ЗНАНИЕ

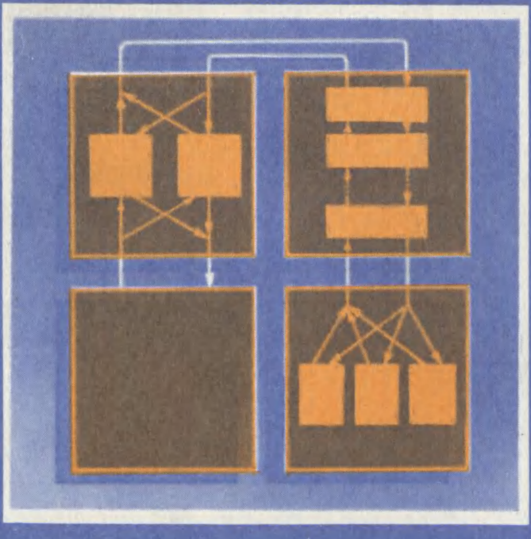
НОВОЕ
В ЖИЗНИ,
НАУКЕ,
ТЕХНИКЕ

СЕРИЯ
МАТЕМАТИКА,
КИБЕРНЕТИКА

9'80

В. В. Калашников

СЛОЖНЫЕ СИСТЕМЫ И МЕТОДЫ ИХ АНАЛИЗА



НОВОЕ
В ЖИЗНИ,
НАУКЕ,
ТЕХНИКЕ

Серия
«Математика,
кибернетика»
№ 9, 1980 г.

Издается
ежемесячно
с 1967 г.

В. В. Калашников,
доктор физико-математических наук

СЛОЖНЫЕ СИСТЕМЫ И МЕТОДЫ ИХ АНАЛИЗА

Издательство
«Знание»
Москва
1980

ББК 32.816
К 17

Калашников В. В.

К 17 Сложные системы и методы их анализа.— М.: Знание, 1980.— 64 с.— (Новое в жизни, науке, технике. Сер. «Математика, кибернетика»; № 9)

11 к.

В брошюре излагается современная точка зрения на предмет теории сложных систем. Объясняется возникновение термина «сложная система» и описываются специфические задачи этой математической дисциплины. Основное внимание уделяется направлениям математических исследований сложных систем и прикладной интерпретации их результатов.

Брошюра рассчитана на научных работников, инженеров, студентов и всех, кто интересуется современным состоянием анализа сложных систем.

20200 ББК 32.816
517

1. СЛОЖНЫЕ СИСТЕМЫ

Около четверти века назад в обиход был введен странный термин «*большие* (*или сложные*) *системы*». Встреченный поначалу весьма неодобрительно и скептически, он постепенно прижился и даже вошел в учебные программы вузов, хотя и поныне существует известный разнобой в его трактовке. Можно предположить, что термин «сложные системы» был вызван известной растерянностью исследователей перед возникшими действительно сложными задачами и содержал одновременно априорное оправдание возможной неудачи исследования. Вместе с тем данный термин как будто специально вызывает аналогию с парадоксом о куче зерна, подчеркивая тем самым известную неопределенность области исследования.

В предлагаемой читателю брошюре дается попытка объяснить смысл понятия «сложная система», раскрыть его содержание. Основное внимание уделяется формулировке задач теории сложных систем, их прикладной интерпретации, изложению ряда подходов к решению данных задач, а также рассматривается математический аппарат, дающий возможность изучать свойства сложных систем.

Разумеется, объем брошюры не позволяет остановиться на всевозможных аспектах теории сложных систем — автор выбрал лишь те из них, которые лежат в сфере его интересов.

Термин «система» появился в научной литературе давно и является фактически таким же неопределенным, как «множество» или «совокупность». Наиболее широко данный термин первоначально использовался в механике, где обозначал материальную систему, т. е. совокупность материальных точек, подчиненных, вообще говоря, некоторым связям. Основной интерес для подобных систем представляют задачи динамики, выявляющие причинноследственный механизм их движения. Становление динамики как науки связано с такими именами, как Галилей, Гюйгенс, Ньютон. Гению Ньютона мы обязаны не только точной формулировкой основных законов механики, но и построением дифференциального исчисления, на котором базируются адекватные математические модели механических систем. Причем модели, использующие дифференциальные уравнения, оказались настолько удачными, что были затем использованы для описания самых разнообразных физических явлений — электрических, электромагнитных, гидравлических и т. д. В последние десятилетия они применялись даже для описания боевых операций [11], глобальных процессов, происходящих в мире [38], и других, в которых одна из главных ролей принадлежит живой природе вообще и человеку в частности.

Системы, изучаемые классической механикой, никто не называл сложными, хотя становление механики насчитывает не одно столетие. Законы динамики (или по современной громоздкой терминологии — законы функционирования механических систем) были получены длительным индуктивным путем. Выдвигаемые гипотезы проверялись на многочисленных опытах. Иные из этих опытов стали классическими. Проверялись также и многочисленные следствия выдвигаемых гипотез. Все это было реализовано во многом благодаря существующей в механике (да и в большинстве других разделов физики) возможности ставить «чистые опыты», т. е. устранять многочисленные мешающие факторы — сводить трение к минимуму, ставить опыты в вакууме, проводить достаточно точные измерения и т. д. Кроме того, условия опытов могли быть воспроизведены с весьма большой точностью в другое время и другом месте. Иначе говоря, длительность эксперимента (те несколько веков, в течение которых производились опыты) была настолько малой, что за это время внешние условия, влияющие на результат опытов, если и изменились, то весьма несущественно.

Ситуация *«повторяемости опытов»* существовала достаточно долго, да существует и сейчас, когда речь идет о явлениях, не подвластных человеку. Новая эра началась с момента, когда ученые захотели исследовать процессы и явления, динамика которых во многом зависит от человека, принимаемых им решений и которые сами влияют на жизнь человеческого общества. Такие процессы и явления описываются, как правило, большим числом параметров —

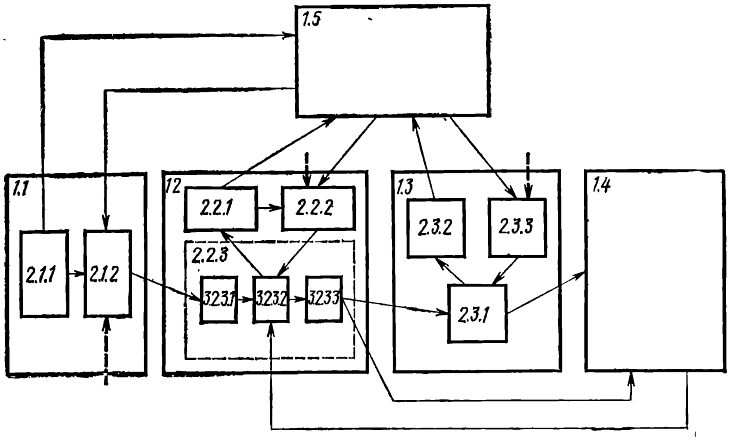


Рис. 1. **Подсистемы 1-го уровня.**
1.1. — система дальнего обнаружения и управления; 1.2. — система многоканальной связи с самолетом; 1.3. — система слепой посадки; 1.4.— система диспетчеризации; 1.5.— бортовая система самолета

Подсистемы 2-го уровня.
2.1.1.— радиолокационный передатчик; 2.1.2.— радиолокационный приемник; 2.2.1.— передатчик; 2.2.2.— приемник; 2.2.3.— вычислительное устройство; 2.3.1.— вычислительное устройство; 2.3.2.— радиолокационный передатчик; 2.3.3.— радиолокационный приемник

Подсистемы 3-го уровня.
3.2.3.1.— устройство обработки информации, полученной с подсистемы дальнего обнаружения; 3.2.3.2.— устройство «ведения» самолетов; 3.2.3.3.— аппаратура выдачи информации

большим в том смысле, что соответствующие уравнения и соотношения, как правило, аналитически разрешены быть не могут. И это — одна, но не главная причина появления термина «сложная система». Предполагалось, что «проклятие размерности» будет хотя бы отчасти снято с развитием ЭВМ. Оказалось, однако, что это не так. Появление ЭВМ и их быстрое развитие показали, что для сложных систем старые модели и методы не годятся, и вот это-то стимулировало развитие теории сложных систем. Основной причиной «осечки» традиционных методов исследования была уникальность изучаемых систем, приводящая к ситуации, когда длительность экспериментов с ними сравнима со сроком жизни самой системы. Причем отнюдь не исключается ситуация невозможности экспериментирования (например, с проектируемой системой). Отметим также практическую невозможность проведения «чистых опытов» со сложными системами — их приходится изучать во всем многообразии действующих факторов. Именно эта осечка и привела к рождению термина «сложная система». Поясним сказанное двумя примерами, упрощенно описывающими реальные системы.

Система управления движением самолетов [10]. Рассмотрим систему, целью которой является регулирование потока самолетов в районе крупного аэродрома, управление посадкой (рис. 1). Такая система состоит из ряда подсистем: дальнего обнаружения и управления; многоканальной связи с бортом самолета; слепой посадки; диспетчеризации. Поскольку эта система взаимодействует (обменивается информацией) с бортовыми системами контролируемых самолетов, то последние также естественно включить в состав всей системы.

Опишем кратко работу всей системы. Самолеты, попадающие в зону «видимости» подсистемы дальнего обнаружения и управления, фиксируются радиолокаторами, и соответствующие сигналы обрабатываются вычислительным устройством с целью отсева самолетов, следующих в другие аэропорты. Если самолет следует в данный аэропорт, то управление им берет на себя подсистема многоканальной связи, состоящая из приемопередающего устройства, обменивающегося информацией с самолетом и упомянутым выше вычислительным устройством. Последнее посредством специальной аппаратуры выдачи передает информацию в подсистему диспетчеризации, в которой определяется порядок посадки самолета, выделяется посадочная полоса и канал, отвечающий за слепую посадку и взлет самолетов.

Именно в подсистеме диспетчеризации решается, когда управление самолетом следует передать от подсистемы дальней связи подсистеме слепой посадки и взлета, в которой и будет производиться дальнейшая обработка информации. Подсистема диспетчеризации состоит из центрального вычислительного устройства и соответствующих устройств ввода-вывода и отображения информации. Подсистема посадки и взлета в соответствии со своим назначением состоит из радиолокационных приемных и передающих устройств и устройств переработки информации.

Все перечисленные подсистемы и составляющие их устройства работают совместно, и именно совместная согласованная работа обеспечивает нормальную работу системы в целом.

Наличие большого числа устройств и связей между ними приводит к невозможности исследования такой системы традиционными методами, хотя характеристики некоторых ее элементов вполне могут быть изучены даже аналитически. Однако такое изучение «изолированных» элементов не дает информации о работоспособности системы в целом. Лишь анализ работы подсистем с учетом их взаимодействия способен дать ответы на вопросы о работоспособности всей системы. Вместе с тем ясно, что эксперименты с подобной системой с целью определения законов ее функционирования невозможны, во всяком случае в достаточно полном объеме, ибо это связано не только с большими материальными издержками, но и с безопасностью людей. Но даже и в экспериментах ограниченного объема исследователь должен считаться с реальными условиями, в которых работает система. В самом деле, постановка «чистых» экспериментов (отвечающих, например, регулярному поступлению самолетов или отсутствию помех в каналах связи и т д.) не только физически невозможна, но и была бы принципиально неправильной с методологической точки зрения, так как «поправки» на реальные условия в данном случае не являются, вообще говоря, ни в каком смысле малыми. Эти и другие особенности функционирования сложных систем будут нами далее обсуждаться более подробно.

Производственно-организационная система строительства плотины ГЭС. [22]. Данная система служит (рис. 2) для обеспечения и управления процессом строительства плотин и состоит из следующих хозяйств (подсистем): карьеры, бетонные заводы, транспортная линия, укладочный комплекс плотины. Мы не включаем в состав системы весьма важные организации и производства, связанные с поставками и установкой металлоконструкций, оборудования и т. п., для того, чтобы упростить изложение. Вместе с тем отметим, что именно «бетонные работы» при строительстве ГЭС являются наиболее трудоемкими и продолжительными.

Система в целом работает следующим образом. Добытый в карьерах камень перерабатывается в щебень и отвозится на бетонные заводы для приготовления бетона. Для различных работ существуют различные марки бетона, и каждый из бетонных заводов имеет возможность готовить любую из марок. Более того, каждая секция, из которых состоят эти заводы, также может готовить

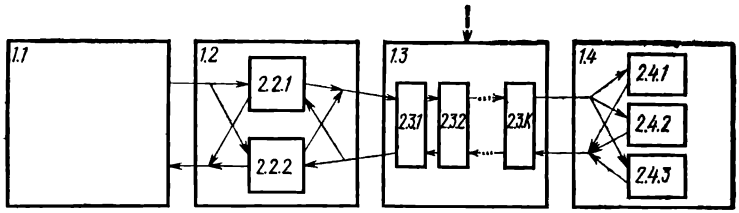


Рис. 2. **Подсистемы 1-го уровня.**
1.1.— карьеры; 1.2— бетонные заводы; 1.3.— транспортная линия «бетонные заводы — плотина»; 1.4.— укладочный комплекс плотины
Подсистемы 2-го уровня.
2.2.1.— 1-й бетонный завод; 2.2.2.— 2-й бетонный завод; 2.3.1.— 2.3.К.— «характерные» участки дороги; 2.4.1.—2.4.3.— разгрузочные устройства

любую из марок бетона (марки отличаются процентным содержанием цемента и добавок), однако для перестройки производства с одной марки на другую уходит дополнительное время. Бетон готовится каждой из секций порциями, которые загружаются в бетоновозы и по трассе отвозятся на место строительства. Разгрузка производится, вообще говоря, в различных местах, различными механизмами и по разным правилам, зависящим, кроме всего прочего, от этапа строительства (с ростом плотины изменяются состав разгрузочных механизмов, правила подъезда к ним и т. д.). Разгруженные бетоновозы едут обратно на бетонные заводы и т. д.

Чтобы дать представление о размерностях решаемых задач, скажем, что на строительстве обычно работа т несколько бетонных заводов (1—3), производящих около 10 марок бетона, автохозяйство насчитывает несколько сотен автомобилей, работают до десятка разгрузочных механизмов, протяженность автотрассы составляет от единиц до нескольких десятков километров при достаточно большой ее загрузке «посторонним» транспортом, создающим помехи перевозке бетона.

Создание подобной системы требует решения следующих вопросов:

а) выбор трассы подвоза бетона к плотине (при этом возможны случаи использования общегосударственных дорог, прокладка новых, прокладка специальных подъездных путей и тоннелей и т. д.);

б) выбор количественного и качественного состава автопарка;

в) выбор типов бетонных заводов, режимов их работы, мест размещения;

г) выбор состава разгрузочных механизмов, режимов их работы и т. д.

Ясно, что и в этом случае натурные эксперименты возможны лишь в весьма ограниченном объеме. В силу же уникальности каждой стройки результаты экспериментов, полученные в одних условиях, будут малополезными в других. Следовательно, и здесь проявляются те же особенности сложных систем, о которых шла речь выше.

Приведенные примеры дают некоторое представление о характере и типах сложных систем, хотя далеко не исчерпывают их разнообразия. В дальнейшем мы будем обращаться к данным примерам для иллюстрации некоторых общих положений.

Попытаемся выявить общие черты, присущие рассмотренным системам, с тем чтобы определить понятие сложной системы и очертить тем самым класс изучаемых объектов.

Даже при поверхностном описании примеров нам пришлось ввести понятие подсистемы как некоторой достаточно автономной части всей системы. Расчленение сложной системы на подсистемы, как правило, обладает значительным произволом и зависит как от принятых технических решений, целей создания системы, так и от взглядов исследователя на систему. Это в первую очередь объясняется разнородностью элементов, составляющих сложную систему. Так, в первом примере вычислительное устройство, обрабатывающее сигналы из подсистемы дальнего обнаружения, мы «приписали» к этой подсистеме, но могли бы выделить и в отдельную подсистему. Более того, каждая из перечисленных подсистем состоит из ряда более «мелких» подсистем и т. д. Во втором примере «транспортная» подсистема, отображающая движение автомобилей по трассе от бетонных заводов к месту строительства, может быть, в свою очередь, расчленена на более мелкие подсистемы, отображающие, допустим, движение автомобилей по отдельным участкам трассы. В первом примере также все из перечисленных подсистем могут быть расчленены еще на более мелкие подсистемы в соответствии с составляющими их устройствами.

В сложных системах одну из основных черт составляет взаимодействие выделенных подсистем. Это взаимодействие возникает из-за внесенного исследователем деления всей системы на подсистемы. Как уже отмечалось, такое деление обладает определенным произволом. Чтобы обеспечить при этом функционирование всей системы как единого целого, приходится тем или иным способом учитывать результат воздействия одной подсистемы на другую. Обычно такое взаимодействие сводится к обмену сигналами между подсистемами, осуществляемому по каналам связи, проложенным от одной подсистемы к другой. Еще раз подчеркнем, что эти каналы связи могут как отвечать реальным каналам, существующим в системе, так и порождаться принятым делением системы на подсистемы. Кроме того, взаимодействие осуществляется между внешней средой и выделенными элементами системы. Учет его аналогичен учету взаимодействия между подсистемами.

Таким образом, *сложная система представляется в виде многоуровневой конструкции из взаимодействующих между собой и с внешней средой элементов* [9, 10], где к элементам 1-го уровня относятся подсистемы, на которые первоначально разбита исходная система, к элементам 2-го уровня — подсистемы, получающиеся из разбиения подсистем 1-го уровня, и т. д. до тех пор, пока получившиеся элементы не признаются «простыми» для исследования.

Рис. 1 и 2, о которых мы уже говорили, дают варианты возможных многоуровневых представлений двух рассмотренных выше систем. Так, в системе управления движением самолетов имеются три уровня. Обмен сигналами осуществляется в соответствии с показанными на рисунке стрелками и отвечает приведенному выше описанию взаимодействия подсистем. Пунктирными стрелками показаны наиболее значительные воздействия внешней среды.

Производственно-организационная система строительства плотины ГЭС представлена в виде двухуровневой конструкции из взаимодействующих элементов. Если каналы связи, показанные на рис. 1, в основном отвечали реальным каналам связи, то в данной системе стрелки отвечают не действительным каналам связи (которых нет в реальности), а служат учету взаимодействия, существующего между выделенными подсистемами. Физически обмен сигналами отвечает в данном случае поступлению либо уходу единицы автотранспорта с рассматриваемой подсистемы. При этом сами подсистемы выделены исходя из реально существующих особенностей. Для простоты изображения принято, что работают два бетонных завода, существуют три места укладки бетона с соответствующими механизмами. Взаимодействие с внешней средой в данном случае может сводиться к отказу (сбоям) в работе заводов и укладочных механизмов, изменению дорожных условий (метеорологических и из-за движения «постороннего» транспорта).

Возвращаясь к вопросу об условности деления системы на подсистемы, отметим, что подсистемы 2.2.1, 2.2.2, 2.4.1—2.4.3 представляют собой не только бетонные заводы и разгрузочные устройства соответственно, но фактически включают в себя и автомашины — как ожидающие загрузки (соответственно разгрузки), так и загружаемые (разгружаемые) в настоящий момент. Аналогично, подсистемы 2.3.1—2.3.K включают в свое описание не только параметры соответствующих участков дороги, но и характеристики движущегося по ним транспорта.

Приведенное выше представление сложной системы в виде многоуровневой конструкции из взаимодействующих элементов имеет своей целью изучить сложную систему «по частям». Какие же это части?

Прежде всего это подсистемы различных уровней, на которые разбивается исходная система. Поскольку подсистемы «верхних» уровней сами зачастую являются сложными и разбиваются далее, то на самом деле изучению подлежат самые «мелкие», неделимые подсистемы, которые условимся называть *элементами*. Самостоятельной частью является так называемая *схема сопряжения* элементов, т. е. схема, реализующая адресацию сигналов, отражающих взаимодействие элементов. Подчеркнем, что системы, не называемые сложными, изучаются как единое целое, без указанной структуризации. В этом, по-видимому, следует искать *основное отличие* между «простыми» и «сложными» системами: оно, это отличие, порождается не только и не столько самими исследуемыми системами как таковыми, а в основном задачами, которые стоят перед исследователями.

Рассмотрим более подробно составные части сложных систем с тем, чтобы представить себе задачи, возникающие при их исследовании.

Обратимся сначала к элементам. Каждый элемент представляет собой *динамическую систему в широком смысле* [9, 10], т. е. (а) систему, функционирующую во времени, (б) изменяющую с течением времени свое состояние под действием внутренних и внешних причин, (в) воспринимающую входные и выдающую выходные сигналы в процессе взаимодействия с другими элементами системы.

Перечисленные свойства, являющиеся общими для различных элементов, не раскрывают, однако, важных частных особенностей, присущих им.

Рассмотрим вновь наши примеры. В системе управления движением самолетов можно выделить несколько разнородных типов элементов — дискретные (вычислительные устройства), дискретно-непрерывные (радиолокационные устройства), непрерывные (передатчики и приемники; следящие системы, входящие в состав систем устройств управления и ведения самолетов). К этому необходимо добавить, что все устройства работают не только в условиях случайных помех, но и в условиях случайных потоков самолетов.

В системе строительства плотины ГЭС также можно выделить разнородные элементы — бетонные заводы, разгрузочные устройства, участки дороги. Влияние случайных факторов здесь тоже весьма существенно (отказы механизмов, случайные потоки транспорта, случайные времена выполнения различных операций). Все сказанное приводит к выводу о том, что для описания элементов сложных систем необходимо иметь набор формальных математических схем, удовлетворяющих, с одной стороны, требованиям (а)—(в) (т. е. являющихся динамическими системами в широком смысле), а с другой — учитывающих различные частные особенности функционирования систем (характер переработки информации, ее восприятия и выдачи, наличие помех и т. д.).

Таким образом, одной из важных задач теории сложных систем является поиск математических схем, достаточно адекватно отражающих специфику работы элементов сложных систем. В то же время данные математические схемы должны допускать исследование аналитическими либо машинными методами.

Далее будет рассмотрен ряд таких схем, нашедших широкое применение в теории сложных систем. Основное внимание мы при этом уделим методам их исследования и содержательной интерпретации анализируемых свойств.

Как неоднократно подчеркивалось, основную черту в функционировании сложных систем составляет взаимодействие элементов, представляемое нами в виде *механизма обмена сигналами*. Данный механизм включает в себя [9, 10] процессы формирования выходных сигналов и реагирования на входные сигналы различными элементами, адресации сигналов и их прохождения по каналам связи.

Учет процессов формирования выходных и реагирования на поступление входных сигналов относится к проблеме построения элементов как динамических систем в широком смысле (см. свойство (в)). Далее удобно считать, что во время прохождения по каналу связи сигнал не подвергается различного рода искажениям, а сам процесс передачи происходит мгновенно. Такие каналы (назовем их *идеальными* [9, 10]), естественно, возникают при расчленении системы в случае, когда реальная система не содержит на самом деле никаких каналов связи, а последние вводятся только для учета существующего взаимодействия элементов системы (см. пример системы строительства плотины ГЭС, рис. 2). В случае же, если в реальной системе присутствуют реальные каналы связи (см. систему управления движением самолетов, рис. 1), являющиеся, естественно, неидеальными, представляется удобным формализовать их как отдельные элементы системы (реализующие задержки и искажения сигналов), соединенные с другими элементами уже идеальными каналами связи. Следовательно, для изучения процесса взаимодействия нам осталось рассмотреть схему сопряжения, реализующую адресацию сигналов в системе по идеальным каналам связи. При задании схемы сопряжения элементы представляются в виде многополюсников, имеющих определенное число *входных* и *выходных контактов* — количества этих контактов различные для различных элементов. Перенумеруем все входные контакты элементов, входящих в систему, и аналогичным образом перенумеруем все выходные контакты. Тогда задание схемы сопряжения означает попросту сопоставление каждой паре (i, j) (где i — «сквозной» номер входного контакта, а j — «сквозной» номер выходного контакта) факта наличия или отсутствия идеального канала связи. Легко видеть, что данная конструкция учитывает и взаимодействие системы с внешней средой — достаточно лишь внешнюю среду представить фиктивным элементом, имеющим определенное число входных и выходных контактов.

Какие же задачи возникают в связи со схемой сопряжения? Прежде всего это задачи *структурного анализа* сложных систем, выявляющие различные отношения между элементами системы [9, 10]. Именно интерес представляет, например, вопрос о существовании цепочки каналов связи, соединяющих различные элементы. Более глубокое исследование предполагает учет направления передачи сигналов, а также их вида. Под видом понимается некоторая содержательная интерпретация назначения передаваемых сигналов. Например, некоторые сигналы отвечают материальным потокам в системе, другие — информационным, третьи служат целям управления и т. д. Такое деление сигналов обычно проявляется в факте их возникновения на определенных выходных клеммах, что и дает возможность использовать для изучения указанных задач структурные методы, позволяющие найти так называемые *типичные структурные конфигурации* (цепи, циклы, контуры и т. п.), играющие важную роль в определении возможностей системы по передаче и переработке сигналов. При этом выявляются еще и такие свойства структур, как связность, иерархичность и др.

Особую роль играют формальные структурные преобразования, когда исходная структура системы преобразуется в другую. Например, некоторая подсистема может расчленяться на ряд более мелких подсистем или, напротив, ряд элементов объединяется в одну подсистему. Такие преобразования играют важную роль на этапе синтеза, когда решается вопрос о возможности создания системы, обладающей заданными свойствами, имея некоторый стандартный набор элементов. Вместе с тем приходится констатировать, что методы структурного анализа сложных систем в настоящее время разработаны слабее, нежели методы анализа динамики отдельных элементов. Отдельные результаты, а также дальнейшие ссылки читатель может найти в [8—10].

Итак, основной задачей теории сложных систем следует считать разработку методов, позволяющих на основе изучения особенностей функционирования, получения характеристик отдельных элементов и анализа механизма взаимодействия между элементами получать характеристики системы в целом. Единственный метод, существующий на сегодняшний день и позволяющий находить характеристики системы в целом, является машинное моделирование. Его особенности мы вкратце рассмотрим в конце брошюры (предполагается, что более подробно вопросам моделирования будет посвящена отдельная публикация). Однако метод моделирования может успешно применяться лишь при том условии, что его реализация в полной мере учитывает особенности моделируемых объектов, т. е. моделирование должно быть «целенаправленным» (в смысле согласования модельных экспериментов со свойствами моделируемой системы).

Имея это в виду, мы далее рассмотрим ряд математических схем, являющихся моделями элементов сложных систем. Для этих схем будут рассмотрены различные общие свойства, учет которых позволяет в конечном итоге организовать упомянутые целенаправленные модельные эксперименты. Поэтому далее такие достаточно традиционные вопросы, как методы нахождения вероятностно-временных характеристик систем, их оптимизация и т. п., решение которых проводится уже *после* формирования модели, рассматриваются бегло. Свойства же, изучаемые более подробно, могут быть использованы именно в *процессе* составления модели.

2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ЭЛЕМЕНТОВ СЛОЖНЫХ СИСТЕМ

Разнородность назначения элементов сложных систем, наличие большого количества действующих на них факторов, функционирование в случайных условиях и т. д. приводят к разнообразию математических моделей, применяемых для описания элементов сложных систем. С одной стороны, эти модели отражают объективные закономерности функционирования систем. Но с другой — вид используемой модели определяется задачами исследования. Так, участок дороги может описываться достаточно сложным стохастическим процессом при исследовании его пропускной способности и критических режимов движения; если же этот участок является элементом сложной системы (см. рис. 2), то иногда его достаточно представить в виде простой задержки времени.

Часто элементы сложных систем классифицируют по содержательному принципу. Так, можно выделить следующие классы:

а) системы автоматического управления, следящие системы;

б) конечные автоматы, релейно-контактные схемы;

в) вероятностные автоматы;

г) системы массового обслуживания;

д) системы передачи и обработки информации;

е) резервированные системы;

ж) системы управления запасами;

и) производственные системы (непрерывные и дискретные);

к) транспортные системы.

Разумеется, подобный перечень заведомо не является полным; более того, содержание его постоянно изменяется с изменением техники, технологии, расширением класса исследуемых объектов и т. д.

Каждый из перечисленных классов имеет свою специфику, порождаемую, во-первых, назначением соответствующих систем, а во-вторых, применяемыми математическими методами исследования [1, 2, 7, 13, 25, 26, 32, 33, 40]. Вместе с тем можно констатировать, что данные классы определены весьма нечетко — например, не существует формальных понятий «система автоматического управления» или «система массового обслуживания» и т. д. В лучшем случае можно указать ряд признаков, отличающих системы из каждого класса. Однако зачастую такую грань между классами провести трудно или даже невозможно: некоторые из систем с равным основанием можно отнести, как, например, в классы систем массового обслуживания, так и резервированных систем или систем управления запасами. Подобное положение вполне естественно — оно лишь указывает на то, что для исследования различных систем зачастую используются одни и те же математические методы. Изучаемые системы при этом могут соответствовать тому или иному классу математических объектов. Поэтому представляет целесообразным вместо классификации элементов сложных систем по *содержательному признаку* использовать классификацию *математических методов*, применяемых при исследовании этих элементов. Среди используемых математических моделей особое место занимают следующие:

1) дифференциальные и разностные уравнения [3, 4, 12];

2) общие динамические системы [23];

3) марковские процессы [15, 24, 37];

4) регенерирующие процессы [29, 37];

5) кусочно-линейные (линейчатые) процессы [5, 10, 19, 27];

Данные модели можно назвать *типовыми математическими схемами*, поскольку они широко используются при исследованиях сложных систем. Наличие разработанных математических методов (и их дальнейшая разработка) исследования этих схем значительно повышает их ценность при использовании в качестве моделей элементов сложных систем. Вряд ли стоит особо останавливаться на том известном обстоятельстве, что любая математическая модель никогда не бывает абсолютно адекватной изучаемому процессу, но отражает лишь основные его черты с учетом задач, стоящих перед исследователем.

Дифференциальные уравнения, пожалуй, наиболее широко используемая сейчас математическая схема. Ее популярность объясняется, во-первых, достаточно хорошо разработанной теорией дифференциальных уравнений, во-вторых, многочисленными успешными «прецедентами» применения этой схемы (о чем мы уже говорили выше) и, в-третьих (но не в последних), знакомством большинства исследователей с данной схемой, поскольку теория дифференциальных уравнений входит в число обязательных курсов технических вузов.

Дифференциальные уравнения используются для описания систем, функционирующих во времени в условиях, когда влияниями различного рода случайностей можно пренебречь — систем автоматического управления, механических, гидравлических и вообще физических систем (если говорить не только об обыкновенных уравнениях, но и об уравнениях в частных производных). Более того, с помощью дифференциальных уравнений можно описать и ряд стохастических систем. Например, переходные вероятности марковской цепи с непрерывным временем описываются дифференциальными уравнениями Колмогорова [24, 37].

Если расширить класс дифференциальных уравнений, включив в него, например, стохастические дифференциальные уравнения, то можно существенно расширить класс описываемых систем. Например, таким образом можно описывать работу систем автоматического управления в условиях случайных помех.

Разностные уравнения являются, по существу, дискретным аналогом дифференциальных уравнений, и потому все сказанное выше относится и к ним.

Общие динамические системы. Отражая динамику систем, схема дифференциальных уравнений тем не менее не отражает возможности восприятия и выдачи этими системами входных и выходных сигналов соответственно. Последняя же является одной из основных особенностей элементов сложных систем (см. выше). Имея в виду учет данной возможности, различными авторами были предприняты попытки обобщения схемы дифференциальных и разностных уравнений. Пожалуй, наиболее удачное обобщение было сделано Р. Калманом [23], включающее как частные случаи дифференциальные и разностные уравнения, конечные автоматы, классические динамические системы и т. д.

Обобщение Калмана отражает существующую в теории сложных систем тенденцию к изучению наиболее общих моделей. И хотя динамическая система Калмана во всей общности в настоящее время не допускает получения содержательных результатов, она служит хорошей методологической основой для исследования различных частных случаев.

Однако, несмотря на, казалось бы, предельную общность динамической системы Калмана, она не учитывает такой важной особенности функционирования элементов сложных систем, как *случайность*. Мы уже отмечали, что в сложных системах случайность, как правило, играет роль не «малой» помехи, а составляет неотъемлемую часть их функционирования. Это обстоятельство может приводить (и, к сожалению, приводит) к тому, что система, спроектированная на основе «средних» характеристик или других детерминированных аппроксимаций, оказывается принципиально неработоспособной в реальных условиях. Так, хорошо известны примеры, когда системы обслуживания, рассчитанные по средним показателям, «захлебывались» при работе с реальными потоками требований.

Учет «случайности» приводит к необходимости рассмотрения процесса функционирования элемента сложной системы как случайного. При этом тип исследуемого процесса определяется как содержанием анализируемой системы, так и задачами анализа.

Мы рассмотрим три класса *случайных процессов*, которые далее будут использованы в примерах: *марковские*, *регенерирующие* и *линейчатые*. Эти классы являются пересекающимися. Более того, линейчатые процессы являются подклассом марковских. На примерах указанных процессов будут проиллюстрированы некоторые общие задачи, возникающие в теории сложных систем, и показаны подходы к их решению. Мы далеки от мысли о том, что три перечисленных класса процессов исчерпывают применяемые стохастические модели сложных систем. Тем не менее широкое применение, а также новизна полученных результатов дают основания считать их удачным иллюстративным материалом.

Следует оговориться, что случайные процессы, рассматриваемые далее, естественно считать одновременно процессами изменения внутреннего состояния, входными и выходными процессами систем либо любыми из этих процессов — мы не будем останавливаться на вопросе разделения изучаемого процесса на соответствующие компоненты. Ниже будут считаться известными понятиями *«вероятность»*, *«условная вероятность»* и *«независимость»* [24, 37]. Состояние рассматриваемого процесса будем обозначать через Y_t. Предположим, что аргумент (время) t принимает значения из множества T. Если $T = [0, \infty)$, то процесс $\{Y_t\}$ называется процессом *с непрерывным временем*, а если $T = \{0, 1, 2, \ldots\}$ — *с дискретным временем*. Иногда рассматривают также $T = (-\infty, \infty)$ или $T = \{\ldots, -1, 0, 1, \ldots\}$, но мы ограничимся лишь указанными выше случаями, считая тем самым, что у процесса функционирования системы есть «начало».

Обозначим множество значений процесса через V, т. е. $Y_t \in V$ при каждом $t \in T$. Часто в качестве V рассматривается евклидово пространство R^n. Однако мы встретимся далее и с более сложными случаями.

Задать случайный процесс — это значит задать совокупность случайных величин $\{Y_t\}_{t \in T}$. Каждая случайная величина Y_t может быть задана своим распределением вероятностей, т. е. совокупностью вероятностей вида $F_t(B) = P(Y_t \in B)$, $B \subset V$*. Однако чтобы задать совокупность $\{Y_t\}_{t \in T}$, недостаточно задать распределения F_t для каждой из величин Y_t, $t \in T$. Поскольку в общем случае эти величины зависимы, то необходимо задать всю совокупность так называемых конечно-мерных распределений, т. е. для каждого конечного набора величин $t_i \in T$, $i =$

* Ни здесь, ни далее мы не будем останавливаться на вопросах о том, что подмножества $B \subset V$, вообще говоря, не произвольны, а принадлежат определенному классу измеримых подмножеств, а также на других «тонкостях» теории вероятностей и случайных процессов. Подготовленный читатель сделает необходимые оговорки самостоятельно. Читателю же, не знакомому с аксиоматикой теории вероятностей [30], нет возможности объяснить ее в рамках данной брошюры. Мы надеемся лишь, что это обстоятельство не отразится на понимании излагаемого здесь материала. Отметим также, что здесь и далее P означает вероятность, а E — среднее значение.

$= 1, 2, \ldots, k, \quad k = 1, 2, \ldots,$ задать вероятности вида $P\{Y_{t_1} \in B_1, \ldots, Y_{t_k} \in B_k\}, B_i \subset V, \quad i = 1, \ldots, k.$

Марковские процессы, представляющие один из важных классов случайных процессов, характеризуются следующим свойством, которое можно назвать отсутствием памяти: при известном текущем состоянии процесса его «будущее» не зависит от «прошлого», т. е. от того, как траектория процесса попала в текущее состояние. Формально это можно записать так.

При любых $k = 1, 2, \ldots, t_1 < t_2 < \ldots < t_k < t$ $P\{Y_t \in B/Y_{t_1} = y_1, \ldots, Y_{t_k} = y_k\} = P\{Y_t \in B/Y_{t_k} = y_k\}$ (здесь y_k имеет смысл текущего состояния, Y_t — его «будущее» состояние, а $y_1, \ldots, y_{k-1}$ — «прошлое» процесса).

В силу этого для задания марковского процесса достаточно задать лишь начальное распределение $F_0(B) = P(Y_0 \in B)$ и переходную функцию (переходную вероятность), равную вероятности того, что в момент t состояние процесса принимает значение из подмножества B пространства состояний при условии, что в момент τ его состояние было y,

$$P(\tau, y, t, B) = P(Y_t \in B/Y_\tau = y), \ \tau \leqslant t,$$

где при $\tau = t$ считается

$$P(t, y, t, B) = \begin{cases} 1, & y \in B, \\ 0, & y \notin B. \end{cases}$$

Для переходных функций марковских процессов справедливо уравнение Колмогорова — Чэпмена

$$P(t_1, y, t_3, B) = \int\limits_V P(t_1, y, t_2, dy')\, P(t_2, y', t_3, B),$$

$$t_1 \leqslant t_2 \leqslant t_3, \qquad (2.1)$$

выражающее, во-первых, тот факт, что переход траектории марковского процесса из точки y в множество B за промежуток времени $[t_1, t_3]$ может рассматриваться состоящим из двух этапов — перехода из точки y в «малое» множество dy' за промежуток времени $[t_1, t_2]$ и перехода из точки $y' \in dy'$ в множество B за время $[t_2, t_3]$, и, во-вторых, дающее количественное соотношение между вероятностями этих переходов, а именно выражающее переходную функцию за промежуток $[t_1, t_3]$ через переходные функции за меньшие промежутки $[t_1, t_2]$ и $[t_2, t_3]$.

Далее мы без дополнительных оговорок будем предполагать у марковских процессов свойство однородности, т. е. считать, что переходная функция $P(\tau, y, t, B)$ зависит на самом деле не от двух аргументов времени τ и t, а лишь от их разности $t - \tau$, что наглядно соответствует инвариантности свойств марковского процесса при сдвигах времени. В этом случае переходную функцию будем обозначать $P(y, t, B)$.

Выше уже было сказано, что переходная функция (плюс начальное распределение) полностью задает марковский процесс. Однако таким заданием на практике пользуются редко. Гораздо более типичной является ситуация, когда известен вид переходных вероятностей за малые промежутки времени $P(y, \Delta t, B)$. В случае дискретного времени роль таких малых промежутков играют единичные интервалы, и в силу уравнения Колмогорова — Чэпмена значения переходной функции при всех t восстанавливаются по значениям одношаговой переходной функции $P(y, B) \equiv \equiv P(y, 1, B)$. В случае непрерывного времени дело обстоит сложнее.

Пример. Процесс рождения и гибели. Данный процесс широко используется в качестве стохастических моделей обслуживания, резервирования, развития популяций и т. д. Обычно он выделяется с помощью следующих допущений: $T = (0, \infty)$; $V = \{0, 1, 2, \ldots\}$; вероятность того, что за «малый» промежуток времени Δt процесс перейдет в состояние $i + 1$, если в начале этого промежутка его состояние было i, равна $\lambda_i \Delta t + o(\Delta t)$, $\lambda_i \geqslant 0$*, т. е. $P\{Y_{t+\Delta t} = i + 1 / Y_t = i\} = \lambda_i \Delta t + o(\Delta t)$, $i \geqslant 0$; аналогично:

$$P\{Y_{t+\Delta t} = i - 1/Y_t = i\} = \mu_i \Delta t + o(\Delta t), \quad i \geqslant 1, \quad \mu_i \geqslant 0$$

и

$$P\{|Y_{t+\Delta t} - i| > 1/Y_t = i\} = o(\Delta t).$$

Таким образом, здесь процесс задан не с помощью переходных вероятностей, а с помощью параметров λ_i, μ_i, дающих эти вероятности лишь при малых Δt:

$$\lambda_i = \lim_{\Delta t \to 0} \frac{P(i, \Delta t, i + 1)}{\Delta t}, \quad i \geqslant 0,$$

$$\mu_i = \lim_{\Delta t \to 0} \frac{P(i, \Delta t, i - 1)}{\Delta t}, \quad i \geqslant 1.$$

И возникает вопрос, можно ли, зная $\{\lambda_i, \mu_i\}$, «восстановить» переходную функцию процесса рождения и гибели

---

* Здесь обозначение $o(\Delta t)$ имеет следующий стандартный смысл: $o(\Delta t)/\Delta t \to 0$, если $\Delta t \to 0$.

и тем самым определить данный процесс и его характеристики?

Оказывается, что это можно сделать не всегда, а лишь в том случае, если процесс *регулярный*, т. е., грубо говоря, если его траектории не могут за конечное время достичь «бесконечности» (с ненулевой вероятностью).

Например, этот процесс будет регулярным, если все λ_i, μ_i ограничены сверху. Необходимым же и достаточным условием регулярности является

$$\sum_{k=0}^{\infty} \frac{1}{\lambda_k \rho_k} \sum_{j=0}^{k} \rho_j = \infty, \qquad (2.2)$$

где

$$\rho_0 = 1, \ \rho_j = \frac{\lambda_0 \cdots \lambda_{j-1}}{\mu_1 \cdots \mu_j}, \ j \geqslant 1$$

(в частности, процесс с конечным множеством состояний регулярен всегда).

Таким образом, приведенное выше задание процесса рождения и гибели, вообще говоря, не является полным.

Оказывается, вместо задания параметров λ_i, μ_i удобнее воспользоваться следующей конструкцией (несмотря на то что она и кажется на первый взгляд более громоздкой). Зададим на множестве $V = \{0, 1, 2, \ldots\}$ произвольную ограниченную действительную функцию f. Поставим ей в соответствие другую функцию, обозначаемую Af, также заданную на V, значение которой в точке $i \geqslant 0$ следующее:

$$Af(i) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\sum_{j=0}^{\infty} P(i, \Delta t, j) f(j) - f(i) \right].$$

Такое соответствие называется *инфинитезимальным оператором*, который будет обозначаться буквой A. В силу условий, наложенных на переходную функцию, легко найти, что

$$Af(i) = \lambda_i [f(i+1) - f(i)] + \mu_i [f(i-1) - f(i)], \quad i \geqslant 0$$

(где полагается $\mu_0 = 0$).

Очевидно, что оператор A однозначно определяется параметрами $\{\lambda_i, \mu_i\}$, и наоборот. В числителе дроби, определяющей A, стоит среднее приращение случайной функции $f(Y_t)$ за время Δt при условии, что $Y_t = i$, т. е. значение $Af(i)$ есть средняя скорость изменения во времени функции, $f(Y_t)$ в точке $Y_t = i$. Таким образом, оператор A можно считать аналогом полной производной по времени. Это утверждение подкрепляется следующей формулой, являющейся аналогом формулы Ньютона — Лейбница из дифференциального исчисления:

$$E\{f(Y_t)/Y_0 = y\} = f(y) + E\left\{\int_0^t Af(Y_u)\, du/Y_0 = y\right\}. \tag{2.3}$$

Эта формула может служить основой для весьма содержательных результатов относительно поведения процесса $\{Y_t\}$.

Мы не обсуждаем здесь точные условия, при которых формула (2.3) имеет место. В общем случае инфинитезимальный оператор A определяется формулой

$$Af(y) = \lim_{t \to 0} \frac{1}{t}[E\{f(Y_t)/Y_0 = y\} - f(y)] \tag{2.4}$$

и применяется к ограниченным действительным непрерывным функциям f *.

В этом общем случае формула (2.3) также справедлива. Она остается справедливой и при замене времени t на случайное, например на время первого достижения траекторией $\{Y_t\}$ какого-либо фиксированного подмножества пространства V (см. [15]). Вновь, как и ранее, мы не приводим соответствующих точных формулировок.

Задавая вид инфинитезимального оператора, мы тем самым задаем вид переходных функций на бесконечно малых промежутках времени (откуда и название оператора). Основной задачей затем является нахождение тех или иных характеристик марковского процесса $\{Y_t\}$ на основе такого задания.

Рассмотрим теперь линейчатые процессы, представляющие собой подкласс марковских процессов и характеризующиеся рядом ограничений, налагаемых прежде всего на вид возможных траекторий. Предварительно рассмотрим пример.

---

* Для того чтобы ввести понятие непрерывности, нужно предполагать, например, что V является *метрическим пространством* с метрикой h. В случае процесса рождения и гибели требование непрерывности было опущено из-за счетности множества V: неявно предполагалось, что метрика h в нем обладала тем свойством, что $h(i, j) > 0$, если $i \neq j$, в такой метрике любая функция непрерывна.

Пример. Работа ЭВМ в режиме «с разделенным процессором». Пусть в ЭВМ через независимые, подчиняющиеся распределению $F(x) = P(\theta_i \leqslant x)$ промежутки времени $\{\theta_i\}$ поступают требующие решения задачи, причем длительность решения i-й задачи равна η_i. Пусть $\{\eta_i\}$ также последовательность независимых одинаково распределенных случайных величин с функцией распределения $G(x) = P(\eta_i \leqslant x)$. Эта длительность является «потенциальной», т. е. i-я задача решалась бы за время η_i в том случае, если бы процессор все время «был занят» только ей. Однако это не всегда так. Поэтому уместнее придать величине η_i «энергетический» смысл, считая ее *работой*, которую должен выполнить процессор, чтобы решить задачу. Режим «с разделенным процессором» состоит в том, что процессор тратит свои возможности «поровну» между всеми наличными задачами. Реализуется же это, разумеется, последовательным предоставлением «очень малых» квантов времени каждой задаче.

В соответствии с этим, если в момент t имелась всего одна задача, которая требовала работы η_1, то через время Δt $(\Delta t < \eta_1)$ эта задача будет требовать работы $\eta_1 - \Delta t$. Если в момент t были две задачи (работы η_1 и η_2), то через время Δt $(\Delta t < 2\min(\eta_1, \eta_2))$ эти задачи будут требовать работ $\eta_1 - \frac{1}{2}\Delta t$, $\eta_2 - \frac{1}{2}\Delta t$ соответственно и т. д. Все это, разумеется, при условии, что за время Δt не поступило новых задач.

Построим марковский случайный процесс $\{Y_t\}$, описывающий функционирование такой системы. Состояние процесса можно было бы характеризовать числом ν_t задач, одновременно решаемых в системе. Однако процесс $\{\nu_t\}$, вообще говоря (если не предполагать показательность функций F и G), будет немарковским, а чтобы «сделать» его марковским, состояние ν_t нужно «дополнить» координатами, содержащими исчерпывающую информацию о предыстории, необходимую для задания будущего течения процесса. Один из возможных вариантов состоит в рассмотрении процесса, состояния которого имеют вид:

$$y = (\nu, x_1, x_2, \ldots, x_{\nu+1}), \ \nu = 0, 1, \ldots; \ x_i > 0.$$

Координаты $x_1, \ldots, x_{\nu+1}$ будем называть *дополнительными* и отметим, что их число в данном случае определяется *основной координатой* ν. Смысл последней уже был определен — это число одновременно решаемых задач. В качестве x_1 выберем время, оставшееся до поступления очередной задачи, в качестве $x_2, \ldots, x_{\nu+1}$ — количество работы, оставшейся до завершения соответственно первой, второй и т. д. — ν-й задачи.

Таким образом, множество V состояний процесса имеет вид объединения $V = \bigcup_{m=1}^{\infty} \Gamma^{(m)}$, где $\Gamma^{(m)}$ — первый ортант* $(m+1)$-мерного евклидова пространства. В нем можно ввести расстояние h, например, по формуле

$$h[(\nu, x_1, \ldots, x_{\nu+1}), (\mu, y_1, \ldots, y_{\mu+1})] =$$
$$= \left[\delta_{\nu\mu} + \sum_{i=1}^{\infty} (x_i - y_i)^2\right]^{1/2}, \tag{2.5}$$

где полагается, что $x_i = 0$ при $i > \nu + 1$, $y_j = 0$ при $j > \mu + 1$, а $\delta_{\nu\mu}$ — символ Кронекера, т. е. $\delta_{\nu\mu} = 1$, если $\nu = \mu$, и $\delta_{\nu\mu} = 0$, если $\nu \neq \mu$.

Траектории процесса $\{Y_t\}$ в соответствии с данным выше содержательным описанием имеют *кусочно-линейный* вид. Если в момент $t = 0$ $Y_0 = (\nu, x_1, \ldots, x_{\nu+1})$, то при $0 \leqslant t < t_1 = \min(x_1, \nu x, \ldots, \nu x_{\nu+1})$ Y_t изменяется детерминированно:

$$Y_t = \left(\nu, x_1 - t, x_2 - \frac{t}{\nu}, \ldots, x_{\nu+1} - \frac{t}{\nu}\right).$$

Момент t_1 является моментом скачка траектории. Если $\lim_{t \to t_1 - 0} x_1(t) = 0$, то момент t_1 является моментом поступления в систему очередной задачи и

$$Y_{t_1} = \left(\nu + 1, \theta, x_2 - \frac{t_1}{\nu}, \ldots, x_{\nu+1} - \frac{t_1}{\nu}, \eta\right),$$

где независимые случайные величины θ и η не зависят от «предыстории» и распределены по законам F и G соответственно. Если $\lim_{t \to t_1 - 0} x_i(t) = 0$, $1 < i \leqslant \nu + 1$, то момент t_1 является моментом окончания решения $(i-1)$-й задачи и

$$Y_{t_1} = \left(\nu - 1, x_1 - t_1, x_2 - \frac{t_1}{\nu}, \ldots, x_{i-1} - \frac{t_1}{\nu},\right.$$

---

* По аналогии с первым квадрантом (в случае плоскости), первым ортантом многомерного пространства называется его подмножество, в котором все координаты положительны.

$$x_{i+\hat{1}} - \frac{t_1}{\nu}, \ldots, x_{\nu+\hat{1}} - \frac{t_1}{\nu}\Big).$$

Далее вновь происходит линейное детерминированное изменение состояния процесса, определяется новый момент скачка t_2 и т. д. Строго говоря, следовало бы задать изменения процесса и в моменты «обнуления» более одной координаты, но мы этого делать не будем, пренебрегая такой возможностью.

Процесс $\{Y_t\}$ является марковским (так как знание его текущего состояния позволяет точно определить следующий момент скачка, величина которого не зависит от предыстории $\{Y_t\}$ и т. д.) переменной размерности.

Данный процесс задан нами *конструктивно*, т. е. были описаны возможные траектории процесса. Однако, пользуясь этим конструктивным описанием, можно записать и соответствующие переходные вероятности. В силу «детерминированности» траектории между скачками имеем:

$$P\{(\nu, x_1, \ldots, x_{\nu+\hat{1}}), t, B\} = \chi_B\Big(\nu, x_1 - t, x_2 - \frac{t}{\nu}, \ldots, x_{\nu+\hat{1}} - \frac{t}{\nu}\Big), \tag{2.6}$$

где множество $B \subset V$, а

$$\chi_B(y) = \begin{cases} 1, y \in B, \\ 0, y \notin B. \end{cases}$$

Казалось бы, что поскольку в формуле (2.6) переходная функция задана для каждого состояния при всех «достаточно малых» t (ведь все $x_i > 0$), то из уравнения Колмогорова — Чэпмена эту функцию можно найти и для всех t. Однако это не так. Дело в том, что для разных состояний интервал времени, на котором определена функция (2.6), различен — чем ближе координаты x_i к 0, тем меньше этот интервал. Именно это обстоятельство и не дает возможности найти из уравнения Колмогорова — Чэпмена переходную функцию для всех t. Содержательно данный математический факт также очевиден — формула (2.6) не учитывает скачкообразных переходов процесса.

Из сказанного выше вытекает, что в данном случае инфинитезимальный оператор (вид которого мы приводить не будем) не задает переходную функцию, которая будет полностью определена лишь после учета скачков.

Предположим теперь, что в рассматриваемом примере задания поступают в соответствии с пуассоновским потоком * (параметр λ). Тогда для «обеспечения» марковского свойства координата x_1 является «лишней». В этом случае инфинитезимальный оператор процесса $\{Y_t\}$ вновь не определяет однозначно его переходную функцию, так как не «отражает» скачки процесса вследствие «обнуления» координат y_i.

Обобщая данный пример, определим теперь *линейчатый марковский процесс*. Состояние процесса имеет вид $y = (\nu, x_1, \ldots, x_{\|\nu\|})$, где *основная координата* ν принимает значения из не более чем счетного множества $\mathcal{N}$. Например, $\mathcal{N} = \{0, 1, 2, \ldots\}$. Количество дополнительных координат, обозначаемое $\|\nu\|$ ($\|\nu\| \geqslant 0$), зависит, вообще говоря, от значения координаты ν. Множество допустимых значений вектора $(x_1, \ldots, x_{\|\nu\|})$ представляет собой выпуклый многогранник (без границы) $\Gamma^{(\nu)}$ в евклидовом пространстве размерности $\|\nu\|$ (в примере $\Gamma^{(\nu)}$ — первый ортант соответствующего пространства). Таким образом, $V = \bigcup\limits_{\nu\in\mathcal{N}} \Gamma^{(\nu)}$. Расстояние h в этом пространстве может быть определено по формуле (2.5). Для каждого $\nu \in \mathrm{N}$ задается вектор $v^{(\nu)} = (v_1^{(\nu)}, \ldots, v_{\|\nu\|}^{(\nu)})$ размерности $\|\nu\|$ *скоростей изменения дополнительных координат*. Кроме того, на V задана *функция λ интенсивностей скачков и вероятность перехода $P_1(y, B)$* из точки $y \in V$ в множество $B \subset V$ в результате этого скачка. Динамика процесса $\{Y_t\}$ описывается следующим образом. Если $Y_t = (\nu, x_1, \ldots, x_{\|\nu\|})$, $\|\nu\| \neq 0$, $\min\limits_{i\leqslant\|\nu\|} x_i > 0$, то за малое время Δt возможны следующие переходы: а) в состояние $(\nu, x_1 + v^{(\nu)}\Delta t, \ldots, x_{\|\nu\|} + v_{\|\nu\|}^{(\nu)}\Delta t)$ с вероятностью $1 - \lambda(\nu, x_1, \ldots, x_{\|\nu\|})\Delta t + o(\Delta t)$; б) в множество состояний $B \subset V$ с вероятностью $\lambda(\nu, x_1, \ldots, x_{\|\nu\|})\, P_1(\nu, x_1, \ldots, x_{\|\nu\|}; B)\, \Delta t + o(\Delta t)$ (при этом мы не сделали необходимых оговорок относительно, например, непрерывности функции λ и т. п., ограничи-

* То есть в этом случае функция распределения F интервалов между поступающими задачами имеет экспоненциальный (показательный) вид: $F(x) = 1 - e^{-\lambda x}$. Величина $1/\lambda$ имеет смысл средней длины промежутка между поступающими задачами. Характерной особенностью пуассоновского потока является то, что функция распределения времени ожидания очередного события (в данном случае задачи) не зависит от времени, прошедшего с момента наступления последнего события,— свойство отсутствия последействия.

ваясь лишь приближенным видом приводимых соотношений).

Если же $\|\nu\| = 0$, то $Y_{t+\Delta t} = \nu$ с вероятностью $1 - \lambda(\nu)\Delta t + o(\Delta t)$ и $P\{Y_{t+\Delta t} \in B\} = \lambda(\nu) P_1(\nu; B)\Delta t + o(\Delta t)$.

Далее, обозначим через $\gamma^{(\nu)}$ границу множества $\Gamma^{(\nu)}$. Зададим при $y \in \gamma^{(\nu)}$ функции $P_2(y; B)$, $B \subset V$, *определяющие скачки процесса* $\{Y_t\}$ при выходе на границу. Именно если $Y_u = (\nu, x_1, \ldots, x_{\|\nu\|}) \in V$ и $x^* = (x_1 + v_1^{(\nu)}(t-u), \ldots, x_{\|\nu\|} + v_{\|\nu\|}^{(\nu)}(t-u)) \in \gamma^{(\nu)}$ при некотором $t > u$, то положим $P(Y_t \in B) = P_2\{(\nu, x^*); B\}$.

Приведенный выше пример показывает, какой вид могут иметь вероятности перехода P_1 и P_2. Хотя линейчатые процессы определяются достаточно громоздко, тем не менее они находят широкое применение при описании элементов сложных систем, поскольку представляют собой содержательное обобщение многих используемых частных конструкций [10, 27], допускают исследование аналитическими методами [5, 10, 19, 27], а также могут служить основой для построения универсальных имитационных систем [8—10, 16, 20, 21].

Случайные процессы из еще одного класса, который будет рассмотрен, называются р е г е н е р и р у ю щ и м и и определяются ниже. Но предварительно рассмотрим примеры.

П р и м е р. Ц е п ь М а р к о в а. Рассмотрим конечную цепь Маркова $\{Y_t\}$ с множеством состояний $V = \{0, 1, 2, \ldots, N\}$ и дискретным временем $T = \{0, 1, 2, \ldots\}$. Предположим, что (рис. 3) $Y_0 = 0$ (пусть состояние 0 — *возвратно*). Отметим на оси времени последовательные моменты S_i, $i \geqslant 0$, такие, что $Y_{S_i} = 0$ (очевидно, по предположению, $S_0 = 0$). Поскольку «будущее» течение марковской цепи зависит лишь от его «настоящего» состояния, то «хвосты» процесса $\{Y_t\}$ при $t \geqslant S_i$ (т. е. процессы вида $Y_u^{(i)} = Y_{S_i+u}$) представляют собой *вероятностные копии* друг друга в том смысле, что их переходные функции, а следовательно, и всевозможные конечно-мерные распределения совпадают. Отсюда получаем, что и длительности интервалов $\theta_i = S_i - S_{i-1}$ независимы и одинаково распределены. Поэтому марковскую цепь можно интерпретировать следующим образом. Зададим совокупность независимых одинаково распределенных пар (циклов) $(\{Y_t^{(i)}\}, \theta_i)$, содержательно представляющих собой «куски» исход-

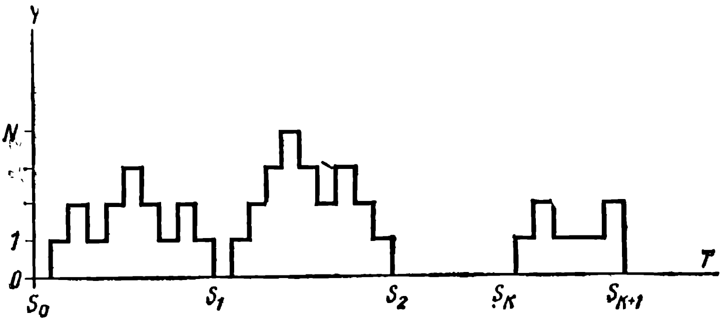


Рис. 3.

ной цепи Маркова, рассматриваемые лишь до момента θ_i первого возвращения ее в состояние 0, т. е. при $t < \theta_i$. И будем полагать $Y_t = Y^{(i)}_{t-S_i}$, если $S_i \leqslant t < S_{i+1}$, или, обозначая

$$N(t) = \max \{i : S_i \leqslant t\}, \tag{2.7}$$

можно записать

$$Y_t = Y^{N(t)}_{t-S_{N(t)}}. \tag{2.8}$$

Следовательно, мы рассматриваем цепь Маркова путем выделения таких *случайных* моментов времени S_i, в которых, как говорят, происходит *регенерация* процесса, т. е. дальнейшее его течение является вероятностной копией всего процесса.

Пример. Процесс рождения и гибели. Аналогичное представление можно получить для рассмотренного выше процесса рождения и гибели, если выбрать, скажем, в качестве S_i моменты попадания процесса в некоторое фиксированное состояние, например 0.

Пример. Линейчатый процесс. Если выбрать в линейчатом процессе в качестве S_i моменты попадания в какое-либо состояние, такое, что $\|\nu\| = 0$, то процесс $\{Y_t\}$ также может быть интерпретирован в терминах циклов. Соответствующие построения полностью повторяют сделанные для цепей Маркова. Однако, например, при рассмотрении системы «с разделенным процессором» (для закона F общего вида) состояний ν, для которых $\|\nu\| = 0$, вообще не было. В этом случае за моменты S_i можно было бы принять моменты «выхода» из состояния $\nu = 0$ (отметим, что в рассматриваемом примере $\|\nu\| = \nu + 1$ и, следовательно, $\|0\| = 1$), т. е. такие моменты t, что

$$\lim_{u \to t-0} \nu(u) = 0, \quad \lim_{u \to t-0} x_1(u) = 0.$$

Вообще марковские процессы часто можно рассматривать как регенерирующие, выбирая за точки регенерации последовательные моменты возвращения процесса в некоторое фиксированное состояние. Однако для этого выделенное состояние должно обладать тем свойством, что время возвращения в него определено и конечно. Например, состояния вида $(\nu, x_1, \ldots, x_{\|\nu\|})$, $\|\nu\| \geqslant 1$ для линейчатого процесса будут, вообще говоря, невозвратными, а любое состояние для броуновского движения, хотя и возвратно, но обладает той особенностью, что за моментом «попадания» в него следует целая «пачка» аналогичных моментов, расстояние между которыми равно нулю. Поэтому для броуновского движения за моменты регенерации следовало бы выбрать, например, моменты попадания в состояние a после пребывания его в состоянии $b \neq a$.

В приведенных до сих пор примерах рассматривались лишь марковские процессы. Однако нетрудно понять, что, скажем, для линейчатого процесса «немарковская» составляющая $\nu(t)$ также будет обладать теми же точками регенерации, что и исходный процесс $\{Y_t\}$.

Сказанное приводит нас к следующему *определению*. Пусть задана совокупность независимых одинаково распределенных пар $(\{Y_t^{(i)}\}, \theta_i)$, называемых *циклами*, первый элемент которой представляет собой случайный процесс, а второй — случайную величину. Отметим, что, вообще говоря (см. примеры), θ_i и $\{Y_t^{(i)}\}$ *зависимы*. Процесс $\{Y_t\}$, определенный формулой (2.8), назовем *регенерирующим*, а моменты $S_i = \theta_1 + \ldots + \theta_i$ — *моментами регенерации*.

Таким образом, способ узнать, является ли процесс $\{Y_t\}$ регенерирующим, состоит в определении возможности представить его в виде (2.8). Очевидно, что такое представление (в случае его существования) может быть не единственным. Искусство исследователя состоит как раз в доказательстве существования представления исходного процесса в виде (2.8) и его нахождении.

Свойство «регенерируемости» позволяет свести изучение сложного, может быть, немарковского процесса $\{Y_t\}$ к изучению последовательности независимых одинаково распределенных элементов $(\{Y_t^{(i)}\}, \theta_i)$, что позволяет воспользоваться хорошо разработанным аппаратом анализа таких последовательностей. Более того, успех исследования здесь в значительной мере определяет последовательность $\{\theta_i\}$ независимых одинаково распределенных случайных величин, называемая *вложенным процессом восстановления* и представляющая собой классический объект изучения теории вероятностей.

3. ЗАДАЧИ И МЕТОДЫ АНАЛИЗА СЛОЖНЫХ СИСТЕМ

После того как мы выяснили основные особенности функционирования сложных систем и привели некоторые математические модели элементов таких систем, рассмотрим задачи, возникающие при исследовании указанных моделей, а также соответствующие математические методы. Основное внимание, естественно, будет уделено содержательной интерпретации рассматриваемых задач. Эти задачи условно можно разделить на два типа. К первому отнесем задачи исследования *качественных свойств* изучаемых моделей. Ко второму — задачи нахождения *количественных соотношений*, характеризующих модели.

Под качественными свойствами будем понимать свойства, позволяющие судить о поведении траектории системы «в целом». Например, уходит ли траектория «в бесконечность» за конечное время, ограничена ли она в каком-либо смысле, возвращается ли она в то или иное подмножество пространства состояний и т. д. Перечисленные свойства относятся к поведению *отдельных* траекторий. Можно рассмотреть также качественные свойства, характеризующие поведение *совокупностей* траекторий. Важнейшим примером такого свойства является непрерывность, наличие которой говорит о том, что при малых изменениях параметров, определяющих работу модели, показатели, характеризующие поведение системы, также изменяются мало. Нетрудно увидеть аналогию между перечисленными свойствами и соответствующими качественными свойствами решений дифференциальных уравнений. Можно с уверенностью утверждать, что в случае сложных систем «вес» качественных исследований возрастает, поскольку результат исследований, с одной стороны, дает информацию, скажем, о принципиальной возможности создать систему с требуемыми свойствами (или о предполагаемом характере ее поведения), а с другой — эта информация зачастую бывает единственно необходимой исследователю в силу невозможности интерпретации громоздких числовых данных, количественно описывающих поведение системы.

Сказанное отнюдь не означает призыва к отказу от количественных исследований сложных систем. Более того, перечисленные качественные исследования могут принести реальную пользу лишь в том случае, если сопровождаются количественными оценками исследуемых свойств. Важно подчеркнуть, что количественные исследования становятся целенаправленными, а следовательно, и эффективными только после выявления характерных особенностей поведения анализируемой системы. Поэтому качественные и количественные исследования дополняют друг друга.

Регулярность. Построение математической модели реального объекта неизбежно приводит к неадекватности описания, причиной чего является невозможность учета многообразия реального мира в математических моделях. Как правило, проверка адекватности трудоемка, в значительной степени неформальна и состоит в сопоставлении результатов анализа модели с экспериментальными данными, следствиями физических законов, имеющимися наблюдениями и т. д. Очевидно, что использование объективных закономерностей функционирования модели способствует более эффективному анализу адекватности. В частности, свойство регулярности может служить одной из основ *анализа адекватности*.

Рассмотрим процесс $\{Y_t\}$ любого из типов, приведенных выше. Возьмем «бесконечно расширяющуюся» систему концентрических многомерных шаров радиуса n (обозначим их B_n) с центром в некоторой точке $y_0 \in V$. Определим случайные моменты времени τ_n первого выхода траектории из шаров B_n (рис. 4). Назовем величину $\tau = \lim\limits_{n \to \infty} \tau_n$ *временем ухода процесса* $\{Y_t\}$ *в бесконечность*. Это время может быть как конечным, так и бесконечным.

Аналогично, отметим на оси времени последовательные скачки процесса $\{Y_t\}$ и обозначим их ξ_n, $n = 1, 2, \ldots$, $\xi_1 < \xi_2 < \ldots$ Здесь также, очевидно, возможны два случая, когда величина $\xi = \lim\limits_{n \to \infty} \xi_n$ либо конечна, либо бесконечна. Назовем величину ξ *временем накопления скачков*. Процесс $\{Y_t\}$ назовем *регулярным*, если $\xi < \infty$ и $\tau < \infty$ (с вероятностью 1). В противном случае процесс назовем *нерегулярным*. Нерегулярное поведение, как правило, *невозможно* в реальных системах, поскольку уход координат в бесконечность за конечное время или наличие конечной точки накопления скачков свидетельствует о бесконечной интенсивности расхода каких-либо ресурсов. Иначе говоря, «нерегулярные» модели нереализуемы физически и, следовательно, не могут быть адекватны никакому реальному процессу: проверка модели на регулярность позволит либо отсеять заведомо неадекватные модели, либо очертить область, где данная модель может быть использована. Нерегулярность модели является признаком того, что в ней, возможно, не нашел отражения ряд реальных факторов, существенно влияющих в рассматриваемых условиях на поведение модели, или динамика системы не отвечает динамике модели. В обоих случаях необходима коррекция исходной модели.

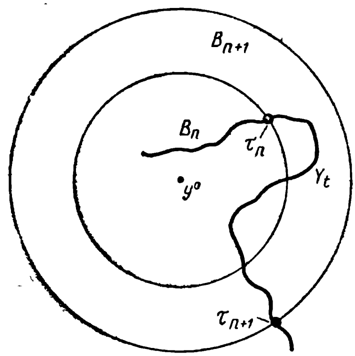


Рис. 4.

Полезно отметить, что моменты τ и ξ могут как совпадать друг с другом (так будет, например, для процесса рождения и гибели ввиду «кусочно-постоянного» вида траекторий), так и, вообще говоря, отличаться (например, у линейчатого процесса может быть $\xi < \infty$ и $\tau = \infty$).

Проиллюстрируем один способ нахождения условий регулярности на примере процесса рождения и гибели, введенного в п. 2. В качестве шаров B_n выберем $B_n = \{0, 1, \ldots, n - 1\}$, $n=1, 2, \ldots$ Тогда поскольку до момента τ_n выхода из множества B_n у процесса $\{Y_t\}$ не может быть точки накопления скачков (т. е. $\xi > \tau_n$), то и вообще $\xi \geqslant \tau$, и, следовательно, достаточно найти условия, при которых $\tau = \infty$ — это и будут условия регулярности. Поскольку $\tau = \lim\limits_{n\to\infty} \tau_n$, то для доказательства факта, что $P(\tau = \infty) = 1$, достаточно показать, что при любом t $P(\tau_n \leqslant t) \to 0$ при $n \to \infty$.

Рассмотрим вместо исходного процесса $\{Y_t\}$ процесс $\{Y_t^{(n)}\}$, который совпадает с исходным до момента τ_n выхода из множества B_n, а начиная с момента τ_n «останавливается» и принимает значение n. Иначе говоря, процесс

$\{Y_t^{(n)}\}$ — это процесс рождения и гибели, у которого состояние n сделано поглощающим (т. е. для него $\lambda_n = \mu_n = 0$). Очевидно, время τ_n выхода из B_n для обоих процессов одинаково. В силу приведенных в п. 2 формул вид инфинитезимальных операторов $A\, f\,(i)$ и $A^n\, f\,(i)$ для процессов $\{Y_t\}$ и $\{Y_t^{(n)}\}$ соответственно совпадает при $i < n$, а $A^{(n)}\, f\,(n) = 0$.

Для чего понадобилось вводить процесс $\{Y_t^{(n)}\}$? В п. 2 мы говорили о том, что инфинитезимальный оператор, вообще говоря, не определяет марковского процесса (и, в частности, процесса рождения и гибели $\{Y_t\}$). Однако у процесса $\{Y_t^{(n)}\}$ конечное число состояний $\{0, 1, \ldots, n\}$, и в этом случае, конечно, все в порядке. Пусть для определенности начальное условие $Y_0 = 0$, и далее мы его будем подразумевать, всякий раз особо не оговаривая.

Попытаемся подобрать в формуле (2.3) (примененной к процессу $\{Y_t^{(n)}\}$) функцию f так, чтобы извлечь из этой формулы необходимое нам утверждение относительно τ_n. Предположим, что удалось выбрать f, удовлетворяющую условиям:

а) $\lim\limits_{n\to\infty} f\,(n) = \infty$, $f\,(k) \geqslant 0$ при всех $k = 0, 1, \ldots$;

б) $\sup\limits_k Af\,(k) \leqslant \Delta < \infty$

(где A — инфинитезимальный оператор процесса $\{Y_t\}$, определяемый формулой $Af(k) = \lambda_k\,[f\,(k+1) - f\,(k)] + \mu_k\,[f\,(k-1) - f\,(k)]$). Тогда $A^{(n)}f\,(k) \leqslant \Delta < \infty$ при всех $k < n$ и $A^{(n)}\, f\,(n) = 0$. Из формулы (2.3) поэтому при любом t получаем неравенство

$$Ef\left(Y_t^{(n)}\right) = f\,(0) + E\int\limits_0^t A^{(n)}f\left(Y_u^{(n)}\right) du \leqslant f\,(0) + \Delta t. \qquad (3.1)$$

Но в момент t состояние $Y_t^{(n)} = n$ с вероятностью $P\,(\tau_n \leqslant t)$ и $Y_t^{(n)} < n$ с вероятностью $P\,(\tau_n > t)$. Поэтому $Ef\,(Y_t^{(n)}) \geqslant f\,(n) P\,(\tau_n \leqslant t)$. Используя (3.1), получаем

$$P\,(\tau_n \leqslant t) \leqslant (f\,(0) + \Delta t)/f\,(n), \qquad (3.2)$$

откуда в силу условия а), наложенного на f, находим, что $P\,(\tau_n \leqslant t) \to 0$ при $n \to \infty$, это и является условием регулярности. Весь вопрос теперь заключается в возможности построения функции f, удовлетворяющей условиям а) и б). Рассмотрим функцию f, такую, что $f(0) = 0$, $f(k) = \sum_{j<k} \frac{1}{\lambda_j \rho_j} \sum_{i=0}^{j} \rho_i$.

Очевидно, она удовлетворяет условию а), если выполнено (2.2). Непосредственная проверка по приведенной выше формуле убеждает, что $Af(k) = 1$ при всех k. Таким образом, условие б) также выполнено, и, следовательно, если процесс $\{Y_t\}$ удовлетворяет условию (2.2), то он регулярен.

Суть приведенного построения заключается, по существу, в замене исходного процесса $\{Y_t\}$ другим — $\{f(Y_t)\}$, который в некоторых отношениях устроен проще исходного (например, его средняя скорость изменения ограничена равномерно по состояниям) и для которого момент τ также является моментом ухода в бесконечность. Такая замена широко используется при изучении классических дифференциальных уравнений, составляя, например, основу второго метода Ляпунова анализа различных качественных свойств [31]. Применительно к стохастическим процессам, в частности к линейчатым процессам, построения, обобщающие приведенные, содержатся в книге [19] *. Естественно, что при этом необходимо дополнительно позаботиться о том, чтобы скачки не накапливались из-за выхода траекторий на границу множеств $\Gamma^{(\nu)}$ — как раз эти скачки «не учитываются» инфинитезимальным оператором линейчатого процесса.

Если рассматривается регенерирующий процесс, то, очевидно, его регулярность порождается регулярностью «составляющих подпроцессов» $\{Y_t^{(i)},\ t < \theta_i\}$. Важно отметить, что выше анализ регулярности фактически был сведен к оценке распределения некоторого *времени первого достижения* (τ_n). Оказывается, что вообще подобные оценки играют ключевую роль в изучении и других качественных свойств [19].

Свойства, связанные с временем первого достижения множества. Рассмотрим в множестве состояний V некоторое фиксированное подмножество $Q \subset V$. Пусть τ_Q — время первого достижения множества Q траекторией процесса $\{Y_t\}$ (если подмножество Q

* Соответствующий метод исследования получил название *метода пробных функций*, поскольку одним из этапов его применения является подбор функции состояния анализируемого процесса, удовлетворяющей определенным условиям.

недостижимо, то будем полагать $\tau_Q = \infty$). Величина τ_Q является случайной, поэтому далее будет идти речь о ее вероятностных свойствах. Наибольшее применение свойства времени первого достижения получили при изучении моделей, описываемых марковскими процессами. В этом случае целесообразно рассматривать условное время первого достижения $\tau_Q(y)$ при $Y_0 = y$. Далее мы предположим, что $\{Y_t\}$ — марковский процесс. В зависимости от целей исследования интерес могут представлять различные свойства времени τ_Q. Например, в теории надежности одной из основных характеристик является функция распределения времени безотказной работы системы. Очевидно, что τ_Q можно трактовать как такое время, если в качестве Q выбрать подмножество «отказовых» состояний. В этом случае интерес представляет как оценка указанной функции распределения, так и различные ее числовые характеристики (моменты, квантили и т. д.).

Если говорить о прикладном интересе, то для функции распределения $P(\tau_Q \leqslant x)$ наиболее важна оценка сверху, а, скажем, для среднего $E\tau_Q$ времени безотказной работы — оценка снизу. Данные оценки являются «пессимистическими», т. е. занижающими реальные показатели надежности системы.

Рассмотрим другой пример — модель, описываемую цепью Маркова. Как известно [24, 37], для таких моделей одной из основных характеристик является время возвращения в фиксированное состояние, которое также можно трактовать, как время достижения, если в качестве Q выбрать одноточечное множество, состоящее из рассматриваемого состояния. Обозначим такое состояние через i и пусть $\tau(i)$ — соответствующее время возвращения. Тогда, если $P(\tau(i) = \infty) > 0$, т. е. траектория процесса с положительной вероятностью не возвращается в состояние i, то это состояние называется *невозвратным*. Если $P(\tau(i) < \infty) = 1$, т. е. время возвращения в состояние i конечно, то это состояние называется *возвратным*. Если при этом и среднее время возвращения конечно ($E\tau(i) < \infty$), то состояние i называется положительным, и в этом случае можно говорить о *стационарной вероятности* состояния i, равной средней доле времени, проводимой цепью в состоянии i. Если к тому же состояние i — *непериодическое* *, то существует *предельная вероятность* $\lim\limits_{t\to\infty} P(Y_t = i)$. Таким образом, в данном случае свойства времени возвращения — основа классификации состояний цепи Маркова. И здесь больший интерес представляют верхние оценки, например, среднего времени возвращения $E\tau(i)$. Указанное обстоятельство имеет место не только для цепей Маркова, но и для других, более сложных процессов.

Таким образом, различные оценки одной и той же величины приводят к различным математическим задачам. Целесообразно подчеркнуть, что в перечисленных примерах самым тесным образом переплелась необходимость исследования качественных свойств величины τ_Q (например, конечности τ_Q, ее моментов и т. д.) и получения количественных оценок распределения τ_Q, моментов этого распределения и др.

Один из способов получения верхних оценок для функций распределения величин τ_Q фактически уже был проиллюстрирован выше (см. (3.2) на примере процесса рождения и гибели, где необходимо положить $Q = \{n, n + 1, \dots\}$ и $\tau_Q = \tau_n$. Данный способ использует свойства пробной функции f, позволяющие достаточно просто получать оценки, используя формулу (2.3). Оказывается, что этот способ распространяется на достаточно широкий класс марковских процессов, например линейчатые [19]. Более того, он позволяет получать не только верхние, но и в ряде случаев нижние оценки функции распределения τ_Q. Достоинством данного метода является то, что он дает возможность получать оценки без каких-либо «асимптотических» предположений — о «высокой» надежности, «малых» потерях, «высоком» уровне и т. п. В то же время получаемые оценки существенным образом зависят от вида используемой пробной функции. Такое положение ведет, с одной стороны, к возможности улучшения оценок за счет вариации пробных функций, а с другой — к необходимости классификации процессов, в соответствии с которой мож-

---

* Состояние i называется *периодическим* с периодом $d>1$, если величина $\tau(i)$ может принимать лишь значения вида d, $2d$, $3d, \dots$ и d — максимальное число, удовлетворяющее этому свойству. Например, если $\tau(i)$ с ненулевой вероятностью принимает лишь значения 8; 16; 20, то i — периодическое состояние с периодом $d=4$. Иначе говоря, d — наибольший общий делитель возможных значений $\tau(i)$.

но было бы искать пробные функции в определенном виде.

Способ оценки сверху среднего времени достижения $E\tau_Q$ необходимый, как было сказано, например, при анализе цепей Маркова, проще всего пояснить именно на примере однородной марковской цепи $\{Y_t\}$ с дискретным временем ($t=0, 1, 2, \ldots$) и множеством состояний $V = \{0, 1, 2, \ldots\}$.

Пусть для определенности множество Q состоит из одной точки $0 : Q = \{0\}$. Обозначим через p_{ij} переходные вероятности рассматриваемой цепи. Предположим, что задана неотрицательная функция f на множестве V, которая удовлетворяет условию

$$Af(i) = \sum_{j=0}^{\infty} p_{ij} f(j) - f(i) \leqslant -\Delta < 0, i \neq 0. \tag{3.3}$$

Величина $Af(i)$ в формуле (3.3) имеет смысл среднего одношагового приращения функции f при начальном состоянии цепи i, а оператор A (дискретный аналог инфинитезимального оператора) называется *производящим*.

Аналогично формуле (2.3) для дискретного времени справедлива формула

$$E\{f(Y_t)/Y_0 = y\} = f(y) + E\left\{\sum_{k=0}^{t-1} Af(Y_k)/Y_0 = y\right\}, \tag{3.4}$$

которая в данном случае легко доказывается по индукции.

Если, как и при рассмотрении свойства регулярности, ввести процесс $\{Y'_t\}$, совпадающий с $\{Y_t\}$ до момента попадания в 0 и остающийся в состоянии 0 «навечно» *, то для него производящий оператор (обозначим его A') имеет вид:

$$A'f(i) = Af(i), \ i \neq 0, \ A'f(0) = 0. \tag{3.5}$$

Но тогда в силу неотрицательности f условия (3.3) и формул (3.4), (3.5) имеем:

$$0 \leqslant E\{f(Y'_t)/Y'_0 = y\} \leqslant f(y) - \Delta E\{\min(\tau_Q, t-1)/Y'_0 = y\},$$

откуда при $t \to \infty$ получаем оценку

$$E(\tau_Q/Y_0 = y) \leqslant f(y)/\Delta, \tag{3.6}$$

которая, помимо всего, доказывает и конечность среднего времени попадания цепи в состояние 0.

Таким образом, существование неотрицательной функции f, удовлетворяющей условию (3.3), достаточно (а мож-

---

* Т. е. переходные вероятности p'_{ij} цепи $\{Y'_t\}$ определяются так: $p'_{ij} = p_{ij}$, $i \neq 0$, $j = 0, 1, 2, \ldots$; $p'_{00} = 1$; $p'_{0j} = 0$, $j \geqslant 1$.

но показать, что и необходимо) для конечности среднего времени достижения точки 0. Отметим, что если условие (3.3) дополнить таким

$$Af(0) < \infty, \tag{3.7}$$

то будет конечным также и среднее время *возвращения* цепи в состояние 0.

Условия (3.3) и (3.7) известны в теории цепей Маркова достаточно давно [42] в качестве критерия эргодичности. Однако использование формулы (3.4) позволило не только упростить доказательство, получить количественные оценки вида (3.6), но и распространить приведенные здесь результаты, например на цепи Маркова с произвольным множеством состояний, линейчатые марковские процессы. Кроме того, были получены условия конечности не только средних, но и моментов вида $E\tau_Q^\gamma$, $\gamma > 0$, а также соответствующие оценки. Монография [19] содержит примеры применения указанных результатов к анализу стохастических систем (резервированных, массового обслуживания, передачи данных и др.).

Если рассматривать регенерирующие процессы, то для них одним из основных свойств является конечность среднего времени цикла $E\theta_i$, из которого практически следует существование предельного (при $t \to \infty$) распределения состояния Y_t этого процесса — те дополнительные ограничения, касающиеся интегрируемости некоторых функций, которые необходимо налагать, на практике всегда выполняются. Весьма часто точки регенерации совпадают с моментами попадания траектории процесса в какое-либо состояние, т. е. здесь также основную роль играют времена первого достижения или их аналоги.

Непрерывность. В общем случае работа элементов сложных систем может быть интерпретирована как *преобразование* **F** набора *исходных данных* **X** в набор *выходных данных* **Y**, т. е. **Y** = **F** (**X**). В качестве исходных данных **X**, например, для систем обслуживания могут рассматриваться распределения, задающие входящие потоки требований, длительности обслуживания и т. д., для марковских процессов — переходные вероятности, для регенерирующих процессов, например, — законы, задающие их динамику на отдельных циклах. Аналогично в качестве выходных данных **Y** для системы обслуживания выбираются длительности ожидания, длины очередей, вероятности потери и т. д., а в общем случае, например, весь случайный процесс $\{Y_t\}$, описывающий динамику системы, или различные вероятностные распределения, характеризующие его. *Задача анализа* заключается в нахождении характеристик $\mathbf{Y}$ при известных исходных данных $\mathbf{X}$.

Однако при *реальных расчетах* показателей работы различных моделей по известным определяющим параметрам возникают проблемы, требующие дополнительных исследований. Эти проблемы порождаются следующими обстоятельствами.

1) Исходные данные $\mathbf{X}$ определяются в результате статистической обработки данных наблюдений и, следовательно, представляют собой лишь статистические оценки истинных распределений и параметров. В результате существует расхождение между истинными распределениями и теми параметрами, на основе которых проводятся необходимые расчеты. Следовательно, вместо отображения $\mathbf{X} \to \mathbf{Y}$ исследователь имеет дело с отображением $\mathbf{X}' \to \mathbf{Y}'$, где $\mathbf{X}'$ — исходные данные, полученные в результате оценки, а $\mathbf{Y}'$ — соответствующие выходные данные. Здесь «искажения» в исходных данных появляются в результате *погрешностей измерений и оценок.*

2) Исходные данные $\mathbf{X}$ полностью заданы, однако при численных (аналитических или машинных) расчетах использованы быть не могут, так как получающиеся вычислительные процедуры для нахождения $\mathbf{Y}$ оказываются слишком сложными или даже нереализуемыми. В этом случае для проведения расчетов обычно вместо $\mathbf{X}$ рассматривают другие исходные данные $\mathbf{X}'$, в известном смысле *аппроксимирующие* $\mathbf{X}$. И вновь вместо отображения $\mathbf{X} \to \mathbf{Y}$ рассматривается отображение $\mathbf{X}' — \mathbf{Y}'$. В данном случае «искажения» исходных данных привнесены математиком в целях *упрощения анализа модели.*

К этому же типу ситуаций относится и моделирование на ЭВМ, при котором возникают и дополнительные погрешности, вызванные конечной разрядностью представляемых в ЭВМ чисел.

В действительности имеет место совместное проявление отмеченных факторов.

Вообще говоря, нельзя исключить и возможность изменения вида отображения $\mathbf{F}$ при численных расчетах. Например, в вычислительном отношении может оказаться проще рассмотрение конечно-линейной системы обслуживания (с «большим» числом приборов) вместо бесконечнолинейной. Такая замена влечет изменение вида отображения $\mathbf{F}$. Тем не менее для простоты в дальнейшем мы будем предполагать вид отображения $\mathbf{F}$ неизменным.

В приведенных выше ситуациях естественным является вопрос: если исходные данные $\mathbf{X}$ и $\mathbf{X}'$ отличаются «мало», то «сильно» ли будут отличаться выходные данные $\mathbf{Y}$ и $\mathbf{Y}'$? Если «малые» отличия исходных данных влекут «малые» же отличия в выходных данных, то естественно говорить о *свойстве непрерывности* модели $\mathbf{Y} = \mathbf{F}(\mathbf{X})$.

Для более четкой постановки проблемы нам потребуются некоторые уточнения и пояснения введенных понятий.

Фундаментальным при изучении феномена непрерывности является этап выбора используемого *типа сходимости*. Именно прежде чем изучать непрерывность, необходимо:

1) в пространстве $\mathscr{X}$ рассматриваемых определяющих параметров, элементами которого являются как $\mathbf{X}$, так и $\mathbf{X}'$, ввести понятие *сходимости* (обозначение: $\underset{\mathscr{X}}{\to}$);

2) выделить *подмножество* $\mathscr{X}^* \subset \mathscr{X}$, которому принадлежат определяющие параметры $\mathbf{X}$ и $\mathbf{X}'$ и в пределах которого следует рассматривать непрерывность. Выделение $\mathscr{X}^*$ ограничивает рамки исследования, налагая на изучаемые модели дополнительные ограничения. Например, можно рассматривать лишь такие определяющие параметры, при которых у изучаемых моделей существует стационарный режим. Иногда при численных расчетах ограничиваются лишь функциями распределения, имеющими дробно-рациональные преобразования Лапласа и т. п.;

3) в пространстве $\mathscr{Y}$ изучаемых «выходных» данных, элементами которого являются $\mathbf{Y}$ и $\mathbf{Y}'$, ввести понятие сходимости (обозначение $\underset{\mathscr{Y}}{\to}$).

После того как «подготовительная работа» проделана, можно дать определение непрерывности.

Определение 1. Модель, описываемая отображением $\mathbf{Y} = \mathbf{F}(\mathbf{X})$, непрерывна в множестве $\mathscr{X}^*$, если для всех $\mathbf{X} \in \mathscr{X}^*$, $\mathbf{X}' \in \mathscr{X}^*$ выполнено соотношение $\mathbf{X}' \underset{\mathscr{X}}{\to} \mathbf{X} \Rightarrow \mathbf{Y}' \underset{\mathscr{Y}}{\to} \mathbf{Y}$.

Легко видеть, что свойство непрерывности в смысле определения 1 зависит от выбранных «допустимого множества» определяющих параметров $\mathscr{X}^*$ и типов сходимости $\underset{\mathscr{X}}{\to}$ и $\underset{\mathscr{Y}}{\to}$.

Поэтому понятие непрерывности относится не к изучаемой системе, а лишь к выбранным аспектам ее функционирования, и наличие или отсутствие непрерывности определяется как рассматриваемыми типами сходимости, так и изучаемыми характеристиками: *одна и та же характеристика может быть непрерывна в одном смысле и не быть непрерывной в другом*. Аналогично, *непрерывность одной характеристики не влечет, вообще говоря, непрерывности другой* (даже в смысле одного и того же типа сходимости).

Так, определенную непрерывность можно назвать *качественной*. Ее наличие позволяет лишь утверждать, что малые изменения в используемых условиях влекут за собой и малые изменения утверждений. Хотя констатация факта непрерывности и доставляет исследователю известную уверенность в использовании модели, но этим, по сути дела, возможность применения свойства «качественной непрерывности» исчерпывается.

Поэтому естественным представляется следующий шаг, цель которого установить *количественную зависимость* между малыми изменениями условий и утверждений. Эффективные оценки «количественной непрерывности» не только автоматически содержат утверждение о «качественной непрерывности», но и открывают путь к решению разнообразных практически важных проблем, таких, как

а) *оценка погрешности*, вносимой в расчеты при замене неизвестных вероятностных характеристик их статистическими оценками, и вследствие этого обоснование таких показателей статистических процедур, как объемы выборок, доверительные границы, уровни доверия и т. д.;

б) *выбор аппроксимирующей системы* при численных расчетах с учетом требуемой точности получаемых результатов и допустимой сложности вычислений;

в) *организация машинных экспериментов* при моделировании — выбор способов генерации определяющих случайных величин, методов сокращения машинного времени с помощью введения, например, зависимых испытаний и т. п.;

г) *проверка адекватности* используемой математической модели изучаемому реальному процессу;

д) *оценка коэффициентов чувствительности моделей* относительно вариации их параметров.

При решении задач «количественной непрерывности» введенные выше сходимости $\xrightarrow[\mathscr{X}]{}$ и $\xrightarrow[\mathscr{Y}]{}$ метризуются. Именно в пространствах $\mathscr{X}$ и $\mathscr{Y}$ вводятся метрики (расстояния) μ $(\mathbf{X}, \mathbf{X}')$ и ν $(\mathbf{Y}, \mathbf{Y}')$ соответственно, значения которых (неслучайные) определяются совместными вероятностными распределениями, задающими пары $(\mathbf{X}, \mathbf{X}')$ и $(\mathbf{Y}, \mathbf{Y}')$. Будем считать, что $\mathbf{X}' \underset{\mathscr{X}}{\to} \mathbf{X} \Leftrightarrow \mu(\mathbf{X}, \mathbf{X}') \to 0$; $\mathbf{Y}' \underset{\mathscr{Y}}{\to} \mathbf{Y} \Leftrightarrow \nu\ (\mathbf{Y}, \mathbf{Y}') \to 0$.

Тогда определение 1 можно «метризовать»:

Определение 2. Модель, описываемая отображением $\mathbf{Y} = \mathbf{F}(\mathbf{X})$ (μ, ν)-непрерывна в множестве $\mathscr{X}^*$, если для любого $\varepsilon > 0$ существует число $\delta = \delta(\varepsilon) > 0$, такое, что при всех $\mathbf{X} \in \mathscr{X}^*, \mathbf{X}' \in \mathscr{X}^*$, для которых $\mu(\mathbf{X}, \mathbf{X}') < \delta$, выполнено неравенство $\nu(\mathbf{Y}, \mathbf{Y}') = \nu(\mathbf{F}(\mathbf{X}), \mathbf{F}(\mathbf{X}')) \leqslant \varepsilon$.

Нахождение или оценка функции $\delta(\varepsilon)$ и представляет собой конечную цель количественного анализа непрерывности. Данная оценка, разумеется, зависит от метрик μ и ν, в терминах которых формулируется непрерывность. Поэтому одной из главных составляющих анализа непрерывности является изучение свойств используемых метрик, а также взаимоотношений между ними. Так как сравниваемыми элементами ($\mathbf{X}$ и $\mathbf{X}'$ или $\mathbf{Y}$ и $\mathbf{Y}'$) являются, как правило, случайные величины, последовательности или функции, то это должны быть метрики между соответствующими случайными элементами. В рамках данной брошюры уместно ограничиться лишь ссылкой на то, что наиболее полное и содержательное изучение вопросов, относящихся к свойствам и взаимоотношениям метрик в пространствах случайных элементов, можно найти в работе В. М. Золотарева [18].

Изложим один подход, позволяющий получать оценки непрерывности для широкого класса систем в случае, когда выходным показателем $\mathbf{Y}$ является случайный процесс $\mathbf{Y} = \{Y_t\}_{t \in T}$. Для простоты ограничимся лишь случаем дискретного времени: $t = 0, 1, 2, \ldots$ Относительно метрики ν предположим, что

$$\nu(\mathbf{Y}, \mathbf{Y}') = \sup_{t \geqslant 0} r(Y_t, Y_t'), \tag{3.8}$$

где r — некоторое расстояние между элементами последовательностей $\mathbf{Y}$ и $\mathbf{Y}'$. Что касается метрики μ, то вид ее нам пока неважен, как неважна и природа элементов $\mathbf{X}$ и $\mathbf{X}'$.

Если расстояние ν имеет вид (3.8), то определение 2 требует фактически *равномерной по времени непрерывности*:

$$\nu(\mathbf{Y}, \mathbf{Y}') = \sup_t r(Y_t, Y_t') \to 0 \text{ при } \mu(\mathbf{X}, \mathbf{X}') \to 0. \tag{3.9}$$

Проиллюстрируем это примером.

Предположим, что $\mathbf{X} = \{X_n\}_{n \geqslant 0}$ — последовательность независимых одинаково распределенных случайных величин с функцией распределения $H(x) = P(X_n \leqslant x)$, $-\infty < x < \infty$, а последовательность $\mathbf{Y} = \{Y_n\}_{n\geqslant 0}$ строится так:

$$Y_0 = 0,\ Y_{n+1} = \max(0,\ Y_n + X_n). \tag{3.10}$$

Пусть $G_n(x) = P(Y_n \leqslant x)$. Аналогичные обозначения будут использоваться и для процесса $\mathbf{Y}'$, только они будут отмечаться штрихами. Определенный таким образом процесс $\mathbf{Y}$ возникает при изучении различных систем обслуживания и управления запасами. Например, он описывает изменение времени ожидания в однолинейной системе с ожиданием, длины очереди в такой же системе, если либо входной поток пуассоновский, либо время обслуживания показательно и т. д. Пусть L — *расстояние Леви* между функциями распределения (расстояние Леви содержательно можно интерпретировать как длину стороны квадрата максимального размера, который может быть вписан между кривыми $A(x)$ и $B(x)$; стороны квадрата параллельны осям координат).

Если $L \to 0$, то говорят, что сравниваемые функции распределения *слабо сходятся* друг к другу.

Выберем

$$\mu(\mathbf{X},\ \mathbf{X}') = L(H,\ H'), \tag{3.11}$$

$$r(Y_n,\ Y'_n) = L(G_n,\ G'_n). \tag{3.12}$$

Тогда

$$\nu(\mathbf{Y},\ \mathbf{Y}') = \sup_n L(G_n,\ G'_n).$$

В основе анализа непрерывности лежит идея выявления у сравниваемых процессов $\mathbf{Y}$ и $\mathbf{Y}'$ «точек близости». Под ними понимаются, вообще говоря, случайные моменты времени σ_k, $k = 1, 2, \ldots$, обладающие тем свойством, что

$$\sup_k r(Y_{\sigma_k},\ Y'_{\sigma_k}) \to 0 \text{ при } \mu(\mathbf{X},\ \mathbf{X}') \to 0. \tag{3.13}$$

Рис. 5. поясняет описываемую идею.

Отметим разницу между соотношениями (3.9) и (3.13). Если в (3.9) требуется, чтобы равномерно *по всем моментам* времени величины $r(Y_n,\ Y'_n)$ стремились к 0 при $\mu(\mathbf{X},\ \mathbf{X}') \to 0$, то в (3.13) такая равномерность требуется лишь *по выбранным случайным моментам* времени σ_k.

Определим, например, для процессов $\mathbf{Y}$ и $\mathbf{Y}'$ вида (3.10) моменты σ_k такие, когда одновременно $Y_{\sigma_k} = Y'_{\sigma_k} = 0$. Тогда $r(Y_{\sigma_k},\ Y'_{\sigma_k}) = 0$ и, следовательно, условие (3.13) заведомо выполнено.

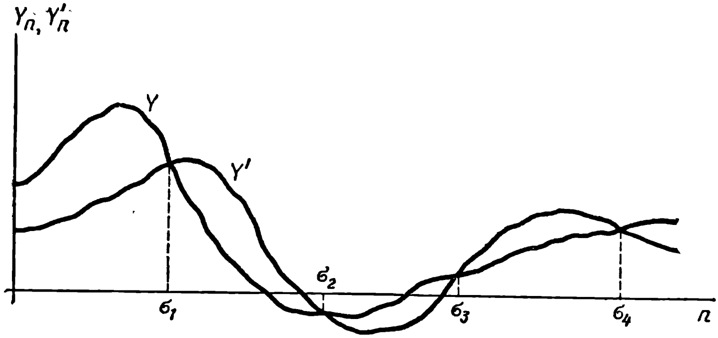


Рис. 5.

Предположим теперь, что случайные моменты σ_k расположены «достаточно регулярно» на оси времени. Потребуем, например, чтобы при всех $\mathbf{X} \in \mathscr{X}^*$, $\mathbf{X}' \in \mathscr{X}^*$ и некоторых $\gamma > 1$ и $g < \infty$

$$\sup_k E(\sigma_{k+1} - \sigma_k)^\gamma \leqslant g. \tag{3.14}$$

Тогда оценка расстояния $\nu(\mathbf{Y}, \mathbf{Y}')$ вида (3.8) зависит от того, успеют ли сравниваемые процессы $\mathbf{Y}$ и $\mathbf{Y}'$ «далеко разойтись» за промежуток времени $[\sigma_k, \sigma_{k+1})$, поскольку близость величин Y_{σ_k} и Y'_{σ_k} гарантируется построением моментов времени σ_k (см. (3.13)).

Если, например, при всех $n \geqslant 0$ (*условие непрерывности на произвольных конечных отрезках времени*)

$$\sup_k \nu(Y_{\sigma_k+n}, Y'_{\sigma_k+n}) \to 0 \quad \text{при} \quad \mu(\mathbf{X}, \mathbf{X}') \to 0, \tag{3.15}$$

то из (3.15) (обобщающего (3.13) в совокупности с (3.14) следует выполнение соотношения (3.9), т. е. равномерная по времени непрерывность. Более того, имея оценки вида (3.14) и количественные оценки соотношения (3.15), можно получить оценки непрерывности.

Возвращаясь к модели (3.10), мы видим, что условие (3.14) является условием конечности момента порядка γ для времени возвращения траектории процесса $\{Y_n, Y'_n\}_{n\geqslant 0}$ в точку (0, 0), иначе говоря, условия выполнения соотношения (3.14) могут быть найдены методами, рассмотренными выше.

Можно показать, что неравенство (3.14) выполнено, если подмножество $\mathscr{X}^* \in \mathscr{X}$ состоит из последовательностей, элементы которых удовлетворяют соотношениям:

$$EX_n \leqslant -a, \;\; E(X_n^+)^\gamma \leqslant b \tag{3.16}$$

при некоторых фиксированных положительных a и b, где $X_n^+ = \max(0, X_n)$.

В данной модели обслуживания

$$r(Y_{\sigma_k+n}, Y'_{\sigma_k+n}) \leq n\mu(\mathbf{X}, \mathbf{X}'). \qquad (3.17)$$

Из неравенства (3.17) вытекает выполнение условия (3.13). В результате в множестве $\mathscr{X}^*$, задаваемом неравенствами (3.16), можно получить оценку:

$$\sup_n L(G_n, G'_n) \leq c\,[L(H, H')]^{(\nu-1)/\nu}.$$

Таким образом, применение изложенного подхода предполагает:

1) выявление «точек близости» $\{\sigma_k\}$ для рассматриваемых моделей;

2) разработку методов проверки условий типа (3.14) (возвратность процессов $(\mathbf{Y}, \mathbf{Y}')$ специального вида);

3) разработку методов оценки непрерывности на произвольных конечных промежутках времени, см. (3.15).

Методы, применяемые на первых двух этапах, относительно слабо связаны с конкретным видом исследуемых моделей — они в основном определяются классом анализируемых процессов. В результате выполнения данных этапов определяется подмножество $\mathscr{X}^*$, на котором возможна непрерывность.

В то же время оценки, получаемые на третьем этапе, в значительной степени определяются структурой и особенностями исследуемой модели. Именно эти оценки фактически определяют вид метрик, в которых имеет место непрерывность. Исследования непрерывности как конкретных стохастических моделей, так и их классов можно найти в работах [6, 17, 19].

Оценка скорости сходимости. Мы условились, что динамику изучаемой модели описывает случайный процесс $\{Y_t\}$. Однако исследователя, как правило, интересуют не сами реализации этого процесса, а некоторые его статистические характеристики. Одной из таких важных характеристик является вероятностное распределение $Q_t(S)$ состояния Y_t в момент времени t, т. е. $Q_t(S) = P\{Y_t \in (S)\}$.

Зная распределения Q_t, можно найти многие практически интересные показатели работы модели.

Однако нахождение «нестационарных» распределений Q_t представляет серьезные, а иногда и непреодолимые технические и математические трудности. Оценка же распределений на машинных моделях также наталкивается на отсутствие эффективных статистических процедур оценки нестационарных процессов. Поэтому на практике часто ограничиваются оценкой финальных распределений $Q = \lim_{t \to \infty} Q_t$. Данные распределения могут находиться как из решения соответствующих уравнений (например, для марковских процессов с конечным множеством состояний для нахождения распределений Q_t необходимо решать систему линейных дифференциальных уравнений Колмогорова, а для нахождения Q — систему линейных алгебраических уравнений), так и хорошо разработанными статистическими методами, например методом Монте-Карло. Однако не всегда учитывается ограниченное значение найденных таким образом распределений. Дело в том, что «стационарный режим» устанавливается при весьма общих предположениях (при постоянстве параметров, определяющих работу модели). Однако модель является лишь некоторым отражением реальной системы. В последней же параметры с течением времени могут меняться; могут меняться и условия, в которых система работает. Поэтому в реальной системе стационарный режим, строго говоря, отсутствует. Следовательно, возникает вопрос о том, насколько найденные стационарные характеристики соответствуют реальным.

Такое соответствие имеет место, если скорость сходимости к стационарному режиму в модели «много больше» скорости изменения параметров моделируемого объекта. При этом в реальной системе возникает режим, который можно назвать «квазистационарным» и вероятностные характеристики которого в известном смысле близки к получаемым в результате моделирования стационарным характеристикам.

Если же скорость сходимости к стационарному режиму в модели соизмерима со скоростью изменения параметров модели, то стационарные характеристики, получаемые в результате моделирования, могут быть весьма далеки от реальных.

Таким образом, оценка скорости сходимости к стационарному режиму позволяет судить о возможности использования стационарных характеристик (нахождение их обычно не представляет серьезных технических и математических трудностей) для получения показателей работы реальной системы.

Необходимо также отметить и те случаи, когда оценки скорости сходимости представляют собой основные показатели работы модели. В качестве примеров можно привести оценки времени «рассасывания пиковых ситуаций» в моделях дорожного трафика, системах передачи и обработки данных, вычислительных системах, оценки времени ликвидации нежелательных (аварийных) ситуаций и их последствий в организационных системах.

Оказывается, что во многих случаях оценку скорости сходимости можно свести к оценке распределения времени первого достижения или его моментных характеристик. Проиллюстрируем основную идею получения таких оценок на примере регенерирующих процессов *.

Итак, пусть $\{Y_t\}$ — регенерирующий процесс и θ_1, θ_2, ... — последовательные периоды регенерации. Построим соответствующую этому процессу *стационарную версию* $\{Y_t^0\}$, т. е. такой регенерирующий процесс, у которого распределения Q_t^0 от времени t не зависят и совпадают с $Q = \lim\limits_{t\to\infty} Q_t$. Важно подчеркнуть, что последовательные циклы процесса $\{Y_t^0\}$ (кроме, быть может, первого) не только независимы между собой и имеют одинаковые распределения, но и распределены одинаково с циклами процесса $\{Y_t\}$. То, что первый цикл в стационарной версии отличается от остальных, является следствием стационарности.

Поскольку до сих пор мы интересовались лишь частными характеристиками процессов $\{Y_t\}$ и $\{Y_t^0\}$, то, вообще говоря, не имело смысла говорить, например, о вероятности $P\,(Y_t = Y_t^0)$, а такого рода вероятности нам понадобятся в дальнейшем. Поэтому прежде всего позаботимся о совместном задании исследуемых процессов. Построим процесс $Z_t = \{Y_t, Y_t^0\}$, компоненты которого являются вероятностными копиями процессов $\{Y_t\}$ и $\{Y_t^0\}$ соответственно. Чтобы не загромождать изложение, мы даже обозначим их одинаково, хотя, строго говоря, следовало бы применить различные обозначения. Вообще говоря, пар $\{Y_t, Y_t^0\}$ с указанным свойством может быть много. Пользуясь этим, попытаемся найти такую пару, для которой существует момент времени ξ, являющийся *общим моментом регенерации* сразу для обоих процессов $\{Y_t\}$, $\{Y_t^0\}$. Тогда,

* Данная идея в той или иной форме использовалась в работах [6, 19, 35, 43] и др.

грубо говоря, начиная с момента ξ процессы $\{Y_t\}$ и $\{Y_t^0\}$ совпадают в вероятностном смысле (а на самом деле можно добиться даже совпадения траекторий), в частности, совпадают и их распределения Q_t и Q_t^0. Таким образом, можно получить оценку

$$|Q_t(B) - Q_t^0(B)| \leqslant P(\xi > t), \tag{3.18}$$

которая фактически и представляет собой оценку скорости сходимости.

Для ее получения необходимо уметь строить такую пару (Y_t, Y_t^0), для которой правая часть (3. 18) была бы поменьше и во всяком случае стремилась бы к 0 при $t \to \infty$. Данная задача может быть нетривиальной. Например, если длительности циклов θ_i имеют плотность, то, скажем, для независимых процессов $\{Y_t\}$ и $\{Y_t^0\}$ момент ξ не наступит никогда с вероятностью 1. Поэтому добиться существования конечного времени ξ можно, лишь вводя достаточно сильную зависимость между $\{Y_t\}$ и $\{Y_t^0\}$.

Оценка же вероятности $P(\xi > t)$ может осуществляться методами, рассмотренными выше для оценок времени первого достижения или их моментов. Например, если существует $E\,\xi^\gamma$, $\gamma > 0$, то в соотношении (3.18) можно воспользоваться неравенством Чебышева:

$$P(\xi > t) \leqslant E\xi^\gamma / t^\gamma. \tag{3.19}$$

Оказывается, что если у регенерирующего процесса длительности циклов обладают конечным моментом порядка $1 + \gamma$, $\gamma > 0$, т. е. $E\theta_i^{1+\gamma} < \infty$, то (при несущественных с практической точки зрения дополнительных предположениях, налагаемых на распределения величины θ_i — типа неарифметичности и т. п.) можно построить такую пару, $Z_t = (Y_t, Y_t^0)$, для которой $E\,\xi^\gamma < \infty$ и, следовательно справедливы оценки (3.18) и (3.19).

З а д а ч и р а с ч е т а р е ж и м о в р а б о т ы э л е м е н т о в с л о ж н ы х с и с т е м. Рассмотренные выше свойства касались в основном качественной картины поведения изучаемых моделей, хотя много внимания уделялось и вопросам получения соответствующих количественных оценок. Как уже отмечалось, качественные свойства используются прежде всего на этапе решения принципиальных вопросов о работоспособности системы. Далее, однако, необходимо рассчитать ее конкретные режимы работы. При этом возникают серьезные математические трудности: даже если и удается найти, например, уравнения, описывающие динамику или стационарный режим таких систем, то решить их можно в крайне редких случаях (либо ввиду их сложного вида, либо ввиду их большой размерности). Решение указанных уравнений дает информацию значительно более подробную, чем это необходимо для исследователя. Это фактически приводит к невозможности ее использования без дополнительной обработки, например получения числовых характеристик для найденных законов распределения. Следует также иметь в виду, что зачастую системы работают в таких режимах, которые приводят к дополнительным трудностям — плохо обусловленным уравнениям, наличию малых параметров и т. д. Типичным в этом отношении является пример *высоконадежных систем*, т. е. систем, элементы которых могут выходить из строя и восстанавливаться, а соотношение между параметрами таково, что выход из строя всей системы происходит крайне редко. Поэтому, скажем, вероятность нахождения системы в отказовом состоянии близка к 0. Задача нахождения этой вероятности может приводить к большим погрешностям. Рассмотрим элементарный пример.

Пусть система описывается цепью Маркова с тремя состояниями — рабочим (0), предотказовым (1) и отказовым (2). И пусть вероятности перехода равны $p_{00} = 1 - \varepsilon_1$, $p_{01} = \varepsilon_1$, $p_{10} = 1 - \varepsilon_2$, $p_{12} = \varepsilon_2$, $p_{21} = 1$ (остальные вероятности равны 0). Стационарные вероятности p_j ($j = 0,1,2$) находятся из следующей системы уравнений:

$$p_j = \sum_{i=0}^{2} p_i p_{ij} \, (j = 0, 1, 2), \sum_{j=0}^{2} p_j = 1. \qquad (3.20)$$

Поскольку уравнения (3.20) линейно-зависимы, то выберем из них три линейно-независимых. Например,

$$\begin{aligned} p_0 &= p_0 (1 - \varepsilon_1) + p_1 (1 - \varepsilon_2), \\ p_1 &= p_0 \varepsilon_1 + p_2, \\ p_0 &+ p_1 + p_2 = 1. \end{aligned} \qquad (3.21)$$

Решение данной системы имеет вид:

$$p_0 = \frac{1 - \varepsilon_2}{1 - \varepsilon_2 + \varepsilon_1 + \varepsilon_1 \varepsilon_2}; \quad p_1 = \frac{\varepsilon_1}{1 - \varepsilon_2 + \varepsilon_1 + \varepsilon_1 \varepsilon_2};$$

$$p_2 = \frac{\varepsilon_1 \varepsilon_2}{1 - \varepsilon_2 + \varepsilon_1 + \varepsilon_1 \varepsilon_2}.$$

Представим теперь, что вероятности перехода получаются в результате обработки, например, экспериментальных данных и, следовательно, заданы с некоторой погрешностью.

Пусть, например, в (3.21) вероятности $1-\varepsilon_1$, $1-\varepsilon_2$ и ε_1 заданы с одной и той же относительной погрешностью δ, т. е. истинные вероятности равны $(1-\varepsilon_1)+\theta_1\delta(1-\varepsilon_1)$, $(1-\varepsilon_2)+\theta_2\delta(1-\varepsilon_2)$, $\varepsilon_1+\theta_3\delta\varepsilon_1$, где $-1\leqslant\theta_i\leqslant 1$, $i=1, 2, 3$.

Найдем из уравнений (3.21), в котором идеальные вероятности заменены на истинные, вероятность p_2, например, по правилу Крамера.

Положим для определенности $\theta_2=\theta_3=0$. Тогда

$$p_2=\frac{\varepsilon_1\varepsilon_2-\theta_1\delta(1-\varepsilon_1)}{1-\varepsilon_2+\varepsilon_1+\varepsilon_1\varepsilon_2-\theta_1\delta(1-\varepsilon_1)} \tag{3.22}$$

и, например, при $\varepsilon_1=\varepsilon_2=\delta=0{,}01$ эта вероятность (в зависимости от θ_1) лежит в диапазоне $0\div 0{,}01$ (формально получаемые отрицательные значения мы не учитываем, числовые данные округлены). В то же время ранее найденная формула дает значение $p_2=0{,}0001$.

Таким образом, даже этот элементарный пример показывает необходимость разработки специальных вычислительных методов для расчета режимов работы сложных систем. В рассмотренном примере (а также во многих других аналогичных ситуациях [28]) целесообразно изучать цепь Маркова не алгебраическими методами, а трактуя ее как регенерирующий процесс и выбирая за точки регенерации моменты пребывания цепи в «высоковероятном» состоянии 0. Тогда за период регенерации возможны следующие переходы: $0\to 0$; $0\to 1\to 0$; $0\to 1\to 2\to 1\to 0$; $0\to 1\to 2\to 1\to 2\to 1\to 0$ и т. д. Однако «в расчет» при малых ε_1 и ε_2 следует принимать лишь первые 3 из перечисленных траекторий — другие значительно менее вероятны. Например, вероятности перечисленных траекторий равны при $\varepsilon_1=\varepsilon_2=\varepsilon$ соответственно:

$$1-\varepsilon,\ \ \varepsilon(1-\varepsilon),\ \ \varepsilon^2(1-\varepsilon),\ \ \varepsilon^3(1-\varepsilon).$$

При этом вероятность p_2 будет равна попросту доле «наиболее вероятных» траекторий, проходящих через точку 2, т. е. $\varepsilon^2+o(\varepsilon)$. Методика определения вероятности какого-либо события в этом случае заключается в подсчете доли «наиболее вероятных» траекторий, приводящих к данному событию на одном периоде регенерации. При этом способе вычислений возможные малые погрешности определения вероятностей перехода не приведут к большим погрешностям в определении искомых вероятностей.

Представим себе, что в рассмотренном примере нас интересует не стационарная вероятность p_2, а случайное время τ_2 первого достижения состояния 2, вернее, функция распределения этого времени или ее числовые характеристики. Что можно сказать относительно τ_2? Во-первых, очевидно, что $\tau_2 \to \infty$, т. е. $P(\tau_2 \leqslant x) \to 0$ при любом x, если $\varepsilon = \min(\varepsilon_1, \varepsilon_2) \to 0$. Во-вторых, нахождение точного вида распределения величины τ_2 даже в рассматриваемом простом случае приводит к громоздким выкладкам и результатам. Поэтому возникает вопрос, нельзя ли подобрать такой *нормирующий множитель* m ($m \to 0$ при $\varepsilon \to 0$), чтобы распределение величины $m\tau_2$ имело некоторый предел при $\varepsilon \to 0$? Если, например, $\lim\limits_{\varepsilon \to 0} P(m\tau_2 \leqslant x) = G(x)$, то вместо нахождения упомянутых громоздких формул можно при конечных ε пользоваться приближенным равенством:

$$P(\tau_2 \leqslant x) \approx G(mx). \tag{3.23}$$

Рассмотрим, как можно решить данную задачу для регенерирующих процессов [36]. Предположим для определенности, что изучается регенерирующий процесс $\{Y_t\}$ и случайный момент времени τ первого достижения подмножества $A \subset V$. Если A трактуется как множество отказовых состояний, то τ — время безотказной работы. Предположим, что динамика процесса $\{Y_t\}$ определяется некоторым параметром $\varepsilon > 0$ и $\tau \to \infty$ при $\varepsilon \to 0$. Здесь ε выступает просто как индекс соответствующего процесса, а не как реальный малый параметр, т. е. при изменении ε, вообще говоря, могут меняться *все* характеристики процесса $\{Y_t\}$. Эти характеристики будем далее отмечать индексом ε. Обозначим через q вероятность того, что на цикле происходит отказ системы, т. е. $q = P(\tau < \theta_1.)$ В силу того что $\tau \to \infty$ при $\varepsilon \to 0$, имеем $q(\varepsilon) \to 0$ при $\varepsilon \to 0$. Пусть $F(x)$ — функция распределения интервалов регенерации: $F(x) = P(\theta_1 \leqslant x)$. Временно предположим, что функции $F(x)$ от ε не зависят. Пусть ν — номер цикла, на котором произошел отказ системы. Тогда имеем:

$$\theta_1 + \ldots + \theta_{\nu-1} \leqslant \tau \leqslant \theta_1 + \ldots + \theta_\nu. \tag{3.24}$$

Очевидно, величина ν является случайной с геометрическим распределением:

$$P(\nu = k) = (1 - q)^{k-1} q, \; k = 1, 2, \ldots$$

(с вероятностью $(1 - q)^{k-1}$ выход процесса в подмножество A не происходит на первых $k - 1$ циклах и с вероятностью q происходит на k-м цикле; произведение взято в силу независимости и одинаковой распределенности циклов). Ситуация здесь такая, что очередной цикл «бракуется» с вероятностью q и считается «незабракованным» с вероятностью $1 - q$. Если бы «отбраковка» производилась независимо от длительности цикла, то из неравенства (3.24) следовало бы, что

$$h_1 (1/q - 1) = E\theta_1 (E\nu - 1) \leqslant E\tau \leqslant E\theta_1 E\nu = h_1/q, \quad (3.25)$$

где $h_1 = E\theta_1$.

Кроме того, так как

$$\lim_{q \to 0} P\left\{\frac{\theta_1 + \ldots + \theta_\nu}{E\theta_1 E_\nu} \leqslant x\right\} = 1 - e^{-x} \quad (3.26)$$

(это утверждение — фактически следствие известной предельной теоремы Пуассона [37]), то из (3.24) — (3.26) получаем, что

$$\lim_{\varepsilon \to 0} P\left\{q(\varepsilon)\frac{\tau(\varepsilon)}{h_1} \leqslant x\right\} = 1 - e^{-x}. \quad (3.27)$$

Однако в действительности «отбраковка» может зависеть от длительности цикла. В приведенном примере с цепью Маркова попадание в состояние 2 могло происходить лишь на циклах, длительность которых $\geqslant 4$. Поэтому вывод соотношения (3.27) в общем случае некорректен. Тем не менее оказывается (см. работу А. Д. Соловьева [36]), что предельное соотношение (3.27) сохраняется, если потребовать, чтобы при $\varepsilon \to 0$ выполнялось $qE\theta_1^2/h_1^2 \to 0$ (вместо ранее предложенного $q \to 0$). Если (в соответствии со сделанным выше предположением) распределение $F(x)$ не зависит от ε, то новое условие требует лишь существования второго момента у величины θ_1. Однако доказанный А. Д. Соловьевым результат гораздо шире — он говорит о том, что если при $\varepsilon \to 0$

$$q(\varepsilon) E\theta_1^2(\varepsilon)/h_1^2(\varepsilon) \to 0 \quad (3.28)$$

(т. е. рассматриваемые процессы $\{Y_t\}$ с изменением индекса ε могут меняться как угодно, лишь бы они оставались регенерирующими и выполнялось (3.28); при этом допускается и изменение множества A), то выполнено и (3.27).

Возвратимся к рассмотренному выше примеру цепи Маркова и обозначим, как и ранее, $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$, τ_2 — время первого достижения состояния 2. Пусть моменты регенерации совпадают с моментами пребывания цепи в состоянии 0. Легко подсчитать, что $h_1 = E\theta_1 = 1 + O(\varepsilon)$, $E\theta_1^2 = 1 + O(\varepsilon)$*, $q(\varepsilon) = \varepsilon_1\varepsilon_2$. Следовательно, из приведенных результатов получается, что $\lim\limits_{\varepsilon \to 0} P\{\varepsilon_1\varepsilon_2\tau_2 \leqslant x\} =$

---

* Символ $O(\varepsilon)$ имеет стандартный смысл $|O(\varepsilon)/\varepsilon| \leqslant C$, где C — некоторая постоянная.

$= 1 - e^{-x}$, т. е. в этом случае в формуле (3.23) можно положить $m = \varepsilon_1\varepsilon_2$, $G(x) = 1 - e^{-x}$.

Однако при практическом использовании приближенных равенств типа (3.23) возникают следующие проблемы. Во-первых, нахождение оценки точности этих равенств при конечных значениях ε. Во-вторых, определение конкретных оценок параметров (q, h_1, $E\theta_1^2$), входящих в условие теоремы.

Начнем со второй проблемы. Решение ее в значительной степени должно учитывать характерные особенности исследуемой системы. Так, одним из эффективных методов оценки вероятности q является ее приближение вероятностью q' отказа лишь по специальным траекториям [28, 36], обычно имеющим монотонный характер,— в приведенном выше примере цепи Маркова такой траекторией была $0 \to 1 \to 2$, вероятность которой равна $\varepsilon_1\varepsilon_2$ и в данном случае точно совпадает с q («немонотонной» траекторией могла бы быть $0 \to 1 \to 1 \to 2$, но она в примере имеет вероятность 0). Для подобных оценок может эффективно использоваться метод Монте-Карло. Например, в работе И. Н. Коваленко [28] разработана методика таких оценок для линейчатых процессов. Не исключены и другие методы оценки q. Что касается оценок величины θ_1, то о них мы говорили выше.

Решению проблемы точности приближенных равенств помогает «равномерная» формулировка сформулированного асимптотического результата. Условия вида (3.28) «показывают нам, какие аномалии в распределениях могут нарушать сходимость, т. е. показывают, когда практически нельзя пользоваться соответствующими приближенными равенствами» [36]. Наряду с этим чисто методологическим результатом в ряде случаев [19, 36] для распределения $P(\tau \leqslant x)$ можно найти двусторонние неравенства, которые имеют то преимущество перед соотношением вида (3.27), что они явно указывают границы возможных значений оцениваемой функции распределения, т. е. как раз и решают задачу оценки точности асимптотических формул.

Рассмотренный метод получения асимптотических распределений применим в случае, когда в системе удается выделить *редкие события*, представляющие основной интерес. В этом плане утверждения, лежащие в основе данного метода, в известном смысле родственны утверждениям предельной теоремы Пуассона из теории вероятностей.

Существует еще одна большая группа результатов, родственная, если угодно, центральной предельной теореме. Речь идет о сходимости случайных процессов (описывающих исследуемые модели и имеющих сложную структуру) к «простым» процессам. В качестве таких процессов обычно выступают *диффузионные* [37], а еще чаще — простейший из диффузионных процессов — винеровский. Диффузионные процессы — это однородные марковские процессы с непрерывными траекториями (ради простоты предположим, что рассматриваются процессы на прямой), которые, грубо говоря, обладают следующими свойствами: (1) среднее приращение процесса за малое время Δt с точностью до $o(\Delta t)$ пропорционально величине Δt, причем коэффициент пропорциональности (*коэффициент сноса*) зависит, вообще говоря, от точки, в которой траектория процесса находилась в начале этого интервала; (2) дисперсия этого приращения также с точностью до $o(\Delta t)$ пропорциональна Δt (коэффициент пропорциональности называется *коэффициентом диффузии*). Отсюда следует *условие непрерывности*:

$$P(|Y_{t+\Delta t}-Y_t|>\varepsilon)\leqslant\frac{\sigma^2(Y_{t+\Delta t}-Y_t)}{\varepsilon^2}\to 0 \text{ при } \Delta t\to 0$$

и любом $\varepsilon>0$.

Можно ослабить приведенные условия, потребовав существования «усеченных» среднего и дисперсии и дополнительно выполнения условия непрерывности.

Поскольку свойства диффузионных процессов достаточно хорошо изучены, то аппроксимация ими исходных процессов может позволить эффективно изучать последние. Таким образом, основной проблемой является выяснение условий, когда такая аппроксимация возможна.

Рассмотрим сначала простейший пример случайного блуждания (к изучению которого, впрочем, можно свести процесс (3.10)):

$$Y_{n+1}=Y_n+X_n,\quad n\geqslant 0, \tag{3.28}$$

где $\{X_n\}$ — последовательность независимых одинаково распределенных случайных величин. Обозначим $EX_n = \delta$, $\sigma^2X_n=a>0$. Предположим, что величина δ мала (т. е. формально будет предполагаться, что $\delta\to 0$). Рассмотрим вместо процесса $\{Y_n\}$ процесс $\{Y'_n\}$, где $Y'_n=\delta Y_n$, т. е. попросту изменим в δ раз масштаб в пространстве состояний исходного процесса. Тогда из (3.28) следует, что

$$Y'_{n+1}=Y'_n+X'_n,$$

где $X'_n = \delta X_n$. Следовательно, $EX'_n = \delta^2$, $\sigma^2 X'_n = a\delta^2$. Теперь изменим еще и масштаб времени. Введем процесс, изменяющийся в моменты, кратные δ^2:

$$Y''_{n\delta^2} = Y'_n.$$

Тогда

$$Y''_{(n+1)\delta^2} = Y''_{n\delta^2} + X'_n. \qquad (3.29).$$

Таким образом, процесс $\{Y''_t\}$ получается из $\{Y_t\}$ в результате как изменения масштаба времени, так и преобразования пространства состояний.

Рассмотрим теперь произвольный промежуток времени Δt, и пусть

$$\Delta Y'' = Y''_{t+\Delta t} - Y''_t.$$

Из (3.29) следует, что число скачков процесса $\{Y''_t\}$ за время Δt не меньше $N = [\Delta t/\delta^2]$ и не больше $N + 1$. Здесь [ ] — целая часть числа $\Delta t/\delta^2$. Следовательно,

$$NEX'_n \leqslant E\Delta Y'' \leqslant (N+1)EX'_n,$$

т. е.

$$\delta^2 [\Delta t/\delta^2] \leqslant E\Delta Y'' \leqslant \delta^2 ([\Delta t/\delta^2] + 1).$$

Таким образом,

$$\lim_{\delta\to 0} E\Delta Y'' = \Delta t,$$

т. е. среднее приращение процесса $\{Y''_t\}$ за время Δt пропорционально (даже равно) Δt при $\delta \to 0$. Аналогично дисперсия величины $\Delta Y''$, равная сумме дисперсий суммируемых величин X'_n, оценивается так:

$$N\sigma^2 X'_n \leqslant \sigma^2 (\Delta Y'') \leqslant (N+1)\sigma^2 X'_n,$$

т. е.

$$\lim_{\delta\to 0} \sigma^2 (\Delta Y'') = a\Delta t.$$

Следовательно, в пределе при $\delta \to 0$ процесс $\{Y''_t\}$ вырождается в диффузионный с коэффициентом сноса 1 и коэффициентом диффузии a.

В общем случае доказательство сходимости получается не таким прямым выводом, а путем достаточно сложных косвенных оценок. Тем не менее и в общем случае аппроксимация диффузионными процессами получается путем двойного предельного перехода — «по пространству» и «по времени». Весьма общие результаты на этом пути получены А. А. Боровковым [6], рассмотревшим, в частности, и линейчатые процессы, возникающие в теории массового обслуживания.

Анализ сложных систем с помощью моделирования. Выше был рассмотрен ряд задач, возникающих при анализе сложных систем. Если говорить о практическом применении приведенных результатов, то следует признать, что они (в рассмотренной нами форме) пригодны в лучшем случае лишь для исследования отдельных подсистем. Исключение составляют структурные методы, имеющие дело не с подсистемами, а со структурой системы в целом.

Вместе с тем, как отмечалось в разделе 1, основную особенность функционирования сложных систем составляет взаимодействие составляющих их подсистем между собой и с внешней средой. В настоящее время комплексное исследование сложных систем, учитывающее такое взаимодействие, возможно лишь с помощью ЭВМ, причем основной способ использования ЭВМ заключается не в проведении численных расчетов по готовым формулам (таких формул либо не существует, либо они чрезвычайно громоздки даже для ЭВМ), а в *моделировании* функционирования сложной системы. Проблемы, возникающие при этом, и методы их решения заслуживают отдельной публикации. Поэтому здесь они будут рассмотрены лишь кратко. Основное внимание будет уделяться именно проблеме *исследования* сложных систем. Возникающие многочисленные и существенные проблемы программирования (вопросы реализации моделей на ЭВМ, общения с ними и т. п.) затрагиваться практически не будут — см. по этому поводу обзор [16].

Что же такое моделирование и какие задачи могут решаться при его проведении?

Уже говорилось, что при исследовании сложных систем роль натурного эксперимента крайне ограничена. И вот своеобразной заменой натурного эксперимента стало моделирование на ЭВМ. Традиционно под моделированием понимался процесс составления машинной программы, воспроизводящей (имитирующей) динамику системы и получение с помощью этой программы данных, для сбора которых в ином случае пришлось бы проводить натурный эксперимент. Данная машинная программа отображает системные события в их причинно-следственных отношениях. Построение подобных программ и их отладка — дело чрезвычайно трудоемкое (ведь системы-то сложные!), поэтому в настоящее время ведутся интенсивные работы по их кардинальному облегчению.

В процессе все более широкого применения моделирования выявились многочисленные проблемы его использования, которые вызваны не способом записи модели, а ее содержанием. Именно поскольку всякая модель неадекватна реальной системе, то кардинальными являются вопросы о *степени адекватности, точности модели*, ее *чувствительности* к вариациям различных параметров и т. д. Ответы на эти вопросы в конечном счете определяют пригодность модели и возможность ее использования в качестве подопытного объекта для вынесения суждений и принятия решений о реальной системе. Далее, если модель уже создана, то возникают проблемы использования ее «внутренних» свойств, таких, как *устойчивость*, *надежность*, *управляемость* и т. д. То, что данные проблемы, вообще говоря, не решаются с помощью прямой имитации, наглядно показывает пример высоконадежных систем: применение для оценки надежности имитации ведет к нереализуемым затратам машинного времени. Поэтому решение указанных проблем предполагает проведение *целенаправленных имитационных экспериментов*, использующих не только возможности ЭВМ по имитации, но и аналитические результаты исследования соответствующих математических моделей [8, 16, 28].

Таким образом, если требовать, чтобы моделирование было *научным инструментом* исследования, то в понятие моделирования следует включать не только имитацию на ЭВМ, но также последовательное изменение и анализ самой модели, проводимые как на начальных, так и на промежуточных стадиях эксперимента, и осуществление целенаправленных имитационных экспериментов, что предоставляет пользователю исчерпывающую информацию о моделируемой системе и помогает принимать решения об ее изменениях.

Проведение моделирования во всей полноте осуществляется с помощью *имитационной системы*. Имитационная система — это программный комплекс, состоящий из двух частей [16, 20, 21]:

1) обеспечивающей проведение машинного эксперимента (программная реализация, ввод-вывод, процесс счета, диалог и т. п.);

2) обеспечивающей формирование модели, постановку машинного эксперимента, анализ свойств модели и принятие решения.

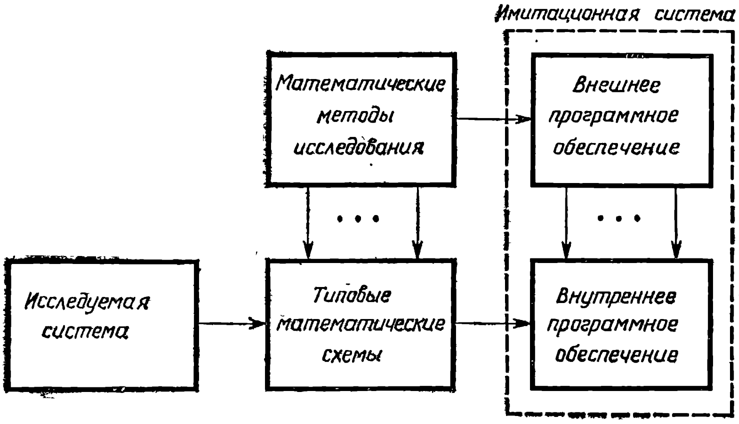


Рис. 6.

Первую часть программного комплекса называют *внутренним программным обеспечением*, а вторую — *внешним*.

Рис. 6 показывает общую схему исследования с помощью имитационной системы. Имитационная система выполняет двоякую роль. Во-первых, ее внутреннее программное обеспечение дает возможность «воспроизводить» на ЭВМ ту или иную математическую схему, описывающую изучаемую систему. Во-вторых, внешнее программное обеспечение имитационной системы, «программно отображая» математические методы исследования формальных схем, дает возможность проводить целенаправленные имитационные эксперименты и анализ систем.

Описанная методика исследования с помощью имитационной системы лежит в основе агрегативного подхода Н. П. Бусленко, суть которого заключается в построении имитационных систем на основе формальных математических конструкций. Изложение этого подхода не входит в задачу данной брошюры, и мы отсылаем читателя к имеющейся литературе [8—10, 20, 21].

* *
*

Итак, мы видим, что термин «сложные системы» не только прижился, но, что более важно, в настоящее время базируется на достаточно надежном фундаменте, составляющими которого являются реально существующие или проектируемые народнохозяйственные и технические объекты, а также современные математические модели и методы их исследования. Разнообразие математических моделей приводит к использованию широкого спектра методов их исследования, в который включаются как аналитические построения, так и машинное моделирование. Фундамент этот постоянно укрепляется: появляются новые модели, новые математические результаты, совершенствуются методы моделирования. Это дает возможность последовательно расширять класс изучаемых реальных объектов. Вместе с тем, естественно, расширение класса решенных задач в еще большей степени расширяет круг задач нерешенных, появляющихся в результате проводимых исследований. Задача данной брошюры будет полностью выполнена, если ее прочтение породит у читателя интерес к исследованию сложных систем, постановкам и решениям таких нерешенных задач.

ЛИТЕРАТУРА

1. Айзерман М. А. Теория автоматического управления. М., Наука, 1966.

2. Айзерман М. А., Гусев А. А., Розоноэр Л. И., Смирнова И. М., Таль А. А. Логика, автоматы, алгоритмы. М., Физматгиз, 1963.

3. Арнольд В. И. Обыкновенные дифференциальные уравнения. М., Наука, 1975.

4. Беллман Р., Кук К. Дифференциально-разностные уравнения. М., Мир, 1967.

5. Беляев Ю. К. Линейчатые марковские процессы и их приложение к задачам теории надежности. Труды VI Всес. совещания по теории вероят. и мат. статистике, 1960. Вильнюс, Гос. изд-во полит. и научн. лит-ры Лит. ССР, 1962.

6. Боровков А. А. Асимптотические методы в теории массового обслуживания. М., Наука, 1980.

7. Бусленко Н. П. Моделирование производственных процессов на ЦВМ. М., Наука, 1964.

8. Бусленко Н. П. Сложные системы и имитационные модели. — Кибернетика, 1976, № 6.

9. Бусленко Н. П. Моделирование сложных систем. М., Наука, 1978.

10. Бусленко Н. П., Калашников В. В., Коваленко И. Н. Лекции по теории сложных систем. М., Сов. радио, 1973.

11. Вентцель Е. С. Введение в исследование операций. М., Сов. радио, 1964.

12. Гельфонд А. О. Исчисление конечных разностей. М., Наука, 1967.

13. Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д. Математические методы в теории надежности. М., Наука, 1965.

14. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. М., Наука, 1966.
15. Дынкин Е. Б. Марковские процессы. М., Физматгиз, 1963.
16. Емельянов С. В., Калашников В. В., Лутков В. И., Немчинов Б. В. Методологические вопросы построения имитационных систем. М., Изд-во МЦНТИ, 1978.
17. Золотарев В. М. О непрерывности стохастических последовательностей, порождаемых рекуррентными процедурами. — Теория вероятностей и ее применения, 1975, т. 20, № 4.
18. Золотарев В. М. Метрические расстояния в пространствах случайных величин и их распределений. — Мат. сборник, 1976, т. 101 (143), № 3.
19. Калашников В. В. Качественный анализ поведения сложных систем методом простых функций. М., Наука, 1978.
20. Калашников В. В., Лутков В. И., Немчинов Б. В., Ривес Н. Я. Вопросы разработки имитационных систем. — Электронная техника, 1980, сер. 9, вып. 1 (34).
21. Калашников В. В., Лутков В. И., Немчинов Б. В., Ривес Н. Я. Внешнее программное обеспечение агрегативной имитационной системы. — Электронная техника, 1980, сер. 9, вып. 2 (35).
22. Калашников В. В., Соловьев М. М., Средникова Л. П. Два примера использования агрегативной имитационной системы. — Электронная техника, 1980, сер. 9, вып. 2 (35).
23. Калман Р., Фалб П., Арбиб М. Очерки по математической теории систем. М., Мир, 1971.
24. Карлин С. Основы теории случайных процессов. М., Мир, 1971.
25. Клейнрок Л. Теория массового обслуживания. М., Машиностроение, 1979.
26. Клейнрок Л. Коммутационные сети. М., Наука, 1970.
27. Коваленко И. Н. О некоторых классах сложных систем. Изв. АН СССР. — Техн. кибернетика, 1964, № 6; 1965, № 1, № 3.
28. Коваленко И. Н. Анализ редких событий при оценке эффективности и надежности систем. М., Сов. радио, 1980.
29. Кокс Д., Смит В. Теория восстановления. М., Сов. радио, 1967.
30. Колмогоров А. Н. Основные понятия теории вероятностей. М., Наука, 1974.
31. Ла-Салль Ж., Лефшец С. Исследование устойчивости прямым методом Ляпунова. М., Мир, 1964.
32. Лоренц А. А. Элементы конструктивной теории вероятностных автоматов. Рига, Зинатне, 1972.
33. Рубальский Г. Б. Управление запасами при случайном спросе. М., Сов. радио, 1977.
34. Сахобов О., Соловьев А. Д. Двусторонние оценки надежности в общей модели резервирования с одной ремонтной единицей. Изв. АН СССР. — Техн. кибернетика, 1977, № 4.
35. Сильвестров Д. С. Метод одного вероятностного пространства в теореме восстановления. — Теория вероятностей и ее применения, 1979, т. 24, № 3.

36. Соловьев А. Д. Основы математической теории надежности. Вып. 1, 2. М., Знание, 1975.
37. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1, 2. М., Мир, 1967.
38. Форрестер Дж. Мировая динамика. М., Наука, 1978.
39. Хедли Дж., Уайтин Т. Анализ систем управления запасами. М., Наука, 1969.
40. Хейт Ф. Математическая теория транспортных потоков. М., Мир, 1966.
41. Emelianov S. V., Kalashnikov V. V., Soloviev M. M., Srednikova L. P. Imitative modelling of industrial enterprises in management, Preprints of the 7-th Triennial World Congress of the IFAC. Pergamon Press, 1978.
42. Foster F. G. On the stochastic matrices associated with certain queueing processes. The Annals of Math. Stat., 1953, v. 24, N 2.
43. Lindvall T. A probabilistic proof of Blackwell's renewal theorem, The Ann. of Probability, 1977, v. 5, N 3.

СОДЕРЖАНИЕ

Владимир Вячеславович КАЛАШНИКОВ
СЛОЖНЫЕ СИСТЕМЫ И МЕТОДЫ ИХ АНАЛИЗА

Главный отраслевой редактор Л. А. Ерлыкин
Редактор Г. Г. Карвовский
Мл. редактор Т. Г. Иншакова
Обложка художника Л. П. Ромасенко
Худож. редактор М. А. Бабичева
Техн. редактор А. М. Красавина
Корректор В. В. Каночкина

ИБ № 2967

Сдано в набор 17.07.80 г. Подписано к печати 22.8 80 г.
Т-13977. Формат бумаги 84×108[1]/32. Бумага 3.
Гарнитура литературная. Печать высокая. Усл. печ. л. 3,36.
Уч.-изд. л. 3,48 Тираж 35 470. Заказ № 1711
Цена 11 коп. Издательство «Знание». 101835. ГСП,
Москва, Центр, проезд, Серова, д. 4.
Индекс заказа 804309
Чеховский полиграфический комбинат
Союзполиграфпрома Государственного комитета СССР
по делам издательств, полиграфии и книжной торговли
г. Чехов Московской области

11 коп.

Индекс 70096