справедливо соотношение

$$f^* = \min_x \max_\alpha f(x,\alpha) \ge \min_x f(x,\alpha).$$

Число f^*, определенное данным способом, называется гарантированной оценкой, а соответствующая альтернатива $x = x^*$ – гарантированной стратегией в том смысле, что каково бы ни было значение параметра неопределенности α, выбор $x = x^*$ гарантирует, что при любом α значение целевой функции будет не меньше, чем f^*.

Выбор гарантирующей стратегии поведения – это рациональный способ принятия решений. В результате использования данной стратегии лицо, принимающее решение, гарантирует исход, защищенный от всевозможных случайностей. Каковы бы ни были неконтролируемые факторы, в результате такого выбора обеспечивается значение целевой функции не меньше, чем f^*. Данный результат может быть улучшен, если принять решение, связанное с определенным риском. Критерии принятия решений в условиях риска рассмотрим несколько позже.

Неопределенности противника

Перейдем теперь к описанию неопределенностей, связанных с существованием активных партнеров или противников, действия которых лицо, принимающее решение, не может полностью контролировать.

В теории принятия решений особое место занимает изучение ситуации, в которой участвует много субъектов (много оперирующих сторон), причем каждый из них стремится достичь своей цели

$$f_i(x_1, \dots, x_n) \to \max_x$$

и имеет для этого определенные возможности, которые описываются вектором x, $x \in X$. Заметим, что формально такая ситуация включает в себя проблему многокритериальности, требующую определения вектора x, при котором достигается максимум критериев $f_i(X)$. В самом деле, если отождествить цель каждого из субъектов с его критерием $f_i(X)$, а в качестве описания множества X принять условия

$$x_1 = x_2 = \dots = x_n,$$

то в результате будет получен частный случай задачи со многими активными партнерами. Общий случай ситуации со многими субъектами гораздо сложнее и требует для своего анализа целый ряд специфических гипотез. Поясним это на примере двух субъектов. Итак, пусть два субъекта *А* и *Б*, располагающие возможностью выбора векторов x и y, стремятся к достижению своих целей, которые будем записывать в виде

$$f(x,y) \to \max,\ \varphi(x,y) \to \max,\ x \in X,\ y \in Y.$$

В частном случае может оказаться, что $f(x, y) = -\varphi(x, y)$; такую ситуацию называют антагонистической. Антагонистические ситуации были предметом множества исследований и сделались основным объектом изучения в теории игр. Чисто антагонистическая ситуация является в известном смысле вырожденной. Наиболее типичен конфликт, в котором интересы партнеров или противников не совпадают, но и не строго противоположны.

Общий случай нетождественности интересов (целей) партнеров (субъектов) называют конфликтом. При изучении конфликтных ситуаций, т.е. при изучении возможных способов выбора, удобно отождествлять исследователя с одним из субъектов. Условимся называть лицо, принимающее решение субъектом A.

В связи с тем, что исход выбора зависит от выбора субъекта $Б$, необходимо принять ту или иную гипотезу о его поведении, которое, в свою очередь, будет зависеть от характера информированности субъекта $Б$. Здесь возможно существование нескольких гипотез (нескольких случаев).

1) Каждый из субъектов не имеет никакой информации о выборе, который сделал другой субъект. В этом случае имеется возможность найти гарантированную оценку. Для субъекта A она будет выражаться формулой

$$f^* = \max_x \min_y f(x, y),$$

для субъекта $Б$ – формулой

$$\varphi^* = \max_y \min_x \varphi(x, y).$$

Решая сформулированные задачи, определяют векторы x^* и y^*, которые реализуют значения f^* и φ^*. Такое решение означает, что сделав выбор $x = x^*$, лицо, принимающее решение, при любых условиях (любом выборе $y \in Y$) гарантирует, что значение целевой функции $f(x, y)$ будет не меньше, чем f^*.

В этой ситуации могут быть предложены различные варианты риска. Например, имеется возможность принять гипотезу о том, что другой субъект использует гарантирующую стратегию $y = y^*$. Тогда необходимо делать другой выбор:

$$f^* = \max_x f(x, y^*).$$

В этом случае определяется вектор $x = x^1$ и соответствующее значение функции $f = f^1$. При этом $f^1 \geq f$, но если противник сделает иной выбор, например, $y = y^1$, то может оказаться, что $f(x^1, y^1) < f^*$. В данной ситуации следует иметь в виду, что риск есть риск, и если исследователь сформулировал гипотезу и она оказалась неверной, то и результат может оказаться не тем, который ожидается.

2) Вторая ситуация характеризуется тем, что исследователь во время принятия решения имеет информацию о выборе субъекта $Б$, т.е. ему известно выбранное субъектом $Б$ значение y.

Тогда стратегию (выбор) x следует искать в виде функции $x = x(y)$. Данная стратегия может быть определена эффективно, для этого требуется решить задачу оптимизации

$$f^{(2)} = \max_x f(x, y).$$

Решая данную оптимизационную задачу, определяем искомую стратегию $x = x(y)$. Для этого случая можно также вычислить гарантированный результат (он будет отличен от f^*):

$$\hat{f} = \min_y \max_x f(x,y)$$

и во всех случаях $\hat{f} \geq f^*$.

Заметим, что выбирая свою стратегию – вектор x – лицо, принимающее решение, никак не можем повлиять на выбор, который сделал другой субъект.

3) Предположим теперь, что субъект $Б$ в момент принятия своего решения будет знать выбор лица, принимающего решение (субъекта A); например, субъект A обязан сообщить свое решение субъекту $Б$. В этом случае исключается возможность оказать влияние на выбор, который сделает субъект $Б$. В самом деле, если исследователь знает целевую функцию субъекта $Б$, то естественно сделать предположение о том, что субъект $Б$ будет делать выбор из условия

$$\varphi = \max_y \varphi(x, y).$$

Решая данную задачу исследователь может определить отклик субъекта $Б$ на свой выбор, который, согласно сформулированной гипотезе, будет оптимальной стратегией субъекта $Б$:

$$y = y(x).$$

Теперь имеется возможность распорядиться выбором вектора x. Подставляя последнее выражение в формулу для целевой функции $f(x, y)$, можно получить

$$f(x, y(x)) = F(x).$$

Свой выбор исследователь может сделать из условия

$$f^{(3)} = F(x).$$

Итак, информация о том, что субъект $Б$ будет знать выбор лица, принимающего решения, а также гипотеза о том, что субъект $Б$ выберет свою оптимальную стратегию, позволяют так воздействовать на его выбор, чтобы он в максимальной степени соответствовал целям исследователя.

Описанная ситуация достаточно часто встречается на практике и ей нетрудно придать ту или иную экономическую интерпретацию. Так, вектор x можно отождествить с ресурсом, а функцию $y = y(x)$ назвать производственной функцией, которая описывает наивыгоднейший для субъекта *Б* способ использования ресурса. Таким образом, субъекту *Б* выделяется такое количество ресурса, чтобы его деятельность наилучшим образом соответствовала целям субъекта *А*.

Критерии сравнения альтернатив

Теория принятия решений в настоящее время представляет собой самостоятельную научную дисциплину. По данному направлению опубликовано множество монографий [3], поэтому не имеет смысла подробно останавливаться на результатах данной теории. Изложим основополагающие идеи и подходы к решению задач теории принятия решений.

Центральным моментом данной теории является введение критерия для оценки выбираемого варианта. В силу неопределенности исхода требуется дать оценку сразу целой строке матрицы (см. табл.13.1). Имея такие оценки для всех строк, и сравнивая их, можно приступать к решению задачи выбора. Для этого требуется ввести критерии сравнения альтернатив.

Самым распространенным является уже рассмотренный ранее максиминный критерий, который гарантирует результат выбора по принципу «наименьшего из зол». Рассмотрим суть данного критерия применительно к табл.13.1 (матрица платежей). В каждой из строк матрицы определяем наименьший из выигрышей $\min_j q_{ij}$, который характеризует гарантированный выигрыш в самом худшем случае. Данное значение считаем оценкой альтернативы x_i. Далее определяем альтернативу x^*, обеспечивающую наибольшее значение этой оценки:

$$x^* = \arg \max_i \min_j q_{ij}.$$

Эта альтернатива называется оптимальной по максиминному критерию. Если платежную матрицу определить не через выигрыш, а через проигрыш, то тот же принцип рассмотрения приводит к минимаксному критерию. Минимаксный критерий использует оценочную функцию, соответствующую позиции крайней осторожности. Правило выбора решения в соответствии с минимаксным критерием можно интерпретировать следующим образом: матрица платежей, на основании которой осуществляется поиск оптимального решения, дополняется еще одним столбцом из наименьших результатов q_{ir} каждой строки. Выбрать надлежит те варианты x_i, в строках которых стоят наибольшие значения q_{ir} этого столбца.

Выбранные таким образом варианты полностью исключают риск. Это означает, что лицо, принимающее решение, не может столкнуться с худшим результатом, чем тот, на который он ориентируется. Какие бы условия ни сложились при развитии ситуации, соответствующий результат не может оказаться ниже q_{ir}. Это свойство заставляет считать минимаксный критерий одним из фундаментальных. В задачах системных исследований применительно к техническим системам данный критерий находит наиболее широкое применение. Однако следует заметить, что положение об отсутствии риска стоит различных потерь. Осознавая данную ситуацию, в теории принятия решений предложено большое количество критериев, учитывающих всевозможные особенности конкретных задач и предпочтений субъектов, принимающих решения. Это такие критерии как Байеса–Лапласа, Сэвиджа, Гурвица, Ходжа–Лемана, Гермейера и т.д. Детальное описание данных критериев можно найти, например, в [3].

Применение минимаксного критерия бывает оправдано в ситуациях, которые характеризуются следующими обстоятельствами:

- о возможности появления внешних состояний ничего не известно;
- приходится считаться с возможностью появления нескольких различных по своему характеру негативных внешних состояний;
- решение реализуется лишь один раз;
- при принятии решения необходимо исключить любой риск, т.е. ни при каких условиях не допускается получать результат, меньший чем q_{ir}.

13.4. Концепция риска в задачах системного анализа

Допущение пусть даже малой вероятности α принятия ошибочного решения не исключает возможности риска. Полное устранение риска при принятии решений практически даже и не требуется; мало того, определенная степень риска вводится сознательно, так как принятие решения без риска, например, с предельно пессимистической позиции, как правило, невыгодно. Однако при этом разумный риск следует отличать от риска азартного игрока. Любой риск, во-первых, должен учитываться по возможности полно, описываться количественными характеристиками и ограничиваться, а во-вторых, ни в коем случае не превышать уровень, при котором результат достигается с достаточной надежностью. В качестве опорного для оценки риска принимается решение, получаемое на основании минимаксного критерия, так как данное решение соответствует позиции крайней осторожности.

В литературе встречается различное понимание термина «риск» и в него иногда вкладывают довольно сильно отличающиеся друг от друга трактовки. Однако общим во всех этих представлениях является то, что под риском понимают неуверенность, произойдет ли нежелательное событие и возникнет ли неблагоприятное состояние. Проблемы риска, тем не менее, часто приходится решать, и выбор варианта решения в общем случае, так или иначе, связан с риском. Поэтому попытаемся найти такое определение риска, которое в достаточной степени соответствовало бы содержанию рассматриваемых задач и в то же время отвечало бы общей концепции теории принятия решений.

С понятием риска часто связывается представление о возможных или грозящих событиях с катастрофическими последствиями и потерями. Отсюда следует точка зрения, что такого события следует избежать любой ценой. При ожидаемых потерях, связанных с жизнью и здоровьем, это представление выражено особенно резко, и оно ясно формулируется в соответствующих инструкциях, например по технике безопасности. Конечно, нужно четко сказать, что полностью свободной от риска техники, несмотря на самые большие затраты, не существует. Однако техническим задачам далеко не всегда сопутствуют такие отягчающие обстоятельства. Ущерб вследствие решения, принятого с учетом риска, может оказаться ничтожно малым по сравнению с затратами на то, чтобы избежать такого ущерба. Учитывая необходимость количественных оценок, можно предложить следующую формулировку понятия риска: величина риска, связанная с реализацией нежелательного события или состояния, есть произведение величины последствий реализации события на меру возможности его наступления.

Обозначим через A некоторое нежелательное событие или состояние, которое может произойти. Пусть данное событие характеризуется некоторой вероятностью наступления $P(A)$ и некоторыми последствиями $U(A)$. Тогда риск, связанный с наступлением события A, будет определяться следующим образом:

$$R(A) = U(A)\, P(A).$$

Последствие в принципе нежелательного события или состояния может в соответствии со своей величиной описываться и оцениваться специфическими параметрами. Диапазон при этом весьма широк – от экономических до этических ценностей. Мерой возможности наступления события служит вероятность q его наступления.

При угрозе материальным ценностям степень риска измеряют в денежном выражении. Если различные последствия нежелательного

события одинаковы или очень велики, то для сравнения достаточно рассматривать одни соответствующие вероятности. Наряду с этим может быть угроза ценностям, которую нельзя выразить количественно, например, когда последствия события нельзя предусмотреть достаточно полно. Примером могут служить последствия выхода из строя прибора, используемого в различных областях народного хозяйства, которые поставщик оценить не может. В этом случае мерой риска остаётся принять вероятность превышения предела нагрузки. При риске, связанном со здоровьем, последствия могут быть частично оценены количественно в таких категориях, как простой в работе или расходы на оплату подменяющего персонала и т.п. При риске, связанном с летальным исходом, количественные оценки последствий в большинстве случаев отсутствуют. При существовании угрозы жизни люди в настоящее время почти всегда, тем не менее, работают. Особые проблемы возникают в случаях, когда опасность грозит и людям, и материальным ценностям одновременно, и желательно меру такого риска сравнить с другими рисками. Риск может быть явно связан с факторами, не поддающимися учету. Так, эстетический вред, наносимый построенным сооружением уникальному ландшафту, практически невозможно оценить.

Заключительная оценка риска бывает проста, когда имеют дело только с угрозой материальным ценностям, а возможный ущерб выражен количественно. При угрозе материальным ценностям и невозможности количественно выразить возможный ущерб нужно этот ущерб оценить приблизительно и продолжать рассмотрение, мирясь с таким недостатком информации. Поскольку нецелесообразно идти на сколь угодно большие затраты, чтобы устранить риск полностью, нужно оценить угрозу людям. Субъективные оценки сильно отклоняются от известных частот реализации тех или иных нежелательных событий. Значения риска субъективно привлекательной деятельности обычно занижаются. Риск события, на которое оценивающему трудно или невозможно оказать влияние, наоборот, обычно переоценивается. Риск события катастрофического характера, как правило, тоже получает более высокую оценку. Кроме того, субъективные оценки меняются со временем. В результате из-за этих некорректностей субъективные оценки не могут быть положены в основу технических решений.

Сравнение рассматриваемой рискованной ситуации с возникавшими в прошлом аналогичными ситуациями дает для оценки риска более надежные исходные данные. Проблема оценки этим, однако, все же не решается. В отдельных случаях, конечно, можно довольствоваться требованием, чтобы допустимый риск был заведомо ниже имевшего место в аналогичных ситуациях ранее. Но в других случаях, особенно при очень высоком уровне затрат, проблема остается нерешенной. Требование четко ограничить допустимые вероятности реализации нежелательного события наталкивается на препятствия, обусловленные следующими положениями:

• такого рода границы должны быть независимыми от экономических затрат, но аналогичная независимость должна существовать также для угрозы безопасности людей и материальным ценностям;

• лицо, принимающее решения, должно для подобных границ принимать общее решение, учитывающее всю специфику частных случаев;

• одно лишь утверждение, что такие границы будут соблюдаться, может освободить лицо, принимающее решение, от обязанности анализировать ситуацию дальше и направлять свои усилия на дальнейшее снижение угрозы безопасности людей. При этом возможны случаи, когда ценой очень небольших затрат опасность может быть еще больше снижена, а этим пренебрегают, поскольку границы уже установлены;

• утверждение, что выдерживаются определенные границы, предполагает качественное единство данных, что на самом деле недостижимо, так как имеют место проблемы самого различного типа;

• ограничения допустимого риска зависят от времени и меняются с изменениями технических и экономических возможностей общества.

Угроза безопасности людей чаще всего состоит из многих составляющих риска, например, из основного существующего риска, риска вследствие ошибок и риска, на который идут сознательно при известных условиях.

Любой математический алгоритм оценки риска должен исходить из того, что твердо установлен экономический эквивалент угрозы. Этот эквивалент должен быть обоснован в том смысле, что он соответствует затратам, которые общество при данных условиях может себе позволить, чтобы предотвратить или уменьшить угрозу. Необходимо воспрепятствовать тому, чтобы, с одной стороны, ценой больших затрат был уменьшен и без того незначительный риск, а с другой – чтобы оставался большой риск, который может быть устранен с небольшими затратами. Установить такой эквивалент – еще не значит добиться успеха. И при многоцелевых решениях эквивалент такого типа не удается получить без влияния субъективных факторов. Тем не менее, эти эквиваленты делают более ясным риск при принятии решения и помогают лучше определить ответственность за выполненную оценку. Решения, связанные с риском, всегда остаются для исследователя сомнительными, так как нельзя заранее определить затраты для четкого разделения во всех случаях оправданного и неоправданного риска. Про-

контролировать, был ли оправдан данный риск, удается всегда только после наступления нежелательного события, и возможно это только при оправданных убытках.

Примеры формирования риска в задачах системных исследований

Технический риск. Технические объекты подвергаются опасности при возрастании нагрузки. Если при этом будет превзойден предел (например, прочности), произойдет выход объекта из строя. В данном частном случае под риском целесообразно понимать вероятность наступления определенного сочетания неблагоприятных событий. Риск целесообразно описывать вероятностью при следующих условиях:

а) если последствия выхода из строя объекта нельзя выразить экономическими показателями;

б) если экономические соображения играют подчиненную роль;

в) если экономические последствия важны, но не поддаются количественной оценке;

г) если последствия столь велики, что без особых рассуждений нужно минимизировать вероятность выхода объекта из строя.

Технический риск характеризуется, таким образом, вероятностью превышения предела. Если X и Y – случайные переменные, причем X характеризует нагрузку, а Y – несущую способность, то для технического риска справедливо соотношение

$$R_m = p(X > Y).$$

Если существуют плотности распределения нагрузки и несущей способности $f_X(x)$ и $f_Y(y)$, то при независимости X и Y можно записать

$$R_m = \int_{-\infty}^{0} \left(\int_{-\infty}^{\infty} f_X(u-v) f_Y(u) du \right) dv.$$

Если, кроме того, известна зависимость плотностей распределения от времени $f_X(x, t)$ и $f_Y(y, t)$, то получим

$$R_m = \int_{-\infty}^{0} \left(\int_{-\infty}^{\infty} f_X(u-v,t) f_Y(u,t) du \right) dv.$$

Зависимость плотности распределения нагрузки от времени отражает характер воздействия факторов во времени на исследуемый объект. Зависимость плотности распределения несущей способности от времени отражает процессы старения в самом исследуемом объекте.

Таким образом, задача определения технического риска сводится к определению плотностей распределения нагрузки и несущей способности.

Технико-экономический риск. В данном пункте рассмотрим случай, когда последствия при конкретных нагрузке X и несущей способности Y можно описать функцией $h(x, y)$. На первый взгляд кажется важным рассмотреть критический случай, когда $x > y$, т.е. когда уровень нагрузки превышает несущую способность. Это условие можно было бы выразить в виде $h(x, y) = 0$ для $x \leq y$ и однозначно оценить критический случай $x>y$ простым утверждением, что при этом $h(x, y) = 1$. Однако реальные данные из практики показывают, что первые признаки разрушения появляются еще до достижения нагрузкой несущей способности, и, наоборот, в других случаях, при нагрузке, превышающей несущую способность, объект продолжает функционировать. Так что ограничение функции $h(x, y)$ всего двумя значениями 0 и 1 может оказаться слишком грубым описанием. Определим технико-экономический риск R_e при независимости нагрузки X и несущей способности Y и известных плотностях распределения $f_X(x)$ и $f_Y(y)$ ожидаемых случайных величин следующим соотношением:

$$R_e = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u,v) f_X(u) f_Y(v) du dv.$$

Для определенного данного значения x нагрузки условное математическое ожидание риска равно

$$R_e(x) = \int_{-\infty}^{\infty} h(u,v) f_Y(v) dv .$$

Угроза безопасности людей. Если при анализе безопасности технических объектов существуют события A_i, при достижении которых возникает угроза здоровью обслуживающего персонала, описываемая количественно функцией потерь $h(A_i)$, то для описания функции риска можно пользоваться выражениями, аналогичными тем, которые были получены для технико-экономического риска. Если функцию потерь сформировать не удается, то в качестве меры риска допустимо использовать вероятность наступления нежелательного события, как это было сделано в случае технического риска. Дополнительно следует, однако, рассмотреть еще ряд возможных случаев.

Угроза при эксплуатации технических средств определяется двумя категориями влияний – событиями, представляющими угрозу, и попаданием в опасную зону. События, представляющие угрозу, и попадание

в опасную зону – явления случайные. В предположении равномерности распределения событий во времени можно получить следующие выражения для вероятности наступления события, представляющего угрозу:

$$P(A) = \frac{T_A}{T}$$

и для вероятности попадания в опасную зону:

$$P(E) = \frac{T_E}{T} .$$

То есть вероятности выражаются как отношения интервалов времени. Здесь приняты следующие обозначения: T_A – суммарная продолжительность события, представляющего угрозу; T_E – продолжительность пребывания в опасной зоне; T – рассматриваемый интервал времени, для которого принимается решение.

Если событие A, представляющее угрозу, и пребывание в опасной зоне E независимы, то вероятность совместной реализации этих двух событий можно оценить по формуле

$$P(A \cap E) = P(A)P(E).$$

Эта формула говорит, что при данных значениях $P(A)$ и $P(E)$ следует считаться с вероятностью совпадения опасностей, т.е. одновременного наступления представляющего угрозу события и попадания в опасную зону в рассматриваемый отрезок времени. Однако отсюда не следует, с какой вероятностью нужно ожидать реализации по меньшей мере одной угрозы. Поэтому при использовании величины как вероятности угрозы возможны серьезные ошибки в интерпретации рассматриваемых ситуаций.

13.5. Принятие решений в условиях стохастической неопределенности

При решении вопросов системных исследований, таких как проектирование автоматизированных систем, организация их эксплуатации и т.п. возникает большое количество задач, в основе которых лежат вероятностные модели объектов или процессов, описывающих исследуемые явления. Примерами таких задач являются задачи оценивания параметров эффективности и надежности технических средств, прогнозирование поведения параметров системы, задачи, связанные с контролем работоспособности и диагностикой неисправностей при функционировании систем и отдельных компонентов, задачи организации оптимального обслуживания технических средств, задачи обоснования срока службы отдельных элементов, узлов, подсистем. В большинстве случаев существуют естественные вероятностные модели, отражающие реальный ход процессов динамического поведения объектов. Эти модели строятся на основе математических, физических или технических закономерностей, отражающих функционирование отдельных объектов, воздействие объектов друг на друга. В ряде случаев для построения вероятностной модели используется объективная информация о поведении объектов в процессе эксплуатации. По результатам обработки этой информации с помощью специальных методов математической статистики производят построение зависимостей. Например, существующие методы проверки статистических гипотез позволяют обоснованно подойти к выбору закона распределения некоторой случайной величины на основании реализовавшихся значений.

И, наконец, существуют ситуации, когда построение вероятностной модели происходит субъективно, на основании интуиции и опыта системного аналитика. При построении вероятностной модели на основании субъективной информации необходимо тщательно анализировать комбинации состояний динамических объектов, возможность появления критических ситуаций, степень вероятности ситуаций. Необходимо также привлекать информацию о сходных процессах, результатах приближенных расчетов.

Следует заметить, что при построении вероятностных моделей процессов предпочтение следует отдавать моделям, учитывающим объективные данные. Субъективные оценки необходимо применять, когда отсутствуют возможности для получения объективных данных. Однако в некоторых задачах учет субъективных вероятностей наряду с объективной информацией бывает весьма полезным. При формировании субъективных вероятностей исследователь должен постараться выразить вероятности рассматриваемых событий через вероятности более простых явлений, которые либо являются заданными, либо поддаются объективной оценке или вычислению.

Общей чертой всех подобных задач является необходимость выбора на основании косвенных или прямых, но обязательно «зашумленных» данных. Основным предположением для формализации решения задач такого типа является предположение о вероятностном характере экспериментальных данных.

Методологической основой для решения задач выбора в такой постановке является теория оптимальных статистических решений. В основе данной теории лежит понятие статистической функции риска.

Рассмотрим постановку задачи. Будем считать заданным вероятностное распределение P на множестве возможных исходов Y, причем значение $P(y)$ определено для каждого исхода y. Предположим, что системный аналитик, не зная результата развития системы, должен принять решение, последствия которого зависят от этого результата (исхода). Пусть X множество всех возможных решений, которые может принять исследователь. Положим, что в результате выбора решения x и реализации исхода y исследователь получает доход r, принадлежащий пространству всех возможных доходов R. Понятие «доход» обычно определяют через полезность, которая служит для численного выражения предпочтений лица, принимающего решения.

Элементы множества R, которые названы доходами, могут быть весьма сложными объектами. Приведем примеры: множество билетов на различные концерты; множество возможных экономических состояний фирмы в определенный момент времени в будущем, измеряемых разностью ее денежных доходов в будущем и в настоящий момент; множество экономических состояний государства и т.д.

Рассмотрим пример, иллюстрирующий понятие дохода в задачах системных исследований. Пусть решается задача выбора варианта реализации структуры сложной системы. Группа проектировщиков должна принять решение о выборе одного варианта реализации системы управления из нескольких рассматриваемых. В качестве множества доходов, получаемых от реализации того или иного варианта, могут служить денежные доходы, получаемые в результате внедрения системы выработки и принятия управляющих воздействий с использованием конкретного варианта структуры системы по сравнению с традиционным способом управления. Под доходом может также пониматься длительность обработки запросов пользователей; длительность обработки результатов функционирования отдельных подразделений, например цехов; точность и достоверность выводов, полученных в результате реализации управляющих воздействий и т.д. Совокупность всех этих составляющих образует вектор дохода конкретного варианта реализации структуры системы.

Для любого множества R у системного аналитика будут предпочтения к тем или иным доходам. В некоторых ситуациях эти предпочтения очевидны. Например, в случае денежных доходов, чем больше доход, тем он предпочтительнее.

Сложнее сформулировать предпочтение в случае, когда доход – векторная величина. При сравнении двух векторных доходов, если каждая компонента первого вектора представляется более желательной, нежели соответствующая компонента второго вектора, то предпочтение отдается первому вектору. Если первый выгоднее только в отношении некоторых своих компонент, в то время как второй вектор предпочтительнее по другим компонентам, то какому из этих векторов отдать предпочтение, не очевидно. Для окончательного решения вопроса следовало бы приписать соответствующие веса отдельным компонентам.

При сравнении двух доходов $r_1 \in R$ и $r_2 \in R$ пишут $r_1 \prec r_2$, если r_2 предпочтительнее r_1, и $r_1 \approx r_2$, если r_1 эквивалентен r_2, т.е. имеет место одинаковая выгодность. Если r_1 не является более предпочтительным, чем r_2, то пишут $r_1 \leq r_2$.

Предполагается, что на основе своих предпочтений среди доходов системный аналитик может задать полное упорядочение множества R. Другими словами налагаются следующие условия.

1. Если r_1 и r_2 – произвольные доходы из множества R, то верно одно и только одно из следующих соотношений:

$$r_1 \prec r_2,\ r_1 \succ r_2,\ r_1 \approx r_2.$$

2. Если r_1, r_2 и r_3 – доходы из R, причем $r_1 \leq r_2$ и $r_2 \leq r_3$, то $r_1 \leq r_3$.

Наконец, будем предполагать, что не все доходы в R эквивалентны между собой, т.е. исключается тривиальная ситуация, а именно, предполагается, что $s_0 \prec t_0$ хотя бы для одной пары $s_0 \in R$, $t_0 \in R$.

В большинстве задач аналитик не вполне свободен в выборе дохода. Обычно он может лишь выбрать из некоторого класса возможных распределений вероятностное распределение на R, согласно которому будет определен его доход.

Например, ставится вопрос о выборе определенного технологического процесса из двух или более возможных таких процессов. Хотя доходы и можно точно выразить через производительность и издержки, производственные характеристики различных процессов могут быть описаны лишь вероятностью.

Другой пример. Системный аналитик хочет получить информацию о значении некоторого параметра. Его доход – это количество информации об этом значении, получаемое после эксперимента. Пусть он выбирает эксперимент из некоторого класса доступных ему, но информация, которую он получит в каждом из экспериментов, носит случайный характер. В любой задаче такого рода аналитик производит выбор не непосредственно среди доходов из множества R, а среди вероятностных распределений на R.

Вероятностное распределение на множестве доходов задают в том случае, когда величина, определяющая доход, имеет характер непрерывной случайной величины. Так, в задаче выбора определенного технологического процесса из нескольких возможных вариантов для каждого процесса могут быть известны средние характеристики. Однако в реальной эксплуатации характеристики технологического процесса могут изменяться в широких пределах. На них оказывает влияние ряд факторов, таких как наличие ресурсов для стабильной работы предприятия, психологический климат в коллективе, состояние здоровья работников, занятых в данном производстве и т.п. Таким образом, доход, получаемый от реализации конкретного варианта технологического процесса, будет величиной случайной, зависящей от большого количества факторов.

То же можно сказать и в случае решения задачи выбора варианта структуры системы управления. Допустим, что проектные документы гарантируют некоторые характеристики качества функционирования системы, такие как быстродействие системы, объем информации, передаваемой по каналам связи, объемы памяти и т.д. Но необходимо отдавать отчет в том, что эти характеристики являются средними. Они могут существенно меняться в зависимости от характеристик конкретных технических средств, например, характеристик надежности. Так, частые отказы одной системы приводят к снижению характеристик качества, в то время как надежная работа другой системы позволяет поддерживать ее характеристики качества на достаточно высоком уровне. В данном случае вектор дохода, получаемого от эксплуатации системы, будет величиной случайной, зависящей от вероятности безотказной работы технических средств (впрочем, как и от ряда других факторов).

Функции полезности

Таким образом, величина дохода, получаемого от реализации того или иного варианта решения, является случайной величиной, зависящей от различных факторов. В случае благоприятного стечения факторов можно получить максимальный доход от выбора определенной альтернативы. В случае неблагоприятного стечения факторов получают минимальный доход. Естественно, что будет некоторое количество промежуточных доходов, зависящих от вероятности реализации того или иного набора факторов.

Поскольку доходы могут иметь различное выражение (денежное выражение дохода, быстродействие системы, объем памяти и т.п.) необходимо обеспечить возможность их сравнения. Как было указано ранее нужно иметь систему предпочтений, систему весовых коэффициентов, с помощью которых можно было бы производить сравнение вариантов реализации различных альтернатив, когда они описываются векторными системами доходов. В качестве такой системы предпочтений выступает функция полезности. Для всякого распределения $P \in \Re$ и всякой вещественной функции q на множестве R обозначим через $E(g/P)$ математическое ожидание функции q (если оно существует) относительно распределения P. Другими словами

$$E(g/p) = \int_R g(r)dP(r).$$

Вещественная функция g, заданная на множестве R, называется функцией полезности, а для любого дохода $r \in R$ число $g(r)$ называется полезностью r. Для всякого распределения $P \in \Re$ число $E(g/P)$ называют полезностью P или средней полезностью.

Определение функции потерь

Рассмотрим теперь пространство X всех возможных решений x, а R пространство всех возможных доходов r, которые может получить исследователь в результате решения x и исхода y. Доход из R, получаемый исследователем при решении x и исходе y обозначим через $\sigma(x, y)$. Будем считать заданным вероятностное распределение P на пространстве исходов Y, причем значение $P(y)$ определено для каждого исхода y. Предположим также, что на множестве R задана функция полезности.

Для всякого вероятностного распределения P_x, для которого функция g интегрируема, среднюю полезность $E(g/P_x)$ можно вычислить по формуле

$$E(g/P_x) = \int_R g(\sigma(x,y))dP(y).$$

Тогда задача исследователя будет состоять в выборе решения x, максимизирующего $E(g/P_x)$.

В задачах принятия решения каждому доходу $r \in R$ принято сопоставлять не полезность, а потери. Функция потерь определяется равенством

$$L(x, y) = -g(\sigma(x, y)).$$

При любом (x,y) число $L(x,y)$ представляет собой ущерб исследователя от принятия решения x, в случае, когда реализовался исход y. Пусть P – вероятностное распределение исхода y. При всяком решении x средний ущерб $\rho(P, x)$ называется риском и определяется по формуле

$$\rho(P,x) = \int_Y L(x,y)dPy.$$

В этом случае системный аналитик должен стремиться к выбору решения x, минимизирующего риск $\rho(P, x)$. Таким образом, сформировано правило выбора решения в случае, когда на пространстве исходов задано распределение вероятностей.

Задачи решения с наблюдениями

Рассмотрим задачи решения, в которых исследователь перед тем как выбрать решение из множества X, наблюдает значение случайной величины или случайного вектора z, связанного с исходом y. Наблюдение z дает исследователю некоторую информацию о значении y, которая помогает ему принять рациональное решение. Будем полагать, что для всех $y \in Y$ задано условное распределение z.

Поскольку решение исследователя зависит от наблюдаемого значения z, он должен выбрать решающую функцию δ, задающую для любого возможного значения $z \in Z$ решение $\delta(z) \in X$. В этом случае функция риска будет определяться равенством

$$\rho(y,\delta) = E\{L[Y,\delta(z)]\} = \int_Y\int_Z L[y,\delta(z)]f(z/y)h(y)dg(z)dp(y).$$

Термин «риск» здесь как и ранее относится к среднему ущербу. Для каждого решения $x \in X$ $\rho(y, \delta)$ обозначает риск от принятия решения x. Распределение $h(y)$ называется априорным, так как оно задает распределение y до проведения наблюдения над z.

Цена наблюдения

В задачах принятия решений наблюдение случайной величины z связано с определенными затратами, которые должны учитываться аналитиками, проводящими системные исследования, при расчете риска от принятия решающей функции, использующей результаты наблюдения z. Это обстоятельство играет особенно важную роль в случае, когда аналитику надо решить, какую из нескольких случайных величин предпочтительнее наблюдать, или ответить на вопрос, производить ли наблюдения вообще. Пусть $c(y, z)$ обозначает цену наблюдения значения z из множества Z. Тогда, если $h(y)$ есть априорная плотность распределения случайной величины y, то средняя цена наблюдения равна

$$E\{c(y,z)\} = \int_Y\int_Z c(y,z)f(z/y)h(y)dg(z)dp(y).$$

Будем предполагать, что для цены $c(y, z)$ верно предположение о средней полезности. Иными словами, будем считать, что эта цена выражена в соответствующих единицах отрицательной полезности так, что существенным является лишь среднее значение вероятностного распределения цены.

Общим риском от наблюдения z и принятия решающей функции δ называется сумма риска $\rho(y, \delta)$ и средней цены наблюдения $E\{c(y, z)\}$. Системный аналитик должен выбрать наблюдение z из некоторого класса доступных наблюдению случайных величин и соответствующую решающую функцию $\delta(z) \in X$, минимизирующую общий риск. Условное распределение y при известном значении z называется апостериорным распределением y, так как оно задает распределение y при зафиксированном значении z.

Сформировав общий риск от наблюдения можно решать задачу о необходимости проведения наблюдений. Если общий риск от наблюдения оказывается меньше, чем риск $\rho(P, x)$, получаемый без проведения дополнительных наблюдений, то наблюдения есть смысл проводить, если же общий риск оказывается больше риска $\rho(P, x)$, то организовывать дополнительные наблюдения смысла не имеет. Таким образом, организовывать наблюдения имеет смысл лишь в том случае, когда цена наблюдения меньше выигрыша, получаемого за счет поступления новой информации.

В рассмотренных задачах делалось предположение о том, что на пространстве решений X и пространстве исходов Y заданы соответствующие распределения. Если вид закона распределения определяющего параметра считается известным, то применение описанных процедур осуществляется согласно приведенным формулам. Если же информации о виде закона распределения у исследователя нет, приходится отказываться от применения параметрического подхода. Здесь важно отметить, что приходится отказываться от необходимости знать вид распределения, а не от того, что выборка подчинена какому-то, пусть неизвестному, но существующему, закону распределения. Предположение о статистичности наблюдений остается в силе. В этом случае для описания распределения на множествах решений и исходов следует применять непараметрические методы.

Незнание функционального вида распределения не означает, что исследователь ничего не может сказать о свойствах распределения. Результаты специально организованных наблюдений, информация, полученная из эксплуатации объекта системного анализа, на этапе его реального функционирования, служит основой для построения непараметрических процедур, решающих задачу выбора. Методы обработки статистических данных с использованием непараметрических процедур рассмотрены ранее [56].

Остановимся на сложностях, которые необходимо осознавать при решении реальных задач выбора, т.е. когда теоретические методы применяются на практике. Неудачное или неправильное применение статистических методов к решению реальных проблем может привести к отрицательному результату. Причины неправильного применения статистических методов известны. Рассмотрим их

1. Статистический вывод по своей природе случаен, поэтому он никогда не может быть абсолютно достоверным. Поэтому при решении задач выбора любая процедура должна сопровождаться оценкой характеристик ее качества. При оценке параметра необходимо вычислять точность, характеризуемую, например, дисперсией. При проверке гипотез необходимо оценивать мощность критерия, с помощью которого осуществляется проверка, вычислять ошибки первого и второго рода. При повышении требований к качеству принимаемых решений необходимо организовать дополнительные исследования объекта системного анализа и тем самым увеличивать объем информации, на основании которой осуществляется принятие решения. Статистический вывод может быть ошибочным, но всегда имеется возможность варьировать характеристики ошибок.

2. Качество решения, принимаемого с помощью процедур статистического вывода, существенно зависит от информации, поступающей на вход. Какие данные в модель заложить, такое решение и получим. В реальной эксплуатации сложных систем встречаются ситуации, когда обслуживающий персонал умышленно скрывает информацию, не записывая все события, происходящие с объектами в оперативные журналы. Например, персонал не заинтересован в ведении журналов учета отказов объектов, так как эффективность функционирования объектов напрямую связана с материальными вознаграждениями персонала. Если они будут записывать все отказы, то это повлечет за собой лишение премий. Естественно, что принятие решений, связанные с планированием деятельности предприятия на основании такой неполной информации будет заведомо содержать ошибку.

3. Отрицательный результат применения теории статистических выводов может быть получен в тех случаях, когда природа явлений, относительно которых принимается решение, не имеет статистического характера. Иными словами встречаются ситуации, когда статистической обработке подвергаются данные, не имеющие статистической природы. Иногда этот факт трудно проверить, особенно при малых объемах выборки. Выяснению факта наличия статистической природы рассматриваемых явлений или процессов следует уделять специальное внимание при организации наблюдений или экспериментов.

4. Снижение качества ожидаемых статистических решений может быть связано с использованием моделей, которые не адекватны описываемым явлениям или процессам. Например, неправомерно применять классические параметрические регрессионные модели в случае, когда ошибка не подчиняется гауссовскому распределению, неправомерно применять модели дисперсионного анализа к негауссовским данным. Часто встречаются ситуации, когда модели, построенные для одних объектов, работающих в условиях воздействия одного комплекса факторов, переносятся на объекты-аналоги, функционирование которых осуществляется при воздействии совершенно другого комплекса факторов. Смена условий функционирования объектов может привести к неадекватности построенной модели.

5. Неудовлетворительный результат применения процедур статистического вывода может иметь место также тогда, когда правильное применение процедуры вывода неверно интерпретируется. Интерпретация статистических результатов лежит вне статистики, за неправильную интерпретацию нельзя осуждать статистику.

В заключение данного параграфа укажем, что в статистических задачах выбора неопределенность бывает двух типов. Первый тип неопределенности связан со стохастической природой явлений и процессов, на основании которых решается задача выбора. Имеется и другая неопределенность, связанная с выбором моделей для описания случайного характера данных, на основании которых осуществляется процедура принятия решений. Например, исследователю заранее неизвестно какое именно распределение из некоторого множества порождало экспериментальные данные. Для решения такого типа задач применяются методы проверки статистических гипотез, которые снижают уровень неопределенности, но полностью ее не устраняют.

13.6. Выбор при нечеткой исходной информации

Идея нечеткого представления информации

Проблемы принятия решений применительно к функционированию сложных систем занимают в настоящее время особое место в информационных технологиях. Математические методы широко применяются для описания и анализа сложных экономических, социальных и других систем. Теория оптимизации создала совокупность методов, помогающих при использовании ЭВМ эффективно принимать решения при известных и фиксированных параметрах. Значительные успехи имеются и

в том случае, когда параметры – случайные величины с известными законами распределения. Трудности возникают, когда параметры систем или ее составных элементов оказываются неопределенными, хотя, может быть, и не случайными, и когда они в то же время сильно влияют на результаты решения.

Специалисты часто сталкиваются с необходимостью расчетов при наличии в уравнениях нечетко заданных параметров или неточной технологической информации. Такого рода ситуации могут возникать как вследствие недостаточной изученности объектов, так и из-за участия в управлении человека или группы лиц. Особенность систем такого рода состоит в том, что значительная часть информации, необходимой для их математического описания, существует в форме представлений или пожеланий экспертов. Но в языке традиционной математики нет объектов, с помощью которых можно было бы достаточно точно отразить нечеткость представлений экспертов.

При построении формальных моделей чаще всего пользуются детерминированными методами и тем самым вносят определенность в те ситуации, где ее в действительности не существует. Неточность задания тех или иных параметров при расчетах практически не принимается во внимание или, с учетом определенных предположений и допущений, неточные параметры заменяются средними или средневзвешенными значениями. Однако обычные количественные методы анализа систем по своей сути мало пригодны и не эффективны для систем, при описании параметров которых используется нечеткая информация. Для систем, сложность которых превосходит некоторый пороговый уровень, точность и практический смысл становятся почти исключающими. Именно в этом смысле точный количественный анализ в реальных экономических, социальных и других систем, связанных с участием человека, не имеет требуемого практического значения.

Иной подход опирается на предпосылку о том, что элементами мышления человека являются не числа, а элементы некоторых нечетких множеств или классов объектов, для которых переход от «принадлежности к классу» к «непринадлежности» не скачкообразен, а непрерывен. Традиционные методы недостаточно пригодны для анализа подобных систем именно потому, что они не в состоянии охватить нечеткость человеческого мышления и поведения.

Теория нечетких (размытых) множеств была впервые предложена американским математиком Лотфи Заде в 1965 г. и предназначалась для преодоления трудностей представления неточных понятий, анализа и моделирования систем, в которых участвует человек.

Для обращения с неточно известными величинами обычно применяется аппарат теории вероятностей. Однако случайность связана с неопределенностью, касающейся принадлежности некоторого объекта к обычному множеству. Это различие между нечеткостью и случайностью приводит к тому, что математические методы нечетких множеств совершенно не похожи на методы теории вероятностей. Они во многих отношениях проще вследствие того, что понятию вероятностной меры в теории вероятностей соответствует более простое понятие функции принадлежности в теории нечетких множеств. По этой причине даже в тех случаях, когда неопределенность в процессе принятия решений может быть представлена вероятностной моделью, обычно удобнее оперировать с ней методами теории нечетких множеств без привлечения аппарата теории вероятностей.

Подход на основе теории нечетких множеств является, по сути дела, альтернативой общепринятым количественным методам анализа систем. Он имеет три основные отличительные черты:

- вместо или в дополнение к числовым переменным используются нечеткие величины и так называемые «лингвистические» переменные;
- простые отношения между переменными описываются с помощью нечетких высказываний;
- сложные отношения описываются нечеткими алгоритмами.

Такой подход дает приближенные, но в то же время эффективные способы описания поведения систем, настолько сложных и плохо определенных, что они не поддаются точному математическому анализу. До работ Л. Заде подобная качественная информация, по существу, просто терялась – было непонятно, как ее использовать в формальных схемах анализа альтернатив. Теоретические же основания данного подхода вполне точны и строги в математическом смысле и не являются сами по себе источником неопределенности. В каждом конкретном случае степень точности решения может быть согласована с требованиями задачи и точностью имеющихся данных. Подобная гибкость составляет одну из важных черт рассматриваемого метода. Для реальных сложных систем характерно наличие одновременно разнородной информации:

- точечных замеров и значений параметров;
- допустимых интервалов их изменения;
- статистических законов распределения для отдельных величин;
- лингвистических критериев и ограничений, полученных от специалистов-экспертов и т.д.

Наличие в сложной многоуровневой иерархической системе управления одновременно различных видов неопределенности делает необходимым использование для принятия решений теории нечетких множеств, которая позволяет адекватно учесть имеющиеся виды неопределенности.

Соответственно и вся информация о режимах функционирования подсистем, областях допустимости и эффективности, целевых функциях, предпочтительности одних режимов работы перед другими, о риске работы на каждом из режимов для подсистем и т.д. должна быть преобразована к единой форме и представлена в виде функций принадлежности. Такой подход позволяет свести воедино всю имеющуюся неоднородную информацию: детерминированную, статистическую, лингвистическую и интервальную.

Разработанные в настоящее время количественные методы принятия решений помогают выбирать наилучшие из множества возможных решений лишь в условиях одного конкретного вида неопределенности или в условиях полной определенности. К тому же, большая часть существующих методов для облегчения количественного исследования в рамках конкретных задач принятия решений базируется на крайне упрощенных моделях действительности и излишне жестких ограничениях, что уменьшает ценность результатов исследований и часто приводит к неверным решениям. Применение для оперирования с неопределенными величинами аппарата теории вероятности приводит к тому, что фактически неопределенность, независимо от ее природы, отождествляется со случайностью, между тем как основным источником неопределенности во многих процессах принятия решений является нечеткость или расплывчатость.

В отличие от случайности, которая связана с неопределенностью, касающейся принадлежности или непринадлежности некоторого объекта к нерасплывчатому множеству, понятие «нечеткость» относится к классам, в которых могут быть различные градации степени принадлежности, промежуточные между полной принадлежностью и непринадлежностью объектов к данному классу.

Вопрос выбора адекватного формального языка является очень важным, поэтому следует отметить преимущества описания процесса принятия решений в сложной многоуровневой иерархической системе на основе теории нечетких множеств. Этот язык дает возможность адекватно отразить сущность самого процесса принятия решений в нечетких условиях для многоуровневой системы, оперировать с нечеткими ограничениями и целями, а также задавать их с помощью лингвистических переменных. Поэтому математический аппарат теории нечетких множеств принят в настоящее время как основной аппарат описания многоуровневых иерархических систем и процессов принятия решений в сложных системах.

Одним из важных направлений применения этого нового подхода является проблема принятия решений при нечеткой исходной информации. Здесь появляется возможность сузить множество рассматриваемых вариантов или альтернатив, отбросив те из них, для которых имеются заведомо более приемлемые варианты или альтернативы, подобно тому, как это делается при использовании принципа Парето.

Терминология теории нечетких множеств

В традиционной прикладной математике множество понимается как совокупность элементов (объектов), обладающих некоторым общим свойством. Например, множество чисел, не меньших заданного числа, множество векторов, сумма компонент каждого из которых не превосходит единицы, и т.п. Для любого элемента при этом рассматриваются лишь две возможности: либо этот элемент принадлежит данному множеству (т.е. обладает данным свойством), либо не принадлежит (т.е. не обладает данным свойством). Таким образом, в описании множества в обычном смысле должен содержаться четкий критерий, позволяющий судить о принадлежности или непринадлежности любого элемента данному множеству.

Понятие нечеткого множества – попытка математической формализации нечеткой информации с целью ее использования при построении математических моделей сложных систем. В основе этого понятия лежит представление о том, что составляющие данное множество элементы, обладающие общим свойством, могут обладать этим свойством в различной степени и, следовательно, принадлежать данному множеству с различной степенью. При таком подходе высказывания типа «элемент принадлежит данному множеству» теряют смысл, поскольку необходимо указать «насколько сильно» или с какой степенью рассматриваемый элемент принадлежит данному множеству.

Один из простейших способов математического описания нечеткого множества – характеризация степени принадлежности элемента множеству чисел, например, из интервала $[0, 1]$. Пусть X – некоторое множество элементов (в обычном смысле). Нечетким множеством C в X называется совокупность пар вида $(x, \mu_c(x))$, где $x \in X$, а $\mu_c(x)$ – функция, называемая функцией принадлежности нечеткого множества C, данная функция принимает значения из интервала $[0,1]$. ***Функцией принадлежности*** называется функция, которая позволяет вычислить

степень принадлежности произвольного элемента универсального множества к нечеткому множеству. Значение $\mu_c(x)$ этой функции для конкретного x называется степенью принадлежности этого элемента нечеткому множеству C. Как видно из этого определения, нечеткое множество полностью описывается своей функцией принадлежности.

Обычные множества составляют подкласс класса нечетких множеств. Действительно, функцией принадлежности обычного множества $B \subset X$ является его характеристическая функция

$$\mu_B(x) = \begin{cases} 1, & x \in B; \\ 0, & x \notin B. \end{cases}$$

В соответствии с определением нечеткого множества обычное множество B можно также определить как совокупность пар вида $(x, \mu_B(x))$. Таким образом, нечеткое множество представляет собой более широкое понятие, чем обычное множество, в том смысле, что функция принадлежности нечеткого множества может быть, вообще говоря, произвольной функцией или даже произвольным отображением.

Задачи достижения нечетко определенной цели

Рассмотрим подход к решению задач выбора, в которых находит применение теория нечетких множеств. Основным предположением в данном подходе является допущение о том, что цели принятия решений и множество альтернатив рассматриваются как равноправные нечеткие подмножества некоторого универсального множества альтернатив. Это позволяет определить решение задачи в относительно простой форме.

Пусть X – универсальное множество альтернатив, т.е. универсальная совокупность всевозможных выборов лица, принимающего решения. Нечеткой целью в X является нечеткое подмножество, которое будем обозначать G. Описывается нечеткая цель функцией принадлежности $\mu_G : X \to [0, 1]$. Допустим, что X представляет собой числовую ось. Тогда нечеткой целью принятия решений может быть нечеткое множество типа «величина x должна быть примерно равна 5» или «желательно, чтобы величина x была значительно больше 10» и т.п. Будем полагать, что присутствующие в подобных описаниях нечеткие понятия вполне точно описаны функциями принадлежности соответствующих нечетких множеств.

Чем больше степень принадлежности альтернативы x нечеткому множеству целей, т.е. чем больше значение $\mu_G(x)$, тем больше степень достижения этой цели при выборе альтернативы x в качестве решения.

Нечеткие ограничения или множества допустимых альтернатив также описываются нечеткими подмножествами множества X. В приведенном примере, когда x элемент числовой оси, нечеткие ограничения могут иметь, например, такой вид «x не должно быть много больше 30», «x должно быть не слишком большим» и т.п. Как и прежде, здесь полагается, что приведенные в качестве примера понятия описаны функциями принадлежности соответствующих нечетких множеств, которые будем обозначать $\mu_c(x)$.

Более общей является постановка задачи, в которой нечеткие цели и ограничения представляют собой подмножества различных множеств. Пусть, как и выше, X – множество альтернатив и пусть задано однозначное отображение φ: $X \to Y$, под элементами множества Y будем понимать оценки показателей качества или эффективности системы. Нечеткая цель при этом будет задаваться в виде нечеткого подмножества множества оценок Y, т.е. в виде функции μ_G: $Y \to [0, 1]$.

Задача при этом сводится к прежней постановке, т.е. к случаю, когда цель – нечеткое подмножество X, с использованием следующего приема. Определим нечеткое множество альтернатив $\overline{\mu_G}$, обеспечивающих достижение заданной цели μ_G. Это множество представляет собой прообраз нечеткого множества μ_G при отображении φ, т.е.

$$\overline{\mu_G}(x) = \mu_G(\varphi(x)),\ x \in X.$$

После этого исходная задача рассматривается как задача достижения нечеткой цели $\overline{\mu_G}$ при заданных нечетких ограничениях.

Перейдем теперь к определению решения задачи достижения нечеткой цели. Грубо говоря, решить задачу, означает достичь цели и удовлетворить ограничениям, причем в данной нечеткой постановке следует говорить не просто о достижении цели, а о ее достижении с той или иной степенью, также следует учитывать и степень выполнения ограничений. В излагаемом подходе оба этих фактора учитываются следующим образом. Пусть некоторая альтернатива x обеспечивает достижение цели (или соответствует цели) со степенью $\mu_G(x)$, и удовлетворяет ограничениям (или является доступной) со степенью $\mu_c(x)$. Тогда полагается, что степень принадлежности этой альтернативы решению задачи равна минимальному из этих чисел. Иными словами, альтернатива, допустимая со степенью, например 0,3, с той же степенью принадлежит нечеткому решению, несмотря на то, что она обеспечивает достижение цели со степенью, равной, например, 0,8.

Таким образом, нечетким решением задачи достижения нечеткой цели называется пересечение нечетких множеств цели и ограничений, т.е. функция принадлежности решений $\mu_D(x)$ имеет вид

$$\mu_D(x) = \min\{\mu_G(x), \mu_C(x)\}.$$

При наличии нескольких целей и нескольких ограничений нечеткое решение описывается функцией принадлежности:

$$\mu_D(x) = \min\{\mu_{G1}(x), \ldots, \mu_{Gn}, \mu_{C1}(x), \ldots, \mu_{Cm}\}.$$

Если различные цели и ограничения отличаются по важности и заданы соответствующие коэффициенты относительной важности целей λ_i, и ограничений ν_j, то функция принадлежности решения задачи определяется выражением

$$\mu_c(x) = \min\{\lambda_1\mu_{G1}(x), \ldots, \lambda_n\mu_{Gn}, \nu_1\mu_{C1}(x), \ldots, \nu_m\mu_{Cm}\}.$$

Один из наиболее распространенных в литературе способов решения задач выбора при нечеткой исходной информации состоит в выборе альтернативы, имеющей максимальную степень принадлежности нечеткому решению, т.е. альтернативы, реализующей правило

$$\max_{x\in X} \mu_D(x) = \max_{x\in X} \min\{\mu_G(x), \mu_C(x)\}.$$

Следует подчеркнуть, что техника, развиваемая в работах Л. Заде и его последователей, основывается на использовании функций принадлежности. Эти функции всегда являются гипотезами! Они дают субъективное представление исследователя об особенностях анализируемой операции, о характере ограничений и целей исследования. Это всего лишь новая форма утверждения гипотез, которая открывает и новые возможности.

В заключение данного параграфа следует отметить, что различие между нечеткостью и случайностью приводит к тому, что математические методы нечетких множеств совершенно не похожи на методы теории вероятностей. Они во многих отношениях проще вследствие того, что понятию вероятностной меры в теории вероятностей соответствует более простое понятие функции принадлежности в теории нечетких множеств. По этой причине даже в тех случаях, когда неопределенность в процессе принятия решений может быть представлена вероятностной моделью, обычно удобнее оперировать с ней методами теории нечетких множеств без привлечения аппарата теории вероятностей.

Получение во всех этих моделях решений в нечеткой форме позволяет довести до сведения специалиста, принимающего решение, что если он согласен или вынужден довольствоваться нечеткой формулировкой проблемы и нечеткими сведениями о модели, то он должен быть удовлетворен и нечетким решением задачи.

13.7. Проблема оптимизации и экспертные методы принятия решений

Рассмотренные до настоящего времени задачи выбора заключались в том, чтобы в исходном множестве альтернатив найти оптимальные. Задача нахождения оптимальной альтернативы заключается в поиске экстремума заданного критерия эффективности. То есть считается, что исследователем сформирован критерий, который выступает в качестве способа сравнения вариантов решения. Предполагается также, что наряду с критерием имеются ограничения, которые также оказывают влияние на результат выбора. Причем следует иметь в виду, что при изменении ограничений при одном и том же критерии результат выбора может оказаться другой.

Идея оптимальности является центральной идеей кибернетики. Понятие оптимальности вошло в практику проектирования и эксплуатации сложных технических систем, получило строгое и точное представление в математических теориях, широко используется в административной практике. Данное понятие сыграло важную роль в формировании системных представлений. Осознавая ведущую роль оптимизационного подхода при решении задач выбора, следует остановиться на ряде ограничений, которые необходимо осознавать при применении данного подхода. Охарактеризуем их.

1. Оптимальное решение часто оказывается чувствительным к незначительным изменениям в условиях задачи. В результате изменения условий или предположений, при которых формировалась модель задачи принятия решений, могут получиться выводы, существенно отличающиеся друг от друга. В связи с этим в теории оптимальности развивается такое направление как исследование устойчивости решения, а также анализ результатов решения на чувствительность к изменению входных параметров, условий и предположений.

2. При решении практических задач оптимизации следует учитывать, что анализируемая система имеет взаимосвязи с другими системами, а зачастую она является подсистемой какой-либо гиперсистемы. В связи с этим требуется увязывать цели анализируемой системы с целями других систем и в особенности с глобальными целями гиперсистемы. В этом случае постановка задачи оптимизации для анализируемой системы может иметь подчиненное значение по отношению к постановкам задач для других систем. Тогда задача сведется к задаче локальной оптимизации. В этом случае локальная оптимизация может привести к результату, отличающемуся от того, который потребуется от системы при оптимизации целевых функций гиперсистемы. Отсюда следует вывод, что необходимо увязывать критерии анализируемой системы с критериями других систем и, в особенности, гиперсистемы.

3. При использовании оптимизационного подхода не следует отождествлять цели системы и критерии, с помощью которых решается задача выбора. Критерий и цель относятся друг к другу как модель и оригинал. Многие цели трудно или даже невозможно количественно описать. Количественный критерий является лишь приближением цели. Критерий характеризует цель лишь косвенно, иногда лучше, иногда хуже, но всегда приближенно.

4. В постановке задачи оптимизации наряду с критериями не менее важную роль играют ограничения. Даже небольшие изменения ограничений существенно сказываются на результате решения. Еще более разительный эффект можно получить, исключая одни ограничения и добавляя другие. Отсюда требуется сделать вывод о необходимости тщательного анализа всех условий, при которых решается задача выбора. Если при постановке задачи не проведен должным образом анализ условий и в результате не сформирован в полном объеме набор ограничений, это может наряду с оптимизацией критерия привести к непредвиденным сопутствующим эффектам.

Подводя итог сказанному, можно сформулировать отношение к идее оптимизации с позиций системного анализа. Оно состоит в следующем: оптимизация – это мощное средство повышения эффективности, но использовать его следует все более осторожно по мере возрастания сложности проблемы. Многие задачи системных исследований могут быть достаточно хорошо формализованы, сведены к математическим моделям, позволяющим ставить и решать оптимизационные задачи. Однако даже после преодоления сложностей формализации системотехнических проблем остаются некоторые особенности, которые сказываются на результате решения. А именно, это неустойчивость оптимальных решений, сильная чувствительность к изменению условий, и неоднозначность постановки многокритериальных задач. Меры преодоления данных обстоятельств состоят в проведении анализа решения на чувствительность, всяческое использование априорной информации с целью повышения уровня достоверности моделей, рассмотрение оптимальных альтернатив по нескольким различным сверткам критериев.

При исследовании социотехнических систем, когда необходимо помимо чисто технических вопросов решать организационные и социальные проблемы, ситуация существенно усложняется. Учет подобного типа вопросов не поддается полной формализации. Следовательно, оптимизационные задачи, которые удается поставить при исследовании сложных систем, неизбежно являются заведомо приближенными, если относятся к системе в целом, либо имеют частичный, подчиненный характер, если описывают хорошо структурированные подсистемы. Ввиду этого оптимизация в системных исследованиях не конечная цель, а промежуточный этап работы. Чем сложнее система, тем осторожнее следует относиться к ее оптимизации. При исследовании сложных систем неизбежно возникают проблемы, выходящие за пределы формальных математических постановок задач. В ряде случаев, по мере необходимости обращаются к услугам экспертов, т.е. лиц, чьи суждения, опыт и интуиция могут помочь в решении проблемной ситуации.

Основная идея экспертных методов состоит в том, чтобы использовать интеллект людей, их способность искать решение слабо формализованных задач. При организации работы группы экспертов необходимо учитывать, что интеллектуальная деятельность людей во многом зависит от внешних и внутренних условий. Поэтому в методиках организации экспертиз и проведении экспертных оценок специальное внимание уделяется созданию благоприятных условий и нейтрализации факторов, неблагоприятно влияющих на работу экспертов.

Важную роль в организации работы экспертов играют факторы психологического характера. Прежде всего, эксперты должны быть освобождены от ответственности за использование результатов экспертизы. Дело не только в том, что лицо, принимающее решения, не должно возлагать ответственности на других, но и в том, что сама ответственность накладывает психологические ограничения на характер выбора, а этого на стадии оценки альтернатив желательно избегать. Следует также принимать во внимание, что решение, принимаемое экспертом, может зависеть от межличностных отношений с другими экспертами, а также от того, известна ли его оценка другим лицам. На ход экспертизы могут повлиять и такие факторы, как личная заинтересованность эксперта, его необъективность, личностные качества. С другой стороны, сложность проблем, решаемых в задачах системных исследований, обычно выходит за рамки возможностей одного человека. В этих условиях коллективная деятельность открывает дополнительные возможности для взаимного стимулирования экспертов.

Поскольку взаимодействие между экспертами может как стимулировать, так и отрицательно сказываться на их деятельности, в разных

случаях используют методики проведения экспертиз, имеющие различные степень и характер взаимного влияния экспертов друг на друга. Известны следующие методы проведения экспертиз: анонимные и открытые опросы и анкетирование, совещания, дискуссии, деловые игры, мозговой штурм и т.п.

В последнее время с целью оказания помощи эксперту в принятии решения развиваются человеко-машинные системы, так называемые системы «искусственного интеллекта». Развитие систем такого типа идет по нескольким направлениям, а именно, разрабатываются базы знаний и экспертные системы и системы поддержки принятия решений. В системах такого типа лицу, принимающему решение, предоставляется помощь в поиске наилучшего решения. Математическое и программное обеспечение таких систем строится на базе набора формализованных процедур, которые лицо, принимающее решение, может использовать в любой момент и в любой степени.

13.8. Коллективный или групповой выбор

В ходе решения задач системного анализа единоличное принятие решения является скорее исключением, чем правилом. Более реальна ситуация, когда решение принимается группой лиц. Причем интересы отдельных личностей в данной группе могут полностью совпадать (кооперативный выбор), быть противоположными (конфликтная ситуация), и могут иметь место промежуточные случаи, создаваться коалиции, достигаться компромиссы в процессе переговоров и т.п. При групповом выборе решений определяющую роль играет проблема согласования индивидуальных предпочтений лиц, участвующих в процессе их принятия. В данной ситуации ставится задача выработки некоторого решения, которое согласует индивидуальные выборы, выражает в каком-то смысле общее мнение и принимается за групповой выбор. Вполне естественно, что данное решение должно быть функцией индивидуальных выборов. Причем различным принципам согласования будут соответствовать совершенно различные функции. Теоретически данные функции могут быть произвольными, учитывать не только индивидуальные выборы, но и другие факторы, в том числе и случайные события. Главный вопрос состоит в том, чтобы правильно отобразить в данной функции особенности конкретного варианта реального группового выбора.

При принятии решения коллективом участников, особенно в ситуации переговоров и посредничества, лица, принимающие решение, редко обладают одинаковой исходной информацией, придерживаются одних и тех же ценностных концепций. Однако даже самая разнородная группа может прийти к соглашению, если ее члены с уважением относятся к многообразию точек зрения, склонны учиться друг у друга, обмениваться информацией. Важным условием успешного принятия консолидированного решения является согласованное принятие некоторых процедур переговоров и условий посредничества. Остановимся на основных принципах выработки и принятия решений в условиях коллективного выбора.

В первую очередь следует остановиться на таком аспекте как роль системы ценностей в анализе решений. Известно, что системы ценностей могут быть различными. Наиболее известными и имеющими широкое распространение в практической деятельности являются две крайние системы, называемые технократическая и гуманистическая. В зависимости от того, какого из ценностных критериев придерживается лицо, принимающее решение, будут сформированы гипотезы и концепции, закладываемые в основу процедур принятия решений. С другой стороны, существуют ценности, универсальные для всего человечества: глобальная ответственность, терпимость, стремление к истине и познанию и т.д. Поэтому при принятии коллективного решения следует постараться воспринять возможные неоднозначные представления о рациональности, уметь выслушать противоположную сторону, а не подвергать позицию оппонента критике и сравнительному анализу. Таким образом, во главу угла ставится проблема рациональности принимаемого решения.

Рассмотрим некоторые модели рациональности. При классификации различных подходов к рациональному принятию решений необходимо, прежде всего, различать целостный и аналитический подходы. Целостная схема принятия решений использует умение воспринимать явление в целом, не выделяя составные части или информацию. Даже если для дальнейшего анализа такое выделение необходимо, оно производится только после того, как явление распознано целиком. Рациональность таких решений может быть подвергнута сомнению, поскольку различные эвристики и интуиция играют здесь определяющую роль. Это ставит вопрос о том, как надо понимать саму концепцию рациональности. Любая общая концепция может быть сначала сужена и в этом урезанном виде значительно усовершенствована с помощью абстрактных построений и математической теории, исследующей лишь определенные аспекты концепции. Однако такое сужение и частичное развитие концепций может нанести значительный ущерб прикладным исследованиям, поэтому предпочтительнее использовать термин «рациональность» в его первоначальном, более широком смысле. Рациональное решение вовсе не должно использовать всю имеющуюся информацию, оно не обязано быть оптимальным, оно должно только учитывать возможные последствия и не причинять ущерба интересам лица, принимающего решение, хотя реальные результаты в коллективе могут быть и нежелательными. В качестве разумного компромисса можно говорить о различных степенях рациональности: о суперрациональности (или возможности разрешить известные парадоксы рациональности), об оптимизационной рациональности, о приемлемой рациональности, процедурной рациональности и т.д. Если следовать такому широкому пониманию вопроса, то адаптивно формируемое решающее правило может приводить к вполне рациональным решениям, а изучение эффективности различных решающих правил и выбор одного из них представляют собой весьма перспективную задачу. Более того, можно утверждать, что большинство повседневно принимаемых решений связано именно с целостным подходом, и он часто оказывается предпочтительным при долгосрочной перспективе.

Однако решения, принимаемые при недостатке информации и изменяющихся условиях, часто требуют аналитического подхода, т.е. систематической оценки возможных альтернатив и соответствующих исходов, а затем выбора одной из них. Известен целый ряд аналитических моделей принятия решений. Наиболее широко употребимой является модель максимизации полезности. Заслуживает упоминания также программно-целевой подход, разработанный В.М. Глушковым, Г.С. Поспеловым, В.А. Ириковым и др. и опирающийся на реальные процессы принятия плановых решений.

Модель максимизации полезности наиболее сильно развита теоретически, имеет подробное математическое обоснование и поэтому воспринимается повсюду как разумная схема аналитического принятия решений. Однако как в теоретическом, так и в эмпирическом плане эта схема приводит к парадоксам.

Некоторые из них означают, что стратегия максимизации полезности не гарантирует рационального поведения в игровых моделях с ненулевой суммой. Следует также отметить, что одна из основополагающих аксиом теории полезности – аксиома независимости от непричастных альтернатив – не подтверждается экспериментальными данными и опровергается более глубоким анализом.

Модель приемлемых решений возникла в результате критики оптимизационного подхода. Реальная практика принятия решений такова: руководители больших организаций, различных институтов, инженеры, проектирующие новые технические устройства, и даже обычные потребители на рынках никогда не прибегают к полной оптимизации из-за нехватки информации и времени. Вместо этого они адаптивно, в процессе обучения, формируют уровни достижимости, которые должны обеспечиваться удовлетворительными, приемлемыми решениями.

В качестве более подходящего описания процесса принятия решений можно принять модель квазиприемлемого поведения, в которой лицо, принимающее решение, проявляет тенденцию к оптимизации, но может в силу ряда причин отказаться от оптимизации, обеспечив себе адаптивно формируемые уровни достижимости.

В программно-целевой модели предполагается, что некоторые цели или программы (фактически уровни достижимости) имеют больший приоритет и должны быть реализованы, а задача состоит в том, чтобы распределить или увеличить ресурсы, преодолеть возможные препятствия и изменить другие уровни достижимости, с тем, чтобы обеспечить реализацию приоритетных программ. Соответствующая математическая модель используется многими исследовательскими группами в разных странах в качестве схемы для описания рационального, целенаправленного поведения. С формальной точки зрения эта схема не противоречит идее максимизации полезности, поскольку приоритетные цели всегда можно использовать в качестве ограничений и максимизировать полезность на множестве допустимых распределений ресурсов. Но по существу данная модель представляет принципиально отличную методологию, которая ближе к модели приемлемых решений. Например, соответствующая модификация целевого программирования может удовлетворительно моделировать программно-целевые действия. Но эта же схема может описываться и моделью квазиприемлемого поведения: достаточно предположить, что некоторые уровни достижимости могут быть более приоритетными и менее изменяемыми.

Рассмотрим постулаты многосторонней рациональности. Если даже формальные схемы принятия решений отражают различные методологические представления о рациональности, то как можно объяснять достижение соглашений при принятии решений в условиях конфликта интересов? Очевидно, должны быть веские причины для согласования интересов. Сформулируем их в виде постулатов многосторонней рациональности, которые следует учитывать при построении интерактивных систем принятия решения.

1. Постулат ограниченной неосведомленности и взаимного обучения. При анализе или обсуждении решений не следует предполагать наличие полной информации или рациональных прогнозов; напротив, необходимо признать собственную (возможно, ограниченную) неосведомленность и быть готовым к взаимному обучению, чтобы установить общую, приемлемую для всех информационную основу. Любая формализация процессов принятия решений при наличии многосторонней рациональности должна учитывать аспект взаимного обучения.

2. Постулат уважения к чужому мнению. Обучение при многосторонней рациональности должно базироваться на уважении к культурным ценностям и представлениям о рациональности, существующим у других участников процесса. В частности, формализация процессов принятия решений в этих ситуациях должна допускать параллельную интерпретацию, предусматривающую наличие разных представлений о рациональности. Не следует принимать свое представление о рациональности как единственно правильное.

3. Постулат законного протокола. При наличии многосторонней рациональности необходимым условием получения взаимно приемлемых решений является соглашение о правилах поведения в данной ситуации. Если, например, одна из сторон в двусторонних переговорах настаивает на сохранении за ней ведущей позиции, а вторая сторона с этим не соглашается, то шансов на принятие взаимно приемлемых решений нет. Таким образом, организационную структуру процесса коллективного принятия решений не следует принимать как данную: ее необходимо предварительно обсуждать и согласовывать.

4. Постулат справедливого посредничества. Если при наличии многосторонней рациональности имеется посредник или же используется какой-либо механизм посредничества, то следует тщательно обсудить и согласовать принципы и условия такого посредничества.

Рассмотрим теперь способы формирования функции, принимаемой за групповой выбор. Один из наиболее распространенных принципов согласования – правило большинства: принятой всеми считается альтернатива, получившая наибольшее число голосов. Правило большинства привлекательно своей простотой и демократичностью, но имеет особенности, требующие осторожного обращения с ним. Прежде всего, оно лишь обобщает индивидуальные предпочтения, и его результат не является критерием истины. Только дальнейшая практика показывает, правильным или ошибочным было решение, принятое большинством голосов; само голосование – лишь форма согласования дальнейших действий. Во-вторых, даже в простейшем случае выбора одной из двух альтернатив может возникнуть ситуация, когда правило большинства не срабатывает, например, происходит разделение голосов поровну при четном числе голосующих. Это порождает варианты: «председатель имеет два голоса», «большинство простое (51%)», «подавляющее большинство (около 3/4)», «абсолютное большинство (близкое к 100%)», наконец, «принцип единогласия (консенсус, право вето)».

При любом из этих вариантов подразумевается отказ от принятия решения, если ни одна из альтернатив не получила необходимого процента голосов. Поскольку в реальной жизни отказ от дальнейших действий, следующих за решением, бывает недопустим, а переход к принятию за групповой выбор выбора отдельного лица (диктатора) нежелательным, разрабатываются различные приемы, сокращающие число ситуаций, приводящих к отказу от дальнейших действий.

ЗАКЛЮЧЕНИЕ

Системный анализ как наука проходит этап становления. В настоящее время отсутствуют работы, в которых были бы проанализированы и обобщены подходы и концепции разных авторов по процедурам проведения системных исследований, дано обобщенное представление о развиваемых методах и применяемых моделях, даны рекомендации по выбору подхода к организации и проведения исследований в конкретных условиях.

Характеризуя современное состояние системных исследований, следует отметить, что они включают три вида деятельности:

- научное исследование вопросов, связанных с проблемой;
- формирование способов разрешения проблемной ситуации;
- внедрение в практику результатов, полученных в ходе исследований.

В системном анализе находят органичное объединение теория и практика, наука и искусство, творческий подход и алгоритмичность действий, формализация и эвристика. В конкретном исследовании соотношение между этими компонентами может быть самым различным. Системный аналитик должен быть готов привлечь к разрешению проблемной ситуации любые необходимые знания и методы. В случае, если он сам не владеет какими-то процедурами, он должен найти исполнителя и выступать в качестве организатора исследования, носителя цели и методологии выполнения всех работ.

Предметом системных исследований являются три типа систем: технические, природные и социотехнические. Наибольшую трудность для анализа представляют собой системы третьего типа. Это обусловлено преобладанием в них субъективного над объективным, эвристического над формальным. Важным отличием данных систем от систем других типов является ярковыраженная зависимость их от времени, подверженность влиянию под воздействием процедур проводимого анализа, их видоизменение в процессе проведения исследований. Процедуры, лежащие в основе системного анализа, помогают создать динамическую модель системы и с ее помощью спланировать и организовать действия всех участников анализа, направленные на достижение конечной цели исследования.

Системный анализ – дисциплина синтетическая. Она использует методы, модели и результаты различных теоретических курсов. Но, несмотря на использование в своем арсенале самых современных методических разработок имеется ряд проблем, которые до сих пор не нашли своего разрешения. Это является следствием того, что объектом системных исследований являются сложные системы. Остановимся на некоторых направлениях, которым уделяется повышенное внимание исследователей, и которые требуют дальнейшей проработки.

Во-первых, отметим такую проблему как построение динамических моделей сложных систем. Необходимость учета особенностей функционирования сложных систем требует привлечения самых современных достижений математических дисциплин. Например, для того чтобы учесть профилактическое обслуживание и возможность восстановления работоспособности системы, приходится привлекать методы теории восстановления и решать интегральные уравнения Вольтерра 2-го рода. Для описания использования запасных частей в работе отдельных элементов и подсистем применяются методы теории массового обслуживания. Следует также отметить, что изменение предположений о характере происходящих процессов в системе может привести к изменению математического аппарата, привлекаемого к формированию модели.

Второе направление связано с обработкой информации и подготовкой исходных данных и параметров разрабатываемых моделей. В качестве исходной информации для формирования входных параметров модели используется информация, получаемая в процессе функционирования систем. Эти данные, как правило, представляют собой либо данные статистической природы, либо нечисловой природы. Последнее особенно характерно для социотехнических систем. В итоге возникает необходимость в разработке методов обработки результатов наблюдений указанного типа. Для анализа данных нечисловой природы используются процедуры теории нечетких множеств. Для анализа статистических данных находят применение робастные методы и непараметрические методы оценивания.

В качестве еще одного направления исследования, требующего внимания системных аналитиков, отметим разработку эвристических процедур на этапе анализа и прогнозирования развития проблемной ситуации или поведения системы под воздействием самой процедуры системного анализа.

Отмеченные направления не охватывают всех проблемных вопросов, но на взгляд автора являются наиболее актуальными на сегодняшний день.

Подводя окончательный итог, отметим, что с практической стороны системный анализ есть теория и практика улучшающего вмешательства в проблемную ситуацию, с методологической стороны он имеет прикладную направленность, ориентированную на изменение окружающей действительности.

ЛИТЕРАТУРА

1. *Перегудов Ф.И., Тарасенко Ф.П.* Введение в системный анализ. – М.: Высшая школа,1989. – 367 с.
2. *Альтшулер Г.С.* Алгоритм изобретения. – М.: Московский рабочий, 1973.
3. *Зарипов Р.Х.* Машинный поиск вариантов при моделировании творческого процесса. – М.: Наука, 1983.
4. *Богданов А.А.* Всеобщая организационная наука (тектология) В 3-х т. – М., 1905-1924. Т.3.
5. *Bertalanffy L.* An Outline of General System Theory//British J. For Phil. of Sci. – 1950. – V. 1. – № 2. – P. 134 – 165.
6. *Винер Н.* Кибернетика. – М.: Сов.радио, 1968.
7. *Винер Н.* Кибернетика и общество. – М.: ИЛ,1958.
8. *Пригожин И., Стенгерс И.* Порядок из хаоса. – М.: Прогресс, 1986.
9. *Моисеев Н.Н.* Математические задачи системного анализа. – М.: Наука. 1981. – 488 с.
10. *Дегтярев Ю.И.* Системный анализ и исследование операций. – М.: Высшая школа, 1996. – 335 с.
11. *Антонов А.В.* Проектирование систем. –Обнинск: ИАТЭ, 1996. – 157 с.
12. Методологические проблемы кибернетики: В 2-х т. – М.: МГУ, 1970.
13. *Саати Т.* Математические методы исследования операций. – М.: МО, 1963.
14. Исследование операций. Методологические основы и математические методы. Пер. с англ. /*Под ред. Дж. Моудера, С. Элмаграби.* – М.: Мир, 1981. – 712 с.
15. Исследование операций. Модели и применения./*Пер. с англ. под ред. Дж. Моудера, С. Элмаграби.* – М.: Мир, 1981. – 677 с.
16. *Острейковский В.А.* Теория систем. – М.: Высшая школа, 1997. – 240 с.
17. *Волкова В.Н., Денисов А.А.* Основы теории систем и системного анализа. – СПб.: Изд-во СПбГТУ, 1997. – 510 с.
18. *Зинченко В.П.* Человеческий интеллект и технократическое мышление.//Коммунист. – 1988. – № 3. – С. 96 – 104.
19. *Холл А.* Опыт методологии для системотехники. – М.: Сов. Радио, 1975.
20. *Мамиконов А.Г.* Проектирование АСУ. – М.: Высшая школа, 1987. – 303 с.
21. *Максимей И.В.* Имитационное моделирование на ЭВМ. – М.: Радио и связь, 1988. – 232 с.
22. *Барзилович Е.Ю., Каштанов В.А.* Некоторые математические вопросы теории обслуживания сложных систем. – М.: Радио и связь, 1971.
23. *Ackoff R.I.* The mismatch between educational systems and requirements for successful management. // Wharton Alumni Magazine. – Spring,1986. – P. 10 – 12.

24. *Checkland P.* Rethinking a System Approach. In: Tomlison R., Kiss I. (Eds.) «Rethinking the Process of Operation and System Analysis». – Pergamon Press, 1984. – P 43 – 66.
25. *Gharajedaghi J., Ackoff R.L.* Toward Systemic Education of System Scientists./ /Systems Reseach. – 1985. – V. 2. – № 1. – P. 21 – 27.
26. Электронные вычислительные машины: В 8 кн. Кн. 8. Решение прикладных задач: Практ. пособие для вузов/*А.Г. Дьячко, Н.М. Когдов; Под ред. А.Я. Савельева.* – М.: Высшая школа, 1993. – 158 с.
27. *Вентцель Е.С., Овчаров Л.А.* Теория случайных процессов и ее инженерные приложения. – М.: Наука, 1991. – 384 с.
28. *Левин В.И.* Логическая теория надежности сложных систем. – М.: Энергоатомиздат, 1985. – 129 с.
29. Надежность и эффективность в технике: Справочник: В 10 т. Т. 4: Методы подобия в надежности/*Под общ. Ред. В.А. Мельникова, Н.А. Северцева* – М.: Машиностроение, 1987. – 280 с.
30. *Антонов А.В., Острейковский В.А.* Оценивание характеристик надежности элементов и систем ЯЭУ комбинированными методами. – М.: Энергоатомиздат, 1993. – 368 с.
31. *Антонов А.В.* Об одном методе проверки однородности информации в случае параметрического оценивания характеристик надежности.//Надежность и контроль качества. – 1993 – № 10. – С. 20 – 32.
32. *Крамер Г.* Математические методы статистики. – М.: Мир, 1975
33. *Египко В.М.* Организация и проектирование систем автоматизации научно-технических экспериментов. – Киев: Наукова думка, 1978.
34. *Заде Л.* Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976. – 168 с.
35. *Антонов А.В., Чепурко В.А.* Планирование эксперимента: Учеб. пособие. – Обнинск: ИАТЭ, 1999. – 100 с.
36. *Мостеллер Ф., Тьюки Дж.* Анализ данных и регрессия. Вып. 1. – М.: Финансы и статистика, 1982. – 315 с.
37. *Шаракшанэ А.С., Железнов И.Г., Ивницкий В.А.* Сложные системы. – М.: Высшая школа, 1977. – 261 с.
38. *Скрипник В.М., Назин А.Е., Приходько Ю.Г., Благовещенский Ю.К.* Анализ надежности технических систем по цензурированным выборкам. – М.: Радио и связь, 1988. – 184 с.
39. *Закс Ш.* Теория статистических выводов. – М.: Мир, 1975 – 776 с.
40. *Хэй Дж.* Введение в методы байесовского статистического вывода. – М.: Финансы и статистика, 1987. – 335 с.
41. *Де Гроот М.* Оптимальные статистические решения. – М.: Мир, 1974. – 492 с.
42. *Антонов А.В., Острейковский В.А., Петренко А.А.* Метод учета априорной информации при определении надежности оборудования ядерных энергетических установок. Препринт ФЭИ-1280. – Обнинск: ФЭИ, 1982. – 12 с.
43. *Моррис У.* Наука об управлении. Байесовский подход. – М.: Мир, 1971. – 304 с.

44. *Parzen E.* On estimation of a probability density function and mode//Annals of Mathematical Statistics. – 1962. – 33. – P.1065 – 1076.
45. *Rozenblatt M.* Remark on some nonparametric estimates of a density function// Annals of Mathematical Statistics. – 1956. – 27. – P. 832 – 837.
46. *Деврой Л., Дьерфи Л.* Непараметрическое оценивание плотности. L1-подход: Пер.с англ. – М.: Мир, 1988. – 408 с.
47. *Ченцов Н.Н.* Оценка неизвестной плотности распределения по наблюдениям// Докл. АН СССР. – Т. 147. – С. 45 – 48.
48. *Островский Е.И.* Экспоненциальные оценки для случайных полей и их применения. – Обнинск: ИАТЭ, 1999. – 350 с.
49. *Ивченко Г.И., Каштанов В.А., Коваленко И.Н.* Теория массового обслуживания. – М.: Высшая школа, 1982. – 256 с.
50. *Банди Б.* Методы оптимизации. Вводный курс /Пер. с англ. – М.: Радио и связь, 1988. – 128 с.
51. Проблемы методологии системного исследования. – М.: Мысль, 1970. – 456 с.
52. Системные исследования. Методологические проблемы. – М.: Наука, 1979. – 384 с.
53. Системные исследования. Методологические проблемы. – М.: Наука, 1981. – 384 с.
54. Системные исследования. Методологические проблемы. – М.: Наука, 1985. – 360 с.
55. Системные исследования. Методологические проблемы. – М.: Наука, 1987. – 495 с.
56. *Антонов А.В.* Системный анализ. Методология. Построение моделей: Учеб. пособие по курсу «Системный анализ». – Обнинск: ИАТЭ, 2001. – 272 с.
57. *Антонов А.В.* Системный анализ. Математические модели и методы: Учеб. пособие по курсу «Системный анализ». – Обнинск: ИАТЭ, 2002. – 114 с.
58. *Мушик Э., Мюллер П.* Методы принятия технических решений: Пер.с нем. – М.:Мир, 1990. – 208 с.
59. *Вентцель Е.С., Овчаров Л.А.* Теория случайных процессов и ее инженерные приложения. – М.: Наука, 1991.
60. Оптимизация числа запасных элементов оборудования, важных для безопасности АЭС /*Антонов А.В., Пляскин А.В., Чепурко В.А.* // Методы менеджмента качества. – 2001. – № 8. – С. 27 – 30.
61. *Байхельт Ф., Франкен П.* Надежность и техническое обслуживание. Математический подход. – М.: Радио и связь, 1988. – 392 с.
62. *Мак-Кракен Д., Дорн У.* Численные методы и программирование на фортране. – М.: Мир, 1977. – 584 с.
63. *Хемминг Р.В.* Численные методы. – М.: Наука, 1972. – 400 с.
64. *Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.* Численные методы. – М.: Лаборатория Базовых Знаний, 2001. – 632 с.
65. *Васильев Ф.П.* Численные методы решения экстремальных задач. –М.: Наука, 1988. –552с.
66. *Копченова Н.В., Марон И.А.* Вычислительная математика в примерах и задачах. – М.: Наука, 1972. – 367 с.
67. Надежность и эффективность в технике: Справочник. Т.8: Эксплуатация и ремонт./*Под ред. В.И.Кузнецова и Е.Ю.Барзиловича.* – М.: Машиностроение, 1990.
68. *Барлоу Р., Прошан Ф.* Статистическая теория надежности и испытания на безотказность/Пер. с англ. – М.: Наука, 1984.

ОГЛАВЛЕНИЕ