

Министерство образования и науки Российской Федерации
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

В.В. ГУБАРЕВ

ВВЕДЕНИЕ В ТЕОРЕТИЧЕСКУЮ ИНФОРМАТИКУ

Часть 1

Утверждено
Редакционно-издательским советом университета
в качестве учебного пособия

НОВОСИБИРСК
2014

УДК 004(075.8)
Г 93

Рецензенты:

Д.Е. Пальчунов, зав. кафедрой общей информатики,
зав. отделом ИДМИ НИЧ НГУ, д-р физ.-мат. наук, профессор;
В.К. Попков, гл. научн. сотрудник ИВМ и МГ СО РАН,
д-р физ.-мат. наук, профессор

Работа подготовлена на кафедре вычислительной техники для студентов вузов, обучающихся по укрупненной группе специальностей «Информатика и вычислительная техника» (бакалавриат, магистратура, аспирантура)

Губарев В.В.

Г 93 Введение в теоретическую информатику : учеб. пособие / В.В. Губарев. – Новосибирск : Изд-во НГТУ, 2014. – Ч. 1. – 420 с.

ISBN 978-5-7782-2477-3

В пособии с единых методических позиций излагаются основные элементарные сведения о моделировании и экспериментировании как методах исследования объектов, метрологии и теории сигналов. Особое внимание с позиции модельного представления объектов уделено элементарным основам следующего математического аппарата описания и исследования объектов: детерминированного, стохастического, нечеткого; интервального; поиска оптимальных решений; динамического хаоса, фракталов, теории игр и массового обслуживания.

Пособие ориентировано на подготовку бакалавров, магистров и аспирантов по направлению 09.00.00 – «Информатика и вычислительная техника». Оно может быть полезным для студентов, обучающихся по направлениям 02.00.00 – «Компьютерные и информационные науки», 10.00.00 – «Информационная безопасность», 11.00.00 – «Электроника, радиотехника и системы связи», 27.00.00 – «Управление в технических системах», а также специалистам в перечисленных и смежных отраслях деятельности.

УДК 004(075.8)

ISBN 978-5-7782-2477-3

© Губарев В.В., 2014

© Новосибирский государственный
технический университет, 2014

ВВЕДЕНИЕ

В настоящее время понятие *информатика* часто употребляется, но, к сожалению, неоднозначно понимается. Дело дошло до того, что дисциплина «Информатика», став обязательной в государственных образовательных стандартах (ГОС) общего среднего и высшего профессионального образования, по своему содержанию, вкладываемому разработчиками в федеральные ГОС по разным направлениям и, что особенно важно, реализуемому в разных образовательных учреждениях и даже в отдельных структурных подразделениях одного учреждения, выглядит очень по-разному. Как следствие, лица, закончившие такие учреждения, пришедшие в бизнес, производство, управленческие структуры, зачастую принимают решения и выделяют значительные средства под работы, связанные с информатикой и ее разделами, по-разному понимая, что это такое и каких результатов следует ожидать, осваивая эти средства. Это особенно важно для выпускников вузов, обучавшихся по направлениям, входящим в укрупненную группу специальностей (УГС) «Информатика и вычислительная техника». Ведь даже само наименование УГС весьма неудачно, оно дезориентирует абитуриентов и выпускников вузов, на что специалисты уже указывали неоднократно (см., например, [1]).

Сегодня сводить информатику к одному из ее разделов, как, например, математику – только к арифметике или алгебре, а тем более, противопоставлять информатику как соборное понятие (аналогичное понятиям «математика», «физика», «биология» и т. п.) какой-то из ее частей, как в сочетании «Информатика и вычислительная техника», является непростительным анахронизмом, весьма не безобидной инерцией старых взглядов, учений, мышлений. Одно дело, если так, по-старому, думает и согласно этому действует неспециалист, – это всего лишь заблуждение. Но если так думает, пишет и преподает ученый, педагог – это невежество, а если управленец – это преступление, вредительство, в лучшем случае неосознанное.

Очень важно в начале преподавания информатики в вузе и на завершающей стадии в школе давать представление о структуре инфор-

матики, о ее основных составляющих и их связи между собой. Это несомненно следует делать в первом семестре для студентов бакалавриата УГС «Информатика и вычислительная техника» в вводных дисциплинах типа «Введение в специальность» или в подобных более узких дисциплинах, важных для общей ориентации в конкретных составных частях информатики, например таких как формальная, техническая, технологическая, прикладная информатика, а также на другом, обобщающем и проблемном, уровне в магистратуре.

Именно на получение обучаемыми первичного представления о том, что такое *формальная информатика*, каков ее аппарат, объекты и предметы, направлено настоящее учебное пособие.

Заметим, что термин *теоретическая информатика*, как, впрочем, многие другие термины (см. далее), не является однозначным. Это связано с тем, что очень часто в приложении к научной дисциплине термин *теоретическая* (например, физика, химия, биология) в отличие от термина *практическая* или *прикладная*, рассматривается как синоним *базовая*, *фундаментальная* (см. далее, § 1.2) либо связывается с понятием *теоретические основы* в прикладных дисциплинах (см. § 1.2). Тем самым в сочетании *теоретические основы* он может относиться к разным составным частям информатики, таким, например, как техническая, технологическая, прикладная информатика. С другой стороны, многие под словом *теоретический* понимают основанный только на умозрительных соображениях, абстракциях, отвлеченно-логических операциях или математических выкладках, расчетах, т. е. использующий какую-то часть формального аппарата (например, математического) либо любые его разновидности. Именно в этом смысле мы и будем использовать этот термин в пособии (см. также § 1.5).

Глава первая

ИСХОДНЫЕ ОПРЕДЕЛЕНИЯ И ПОНЯТИЯ

§ 1.1. ЧТО ТАКОЕ МОДЕЛЬ? ТЕРМИНЫ КАК МОДЕЛИ ОПРЕДЕЛЯЕМЫХ ИМИ ПОНЯТИЙ

Одним из центральных первичных понятий информатики является термин *модель* [1–3].

Этот термин неразрывно связан с еще четырьмя понятиями: *объект*, о модели которого идет речь; *цель* введения и использования модели объекта; *субъект* и *среда*, в которой субъект производит любые операции с моделью. Иными словами, когда мы говорим о модели объекта, мы имеем в виду четырехместное отношение объект–модель–субъект–среда (см. [2] и рис. 1.1 в [3, 4]).

К ним примыкает еще одно используемое для краткости обобщенное понятие *исследование* (объекта, модели), под которым в зависимости от контекста будем понимать собственно исследование (изучение, познание), а также описание, представление, управление, проектирование, анализ, синтез, передачу и т. п. (объекта, модели). Под ***объектом*** (object) понимается некоторая вещь, процесс, сигнал, свойство, событие, явление и т. п., т. е. некоторый ***оригинал (объект-оригинал)***, который мы замещаем моделью – новым, вспомогательным, используемым для исследования ***объектом-образом***, существующим в действительности, т. е. имеющим объективный характер, либо являющимся виртуальным (например, проектируемым, синтезируемым) – не существующим, но могущим существовать при определенных условиях, или абстрактным – мысленным, существующим в воображении, представляемым в виде набора символов, графиков, изображений и т. п. В частности, объектом-оригиналом может выступать модель некоторого первичного объекта, которую мы хотим заменить новым образом, новой моделью, более подходящей для новой цели или в новых

условиях исследования субъектом первичного объекта. С другой стороны, в качестве модели может выступать сам объект-оригинал, когда, например, мы используем в исследовании только часть его особенностей, свойств, характеристик, а выводы будем делать обо всем объекте.

Под *субъектом* понимается человек (или заменяющее его средство), работающий с моделью (создающий, исследующий, использующий ее).

Цель – то, к чему стремится, чего хочет добиться субъект, создавая модель объекта, т. е. модель того желаемого будущего, которое хотел бы получить субъект, располагая моделью.

Модель объекта есть его целевой образ

Под *целевым образом объекта-оригинала* понимается результат такого отображения, отражения (воспроизведения, представления, передачи, воплощения, приведения в соответствие) его моделью, которое передает все необходимое для цели создания и применения модели (моделирования, см. далее): существо, механизмы и особенности его строения и свойств, функционирования (жизни), закономерностей и т. п. Желательно, чтобы этот образ как результат отображения объекта-оригинала обладал необходимым качеством, мог использоваться в той среде, в которой будет применяться субъектом, отвечая при этом соответствующему назначению и задачам, требованиям качества (см. далее § 2.1, требования к моделям, и рис. 1.1).

Под особенностями объекта здесь и всюду далее мы будем понимать отличающие его от других объектов принципы строения и функционирования, структуры (состав и связи элементов), механизмы, правила и закономерности появления, существования, развития, функционирования, «смерти», характеристики и параметры, реакции на внешние и внутренние воздействия и т. п.

Наконец, под *средой*¹ будем понимать: термины, цели и задачи исследования; исходные и полученные данные и результаты; условия «жизни» объекта и получения сведений о нем; измерительные шкалы; виды средств (аппаратные, программные, лингвистические, логистические, метрологические и др.) и технологий сбора, обработки, анализа, исследования, моделирования, интерпретации, применения; правила, способы, модели и их характеристики; методы, алгоритмы и т. п.

¹ В зависимости от контекста – при моделировании, экспериментировании, обработке, ...

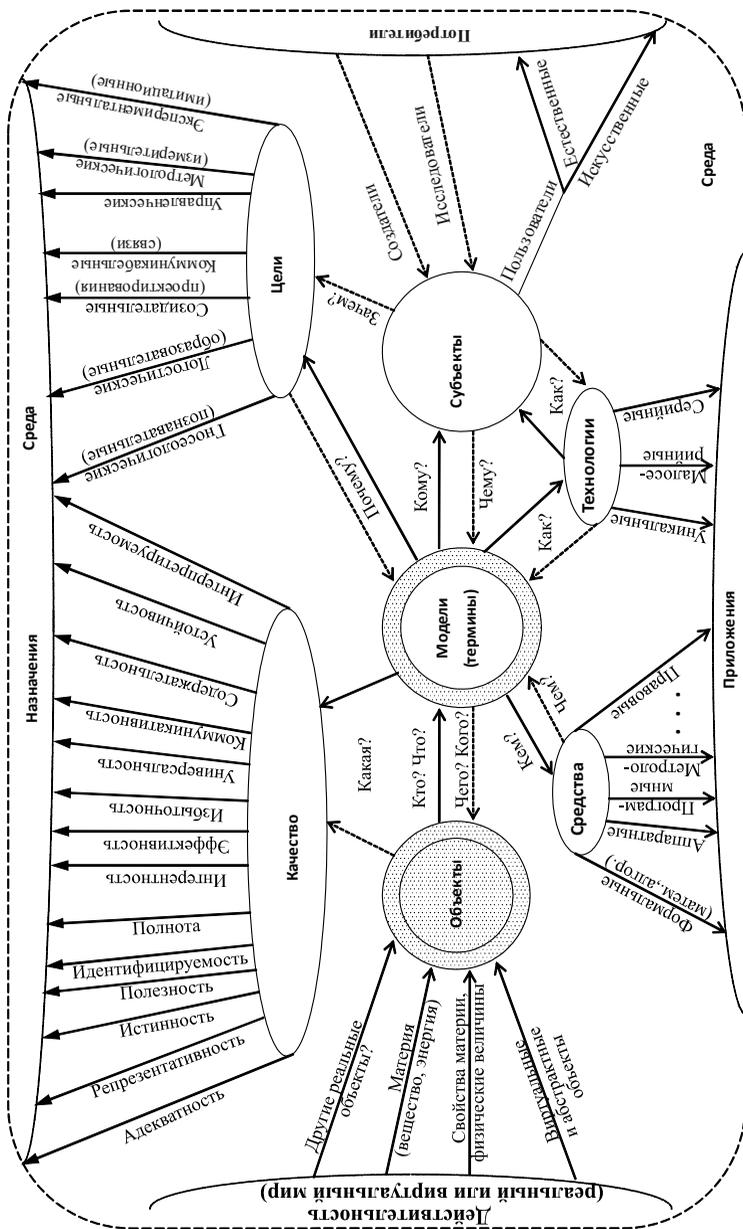


Рис. 1.1. Схема отношений объект–модель (термин)–субъект–среда:
 ---- создание (построение), исследование; — применение (потребление)

В связи с изложенным сделаем два замечания.

1. В системном подходе (см. далее) понятие *среда* для исследуемого объекта трактуется как все наружное по отношению к нему, т. е. все то, что окружает объект, во что он встроен. Ясно, что это модельное понятие, и поэтому под *средой* реально понимается лишь только та часть общей среды, которая влияет на объект, взаимно связана с ним с точки зрения цели и задач исследования объекта. В системологии (см. далее) эта часть среды часто называется надсистемой, элементом которой является системно модельно представляемый объект. В моделировании же (см., например, [2]) она называется микро-, культурной, внутренней средой, в которую встроена или в которой функционирует система из трех элементов: модельно представляемый объект, его модель и работающий с объектом и моделью субъект (см. рис. 1.1). В дальнейшем именно в этом микропонимании мы будем использовать слово *среда*.

2. Описанное понимание модели как любого целевого (в указанном ранее контексте) образа объекта-оригинала отражает современное положение дел. Первым же, историческим, можно считать следующее понимание термина модели: модель есть вспомогательный объект (средство)-заменитель, который в конкретной ситуации способен заменить сам объект-оригинал и при этом, во-первых, более удобен в применении, чем объект-оригинал; во-вторых, воспроизводит все интересующие пользователя модели свойства и характеристики объекта-оригинала [2]. Ясно, что такое понимание является частным случаем современного.

Важна для данной дисциплины и трактовка слова *термин*. Под ним обычно принято понимать слово или словосочетание, точно обозначающее определенное понятие, применяемое в какой-то специальной профессии или области человеческой деятельности: науке, технике, искусстве, ...

Термин есть не что иное, как модель определяемого им понятия. Поэтому к нему применимы все свойства модели, к рассмотрению которых, а также места (статуса) и назначения модели, мы и перейдем [3].

Свойство целевости (место и роль модели в деятельности субъекта). Модель есть неотъемлемый элемент любой *целенаправленной*, в частности познавательной, деятельности человека в силу следующих обстоятельств. Во-первых, любая деятельность человека направлена на достижение определенной цели. Во-вторых, цель есть организующий элемент деятельности, поскольку она служит моделью (образом) жела-

емого будущего, т. е. того состояния, на достижение (реализацию) которого направлена деятельность. В-третьих, человек, делая что-то, имеет, как правило, явно или неявно схему, образ (модель, алгоритм, правила) деятельности. В-четвертых, в познавательных задачах модель как искомый образ исследуемого объекта есть продукт и/или инструмент целенаправленной деятельности субъекта-исследователя. Наконец, в-пятых, важно иметь в виду, что человек (субъект) не мыслит иначе, как категориями моделей, и поэтому любые утверждения могут быть сделаны им лишь только относительно модели.

Модель есть неотъемлемый элемент (цель, образ, инструмент, продукт) человеческой деятельности. **Цель** – это модель желаемого будущего, а **модель** – результат целевого отображения объекта-оригинала.

Свойство встроенности (отношения «объект–модель–субъект–среда»). Очевидно, что модель (в частности термин) не может существовать без субъекта (ее создателя, исследователя или пользователя) и без ингерентной (согласованной с моделью) среды (целей и задач, показателей качества, условий применения, лиц, средств и технологий, с помощью которых создается, функционирует и применяется модель), обеспечивающей ее функционирование и порождающей интерес к ее использованию микросреды, «культурной» среды, макросреды (рис. 1.1). Приведите примеры, поясняющие это.

Модель не является изолированной, а должна рассматриваться только во взаимосвязанном многоместном отношении объект–модель–субъект–среда.

Свойство дуальности (предназначение и истинность моделей). Суть свойства в проявлении моделью ее двойственности, двоякого характера или даже тройственности, трехсторонности, когда одно проявление не совпадает (в двойственности даже по сути) с другими, исключает их. Поясним его.

Все многообразие моделей (в том числе терминов) условно (!) можно разделить на три класса [1–4]: исследовательские (познавательные), созидательные (прагматические) и договорные.

Целевая ориентация *исследовательских* моделей такая же, как в познавательной деятельности: модель должна как можно лучше представить действительность (объект-оригинал). При обнаружении расхождений между познавательной моделью и действительностью уси-

лия субъекта должны быть направлены на изменение¹ модели, повышение степени приближения ее к действительности. Познавательные модели – физические законы, теории, математические модели. Познавательные термины – «жизнь», «земля», «фотокопия». Направленность *созидательной* (прагматической) модели – как можно полнее представить образ будущего (проектируемого, создаваемого) объекта. В случае обнаружения отклонения создаваемого объекта от модели как его образца, эталона задачей создателя объекта является приближение объекта к модели, а не модели к объекту, как в исследовании. Созидательные модели – юридические законы, стандарты, технические задания на изделия, цели, планы, уставы, алгоритмы. Созидательные термины – «коммунизм», «проект».

Договорные (конкордовые) модели устанавливаются одними и признаются другими (например, буквенные или математические символы +, –, :, ∫, ∑, Π, модели математики) либо являются продуктами соглашений, договоров двух и более физических, юридических (в частности официальных органов) лиц и сообществ (например, купюры денег, сигналы светофоров и т. п.). Направленность таких моделей – упрощение человеческих коммуникаций. Изменение договорных моделей происходит при потере ими своих преимуществ или недостатке возможностей перед другими моделями, средствами, служащими тем же целям.

Свойство дуальности моделей проявляется в том, что в зависимости от цели, назначения одна и та же модель может выступать как познавательная, созидательная или договорная.

Второй пример двойственности – когда модель может являться одновременно «истинным», «верным» и «неверным» отражением оригинала. «Истинным», «верным» – значит, отражающим объект таким, какой он есть, существует вне и независимо от субъекта всегда, т. е. безусловно или условно (при определенных условиях), а «неверным» – значит, не имеющим отношения к оригиналу, не отражающим оригинал таким, какой он есть на самом деле. Примером проявления такой дуальности модели служит корпускулярно-волновой дуализм: согласно ему любые микрообъекты (фотоны, электроны, протоны, атомы и т. д.) обладают свойствами и частиц (корпускул), и волн. Так, волновая модель света в одних условиях является «истинной», «верной» и в

¹ В широком смысле – самой модели, микросреды, цели построения и условий ее применения.

то же время «неверной», не относящейся к объекту-оригиналу, или «условно верной», если свет рассматривается как корпускула с точки зрения иного подхода.

Свойство тройственности – когда в зависимости от назначения модель используется либо как исследовательская, либо как созидательная, либо как договорная.

Внимание! Приведите пример.

Третий вид дуальности рассмотрим на примере характеристик случайных сигналов как моделей некоторого физического объекта и/или экспериментальных данных о нем. Характеристики могут быть истинными моделями, если данные соответствуют свойствам объекта и/или модели, и неверными – в противоположном случае. Так, математические ожидания или средние арифметические могут быть истинными моделями характеристики положения, центра рассеяния данных, если они (данные) имеют распределение Гаусса, и неверными – для данных, которые имеют распределение Коши.

Исследовательские (познавательные) модели (термины) отражают реально существующее; **созидательные (прагматические)** – виртуальное, не существующее, придуманное, в том числе желаемое, но возможно осуществимое; **договорные** – соглашение об их применимости. Модель объекта может быть познавательной, прагматической, договорной; «истинной», «верной» или «ложной», «неверной» в зависимости от условий, свойств объекта и целей моделирования.

Свойство неоднозначности (отсутствие взаимной однозначности между моделями и объектами). Может ли между объектом-оригиналом и его моделью существовать взаимно однозначное соответствие и надо ли к этому стремиться? Для ответа на этот вопрос отметим следующие особенности отношений модель–объект:

- модель есть не сам объект-оригинал (хотя в физических исследованиях объект-оригинал и может выступать как модель!), а лишь его объект-образ, приближенно (насколько?) отражающий объект-оригинал;
- сам объект-оригинал не всегда имеет четко очерченные границы, отделяющие его от других «родственных» ему объектов и среды. Например, что представляют собой объекты, моделями которых явля-

ются термины «жизнь», «лес», «энтропия», «земля», «семья», «красивая девушка», «атмосфера» и т. п.?

- модель есть целевой ингерентный среде объект-образ объекта-оригинала, т. е. образ, адекватный назначению, цели, задачам, согласованный с требованиями к качеству, средствам, условиям и правилам моделирования. Следовательно, объект-оригинал может иметь множество моделей, отвечающих разным назначениям, целям, задачам, условиям;

- одна и та же модель (особенно термин) может соответствовать многим объектам (например, математическая модель каких-либо объектов или слова «коса», «энтропия»).

Сказанное означает, что, во-первых, между объектом и его моделью не может существовать строгого взаимно однозначного соответствия; во-вторых, создатели модели должны стремиться к таким ее построению и применению, которые отвечали бы требованию однозначной (лучше взаимно однозначной) интерпретируемости получаемых при моделировании результатов (адекватности цели, задачам, средствам, условиям); в-третьих, становится понятной сложность задачи построения таких моделей, особенно относительно универсальных по целям, задачам, объектам.

<p>Взаимно однозначное соответствие между моделью и объектом само есть прагматическая модель как одно из требований к моделированию, желаемое, но не всегда достижимое.</p>
--

Свойство финитности (ограниченность моделей). Из свойства неоднозначности следует, что существует определенная ограниченность в применении моделей для познания и описания реального мира и синтеза новых объектов, которая, однако, не тождественна отрицанию возможности познания Вселенной или синтеза ранее несуществующего «бесконечного» конечными средствами. Эта ограниченность – следствие финитности (конечности) модели, проявляющейся в том, что модель отображает оригинал лишь в конечном числе главных (наиболее существенных) отношений (структур, свойств, правил функционирования), реализуется при конечных ресурсах в процессе ее построения и оперирования моделью, условиях ее применения, в частности за счет наличия помех, сбоев, ограниченной области применимости модели, ее упрощенности и т. д.

Область применения моделей всегда ограничена (финитна) из-за особенностей самой модели, ее ингерентности (согласованности со средой), условий оперирования с ней и ее использования.

Свойство динамичности («жизнь» моделей). Как уже указывалось, модель может быть полезной, лишь если она ингерентна, согласована со средой. Но среда не является стационарной, она эволюционно или революционно изменяется во времени. В связи с этим и модели проходят свой «жизненный» цикл: рождаются, развиваются, сотрудничают и соперничают с другими, «умирают», уступая место более качественным, совершенным. Жизненный цикл моделей может сопровождаться их естественными и/или искусственными эволюционными и революционными изменениями. Сказанное полностью относится к терминам.

Модели (термины) динамичны и имеют свой жизненный цикл.

Кроме того, многие термины обладают еще одним важным свойством – *расплывчатостью* при интерпретации.

Итак, термины как модели описываемых ими понятий обладают по крайней мере следующими свойствами: целевостью, встроенностью, дуальностью, неоднозначностью, финитностью, динамичностью, расплывчатостью. Поэтому как при рождении (создании, введении) термина, так и при его использовании необходимо быть очень аккуратными, тщательными, точно определять фактический смысл данного термина, отвечать на все вопросы и выбирать соответствующий путь по каждому элементу схемы, представленной на рис. 1.1. Важно также выяснить, какой смысл вложил в термин его создатель, т. е. каковы его ответы на вопросы, приведенные на рис. 1.1, как интерпретируется сейчас этот термин «культурной средой» и как пытается его интерпретировать читатель. При введении определения (дефиниции) и понимании термина очень важным, первичным является выделение соответствующего термину ключевого слова, которое означает суть термина, его существо, а затем уже других слов из его определения, дополняющих это слово. Например, *модель* объекта – это *образ* (целевой, вспомогательный, искусственный или естественный и т. д.); *математика* – наука о ...; *система* – модель или модельное представление...; *моделирование* – метод либо сам процесс исследования, ...; *сигнал, данные* – носители информации, а не сама информация и т. п.

§ 1.2. ЧТО ТАКОЕ НАУКА?

Одно из часто используемых в пособии слов – наука. Поясним его. Отметим два аспекта, связанных с этим термином.

С одной стороны, под наукой понимается:

1) система знаний о законах и закономерностях развития природы и общества, а также способах воздействия на окружающий мир (добавлю: и его преобразования);

2) отдельная отрасль (дисциплина) таких знаний;

3) навыки, знания, получаемые человеком в результате обучения или жизненного опыта [4].

С другой стороны, наука – это сфера человеческой деятельности, функция которой – выработка и теоретическая систематизация объективных знаний о действительности [3].

1. Итак, наука есть область человеческой деятельности. Она состоит из системы научных дисциплин, каждая из которых имеет свои объекты и предметы. Объекты – это то первичное, на что направлены усилия ученых этой дисциплины, а предметы – то, что исследуется в объектах. Кроме объектов и предметов каждая научная дисциплина должна иметь свою (собственную или заимствованную из других дисциплин) **методологию** (совокупность подходов), **методику** (совокупность методов), **методы** (совокупность способов, правил, приемов), **технологии**, а также **понятие и критерии истины** (точнее, истинности, верности результатов). При отсутствии критериев истинности результатов дисциплину, относимую к научной, нельзя считать таковой. Правильнее было бы к таким сферам человеческой деятельности придумать свой термин, аналог терминам «вера», «искусство», «мастерство» и т. п.

Внимание! Приведите примеры таких «квазинаук».

Наличие объектов, предметов, методологии и критериев истинности результатов лишь необходимый, но недостаточный набор показателей, по которым какую-то сферу человеческой деятельности можно считать научной. Помимо этих атрибутов любую научную дисциплину характеризуют свои отличительные особенности, отвечающие критериям новизны и научности результатов, такие как непротиворечивость, интерсубъективность (независимость от субъекта), проверяемость либо опровержимость результатов, их пригодность для предвидения и т. п. Именно наличие этих четко очерченных атрибутов позволяет считать некоторую область знаний и деятельности человека наукой и

по отличию хотя бы одного из них отделять различные научные дисциплины друг от друга. Например, по объектам исследования отличать физику от химии и математики, по методам, методологии исследования и понятию истинности отличать математику от всех других научных дисциплин.

2. Наукой (научной дисциплиной) может быть названа некоторая отдельная дисциплина, объекты и предметы которой являются первичными, не разлагаемыми на более простые в рамках данной науки (например, арифметика, алгебра, геометрия, теория вероятностей; механика, оптика, термодинамика и т. п.), либо совокупность дисциплин, объекты и предметы исследования в которых являются сборными (сборными, обобщающими, агрегативными), включающими в себя объекты и предметы первичных дисциплин (например, соборная наука математика включает в себя арифметику, алгебру, геометрию и т. д., а физика включает механику, оптику, термодинамику и т. д.).

3. Название научной дисциплины, как правило, отражает, с одной стороны, ее фундаментальную и прикладную составляющие, с другой – практическую область человеческой деятельности, базирующуюся на достижениях именно этой научной дисциплины. В связи с этим говорят о триединстве области деятельности, отражаемой термином по названию научной дисциплины, например информатики (см. далее), где объекты (внимания, исследования, разработки) дисциплины едины для всех трех проявлений этого единства (как фундаментальной науки, прикладной науки, добывающих соответствующие знания, так и области практической деятельности, применяющей эти знания).

4. Операндами (промежуточными «сырьем» и «продуктами» технологических операций) и результатами научной деятельности являются познавательные (исследовательские, гносеологические) или созидательные (прагматические) модели, а также договорные, удобные для коммуникаций. Поскольку модели являются целевыми образами объекта-оригинала, они отражают лишь те особенности объекта-оригинала, которые соответствуют цели моделирования, причем только в рамках принятых целей и допущений, применяются в условиях ограниченности ресурсов, обладают финитностью, дуальностью, неоднозначностью и другими свойствами модели. В силу этого один и тот же объект-оригинал зачастую изучается разными научными дисциплинами, оперирующими разными модельными представлениями об объекте-оригинале, построенными при разных допущениях и целях исследования.

5. Что касается целей научных исследований, следует иметь в виду их направленность и назначение. Например, для построения каких модельных представлений реальности, какой научной картины мира ориентирована данная научная дисциплина: *куомодной*, отвечающей на вопрос «Как, каким образом устроен мир?», *каузальной* («Почему так устроен мир?»), *телосной* («Зачем, с какой целью так устроен мир?»), *исторической* («Всегда ли он был так устроен?») и т. д.

6. По подходу к исследуемому объекту и предмету научные дисциплины можно разделить на *редукционные*, *анализовые* (менее удачное название – аналитические) и *идентичные*, *традукционные* (одинаковой общности, сложности), *синтезовые* (синтетические). В дисциплинах первой группы объекты исследования представляются как простые (аддитивные) физические системы, т. е. как целое, состоящее из частей и допускающее:

а) рассмотрение его без обязательного учета связи с окружающей средой или даже изолированно от нее;

б) наличие принципа суперпозиции;

в) познание (в том числе описание) и синтез методами анализа, т. е. расчленением целого на части, изучением (описанием) и синтезом модельного представления его по частям. Для дисциплин второй группы обязательно системологическое (модельное системное) представление объектов, т. е. представление их модельно именно как системы, обладающей соответствующими свойствами системы, в частности свойствами эмерджентности, а не аддитивности. Из него следует невозможность сведения сложных систем к простым, т. е. невозможность только аналитического познания (а следовательно, описания и синтеза) таких объектов. Ко второй группе дисциплин относятся частично кибернетика и особенно синергетика. Из изложенного ясно, что модели и методы первой группы дисциплин не заменяют, а дополняют модели и методы второй группы, обеспечивая тем самым наилучшее представление действительности, когда разные модели являются ингерентными и касаются разных аспектов строения, состава, свойств, правил, механизмов и условий жизни (функционирования) таких объектов, в приложении к которым они являются адекватными среде, объекту, цели, условиям, средствам, технологиям исследования.

7. Ранее мы отметили триединство науки, а именно наличие в ней трех взаимосвязанных частей: фундаментальной, прикладной и практической. Назначение фундаментальной и прикладной наук (частей науки как целого) – выявление, открытие, исследование (включая

управление, проектирование) соответствующих объектов-оригиналов и предметов в них, в то время как назначением практической деятельности являются разработка, создание и эффективное применение ее объектов и предметов на базе результатов первых двух частей данной науки. Например, в кибернетике практическая деятельность связана с разработкой конкретных методов и алгоритмов, позволяющих управлять кибернетическими системами для того, чтобы они функционировали заранее заданным образом, в синергетике – позволяющих изменять условия функционирования синергетических систем, «управляя» факторами («причинами») самоорганизации, а в комбинации кибернетико-синергетических приемов (в синергетическом управлении) – обеспечивая искусственный целевой (направленный) способ самоорганизации путем уменьшения избыточных степеней свободы и формирования желаемых *аттракторов* (точек притяжения поведений) за счет выбора (усиления, ослабления) соответствующих обратных связей в системе и самодостраивания ее, самодоводки до желаемого состояния (см. далее).

Отличие фундаментальной составляющей в том, что она направлена прежде всего на получение обобщающих результатов, отражающих глубинные основы объектов и ориентированных на внутренние потребности интересов науки, в то время как прикладная наука ориентирована на получение результатов, пригодных к практическому применению, выступает посредником между фундаментальной наукой и соответствующей областью или областями практической деятельности. Сходство же фундаментальной и прикладной наук в том, что они занимаются именно *исследованиями*, получением знаний – познавательных и созидательных моделей и разработкой теоретических основ, в то время как практическая деятельность занимается *разработкой, созданием и применением* конкретных объектов, методов, средств, технологий на базе результатов фундаментальных и прикладных исследований. В этом смысле деление объектов и предметов между фундаментальной, прикладной науками и областью практической деятельности является относительным, условным. Предметы практической деятельности могут выступать, например, объектами исследования прикладной науки с использованием методологии как своей, так и других наук, в то время как предметы прикладной науки могут быть объектами разработки практической деятельности или предметами фундаментальной науки в зависимости от общности и глубины познания. Компактно это представлено на рис. 1.2.

ЧТО ТАКОЕ НАУКА?

1. СОСТАВ	<ul style="list-style-type: none"> ● объекты исследования ● предметы исследования ● методология (методы, средства и приемы) ● понятие истины (истинности)
2. КРИТЕРИИ НАУЧНОСТИ (НАУЧНОГО ПОЗНАНИЯ)	<ul style="list-style-type: none"> ● непротиворечивость ● интерсубъективность (независимость от субъектов) ● проверяемость (подтверждаемость (верификация) + опровержимость (опытом), фальсификация) ● пригодность для предсказания и изменчивость (вложенность, ингерентность)
3. РЕЗУЛЬТАТЫ	<ul style="list-style-type: none"> ● <i>фундаментальной науки</i> – модели прошлого, настоящего и будущего исследуемого объекта как целевые образы, адекватно отражающие объект ● <i>прикладной науки</i> – модели, составляющие теоретические основы создания объектов, обладающих требуемыми особенностями, назначением, качеством
4. ВИДЫ НАУЧНЫХ ЗНАНИЙ	<ul style="list-style-type: none"> ● <i>редукционные (анализовые)</i>: допускают обособленные аналитические методы познания, используют принцип суперпозиции. Пример – классические физические знания ● <i>идентичные, традукционные, синтезовые</i>: основаны на системологическом представлении (целовость, цельность, эмерджентность (эмергентность), ...) и невозможности только аналитического познания
5. ТИПЫ НАУЧНЫХ ДИСЦИПЛИН	<p>5.1.1. Строго формализованные – исходные понятия строго формализованы (например, математика)</p> <p>5.1.2. Не формализованные – исходные понятия не строго определены, не формализованы или вовсе стартово не определены и являются объектом либо предметом исследования в этой дисциплине</p> <p>5.2.1. Прямые – дисциплина и ее наименование являются первичными, неразложимыми на более простые в рамках данной науки</p> <p>5.2.2. Сборные (агрегированные) – дисциплина и ее исследования являются обобщенными, вторичными и включают в себя первичные дисциплины, более простые, входящие в нее</p> <p>5.3.1. Фундаментальные – ориентированные на познание общего в науке, построение их исследовательских моделей и, как вспомогательных, договорных</p> <p>5.3.2. Прикладные – ориентированные на построение теоретических основ описания объектов науки, их созидательных моделей и, как вспомогательных, договорных</p>

Рис. 1.2. К термину «наука»

§ 1.3. ИНФОРМАЦИЯ («РАБОЧЕЕ» ПОНЯТИЕ) И ЕЕ НОСИТЕЛИ

Несмотря на то что слово *информация* широко используется, до сих пор нет его общепринятого понимания. Более того, очень часто слово *информация* отождествляется с ее носителями¹, а теория информации – с теорией передачи и кодирования данных. Более подробно проблемы с этим термином рассмотрим в § 4.1. Здесь же введем «рабочее» определение, противопоставляющее его другим, связанным с носителями информации. Для этого вначале приведем, а затем прокомментируем необходимые понятия, используемые в дальнейшем.

Сигнал: 1) в теории эксперимента: «внутриобъектный» физический носитель (источник, поставщик) информации, недоступный непосредственному восприятию субъектом; 2) в теории связи и управления: физическое средство хранения и передачи (переносчик) информации в пространстве и времени.

Данные – модельное представление исследуемых, в том числе физических, объектов в виде доступной для восприятия субъектом («внеобъектной») совокупности (набора) символов, записей, чисел, изображений, ..., в частности числовых или графических значений отсчетов сигналов, хранящейся на материальных носителях (хранителях!) информации и рассматриваемой как носитель (источник!) информации об объекте безотносительно к ее содержательному смыслу, т. е. в отрыве от содержащейся в ней (совокупности) информации.

Знания – модельное представление об исследуемом объекте в виде проверенных практикой, требующих постоянного дополнения и оценки истинности результатов познания действительности: гипотез, идей, теорий, выводов, понятий, законов, закономерностей, концепций, ... С точки зрения связи терминов *знания* и *информация*, знания следует разделить на два вида, типа, рода: *знания 1* (знания 1-го рода, анзнания [1]) – операнды (объекты) сбора, обработки, анализа как носители информации; *знания 2* (знания 2-го рода, знания [1]) – элемент, разновидность, часть информации, семантика, полезные сведения о действительности, об объекте.

¹ Подобные факты встречаются, к сожалению, нередко. Так, часто модель реальности отождествляется с самой реальностью, как в школьных учебниках при формулировке и интерпретации физических законов, в частности третьего закона Ньютона (см. [3, с. 5]).

В связи с изложенным обратим внимание на принципиальное отличие сигналов, данных и знаний-1 как носителей информации: сигналы есть физические элементы действительности, физического объекта, т. е. физические носители информации, а данные и знания – модели объекта, т. е. «модельные» носители информации.

Сигналы – физические носители информации
Данные, знания-1 (анзнания [1]) – модельные носители информации

Контент – любое информационно значимое наполнение информационного ресурса.

Данные – сигналы ∨ данные ∨ знания-1 (анзнания) ∨ контент.

Протознания – часть информации, которая может быть преобразована в знания об объекте согласно цели его исследования.

Новости: 1) новое, только что полученное сообщение, известие; 2) то содержательное, что является для субъектов новым, ранее неизвестным.

Информационный мусор – содержащиеся в Данных семантические сведения, не имеющие полезных для пользователя знаний об исследуемом объекте, но многократно увеличивающие издержки пользователя на их хранение, передачу, обработку как составной части информации.

Информация: 1) совокупность полезных содержательных сведений, имеющихся в Данных об исследуемом объекте; 2) сведения, смысл, программа действий, конструкций и т. п., содержащиеся в сигналах, данных, знаниях, контенте, кодах как ее носителях; 3) семантика сигналов, символьных сообщений, чисел, записей, изображений; 4) модели (знания: теоремы, законы, образы; программы, конструкции, ...) ∨ протознания (протомодели) ∨ информационный мусор.

Сбор Данных – действия с Данными, направленные на их съем, восприятие, получение, передачу, накопление, ..., не приводящие (по своему назначению) к изменению имеющейся в них информации, их смыслового содержания, семантики.

Обработка Данных – действия с Данными, направленные на преобразование содержащейся в них и интересующей исследователя информации к компактному, удобному для хранения, передачи, исследования и анализа виду, приводящие к априори допустимому изменению семантики, ценности, секретности, избыточности, защищенности, эстетического содержания и других особенностей информации, находя-

щейся в них. Обработка не добавляет содержащейся в Данных информации.

Анализ, исследование Данных – действия с Данными, направленные на извлечение имеющейся в них информации об исследуемом объекте, на получение по имеющимся Данным новых Данных, содержащих в себе извлеченную из первых информацию об объекте, на выявление имеющихся в них особенностей и закономерностей, упрощение интерпретации получаемой информации и обработку Данных.

Интерпретация результатов обработки, исследования и анализа Данных – истолкование, разъяснение смысла, значения результатов, их «перевод» на язык, в термины, образы, ..., доступные, понятные пользователю (результатов).

Применение результатов обработки, исследования и анализа Данных – использование результатов пользователем для решения теоретических и практических задач.

Технологический процесс – последовательность направленных на получение заданного результата (продукта) физических и мысленных действий (технологических операций), каждое из которых основано на использовании каких-либо естественных процессов (физических, химических, биологических и др.) и человеческой деятельности или работы заменяющих его (человека) автоматов.

Прежде всего еще раз обратим внимание на разделение понятий *информация* и *ее носители*: сигналы, данные, знания-1¹, контент, а также надные = сигналы ∨ данные ∨ анзнания. Теперь поясним разницу между терминами сигнал, данные, знания как носители информации.

Поясним эти отличия на примере температуры тела (или давления, роста, массы, ...) человека. Температура тела человека служит сигналом о состоянии его (человека) здоровья, находится «внутри» человека («внутри объекта»), «живет» с ним, ежемоментно представляя своим истинным значением состояние организма, свойственное именно этому индивиду, недоступна (без измерения!) кому-либо. Чтобы информация, содержащаяся в этом сигнале о состоянии здоровья человека, была доступна для применения, температуру надо измерить. В результате измерения получаем данные, представляющие собой значения

¹ В [1, 3–5] для обеспечения четкости разделения понятий *знание-1*, как носителя информации, и знания-2, как ее части, для знания-1 предложен термин *анзнания*, оставив термин *знания* только для обозначения понятия знания-2.

отсчетов сигнала (измеренные, конечно, с какой-то погрешностью) в фиксированные моменты времени (также измеряемые с погрешностью). Они тоже несут информацию о состоянии здоровья пациента. Однако она, с одной стороны, уже, как правило, несколько искажена за счет неточности измерения значений как самой температуры, так и момента времени, к которому относится этот результат измерения, с другой, – оторвана от объекта. Что это значит? Во-первых, такое же значение температуры могут в то же время иметь разные люди, тогда как температура конкретного человека (ее динамика, значение, состояние) уникальна. Во-вторых, измеренное значение характеризует состояние пациента в момент измерения, а не сейчас, вчера, завтра, ежемоментно. В-третьих, результат может быть представлен в разных шкалах, придуманных людьми, в том числе слабых, типа усредненных понятий: большая, маленькая, нормальная, весьма относительно отражающих индивидуальные особенности человека. Для одного пациента нормой может быть 36,7 °С, для другого – 36,5, для третьего – 37,0 °С. Причем, эта норма – усредненная в течение суток. Знаниями здесь служат усредненные сведения о том, какому виду болезни какие значения температур могут соответствовать, какова динамика поведения температуры при развитии болезни, ее ожидаемые и реальные реакции на лекарства и т. д. При этом знания-1 – набор усредненных сведений об этом, собранных, хранящихся, например, в справочниках, а знания-2 – конкретные сведения о состоянии наблюдаемого пациента, их отражение в значениях температуры и многое другое, на основании чего врач будет принимать решение о состоянии здоровья пациента (рис. 1.3).

Уяснив принципиальную разницу между понятиями *Данные* и *информация*, следует отметить, что допустимо (но не желательно!) использовать их (что часто и делается!) как синонимы. Это имеет место, когда из контекста ясно, о чем идет речь, т. е. когда мы реально совершаем операции с Данными, а имеем в виду информацию, в них содержащуюся. Заметим, кстати, что информация в ряде таких операций может не претерпевать изменений. Поясним сказанное.

В процессе эксперимента мы получаем (измеряем, снимаем, воспринимаем), накапливаем, передаем, храним, упорядочиваем, структурируем, представляем (отображаем) и совершаем другие действия над Данными, не приводящие (по функциональному назначению этих действий) к изменению смыслового содержания, ценности, секретности и других особенностей информации, содержащейся в них. Условно все эти действия (операции) можно, учитывая некоторую некорректность

замены совокупности различных по смыслу терминов одним, определить единым обобщающим понятием – *сбор Данных* (об объекте), что в рассматриваемом контексте эквивалентно (но не равнозначно!) словосочетанию *сбор информации* (об объекте). Заметим, что если при сборе осуществляется отказ от всех возможных Данных (генеральной совокупности) в пользу специально получаемой их части – *выборки*, мы имеем дело с выборочными методами исследования.

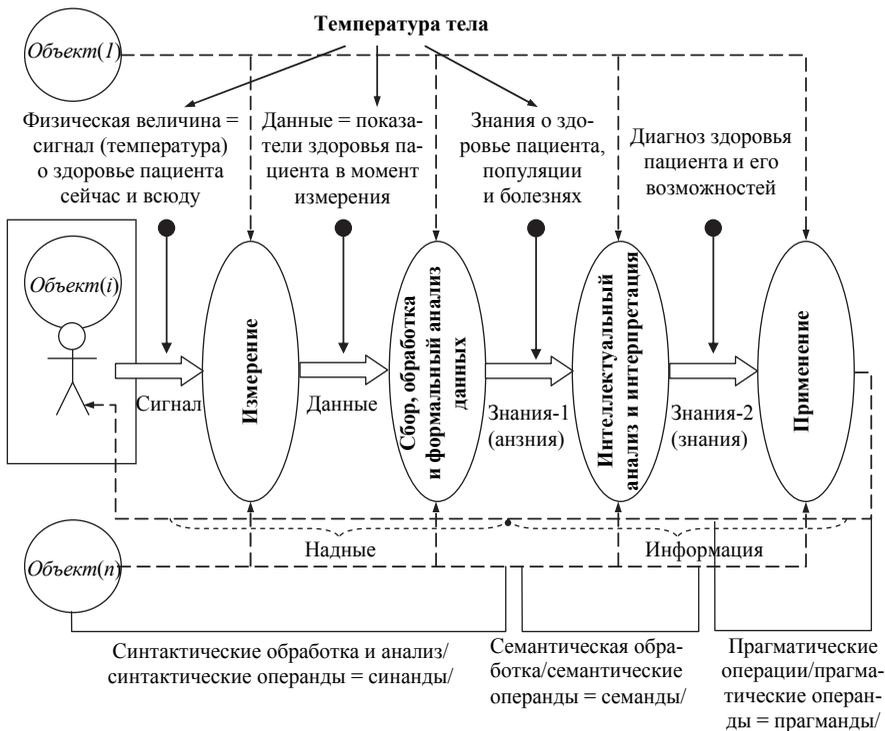


Рис. 1.3. Пояснение понятий: сигнал, данные, знания; синанды, семанды, прагманды

Аналогично, вместо терминов *преобразование, сокращение объема, сжатие, распознавание, кодирование, генерирование, защита* и других операций над Данными, обозначающих формализованные действия, направленные на приведение интересующей исследователя информации, которая содержится в Данных, к более удобному, компактному

для анализа виду (т. е. действия, связанные с частичным априори прогнозируемым изменением смысла, семантики, ценности, полезности, важности, секретности, эстетического содержания и других особенностей информации, содержащейся в этих Данных), будем применять термин *обработка Данных*, что эквивалентно (в употребляемом контексте) словосочетанию *обработка информации*. Например, при обработке выборки результатом ее чаще всего являются выборочные характеристики, используемые самостоятельно или как оценки характеристик всей генеральной совокупности. Подобная обработка Данных называется их *сверткой*, или *сверткой информации* [5].

Из изложенного ясно, что сбор и обработка Данных или информации относятся к синтаксическим операциям. При сборе и обработке их компоненты выступают как *синанды* – операнды синтаксических операций. В этом и проявляется эквивалентность Данных и информации.

Операции, связанные с исследованием Данных и извлечением информации из них, с получением согласно поставленным целям по имеющимся Данным новых знаний (теорий, идей, выводов, решений и т. п.), объединим понятием *анализ Данных*, или *анализ информации*. Анализ является семантической операцией, в которой компоненты выступают как *семанды* – операторы семантических операций. Заметим, что иногда под анализом Данных понимают такие процедуры получения сверток, которые не допускают формального алгоритмического подхода [5].

При формализованной (механической) обработке (как, впрочем, при сборе и обычном, сверточном, анализе) Данных содержащийся в них объем информации не может быть увеличен. При этом происходит лишь преобразование информации к виду, более удобному (по сравнению с полученным во время или после сбора) для дальнейших операций анализа, интерпретации и применения результатов обработки. Заметим также, что анализу Данных предшествуют их сбор и обработка. Поэтому словосочетание *анализ Данных*, или *анализ информации*, часто используется в расширенном понимании, объединяющем все операции с Данными, или информацией. Мы постараемся при необходимости делать различие между понятиями *сбор*, *обработка* и *анализ*.

Следующее важное понятие – *интерпретация результатов* обработки и анализа Данных и информации. Под ним понимается истолкование, разъяснение смысла, значения, перевод результатов на язык, в термины, образы, доступные, понятные пользователю. Обратим внимание на два обстоятельства. Во-первых, на некорректность словосочетания «интерпретация сигналов или данных». Интерпретировать можно только

информацию, в частности знания, т. е. то, что содержится в сигналах, данных, является результатом их обработки, анализа, а не сами сигналы, данные, знания-1 как носители информации. Во-вторых, на то, что интерпретация относится к семантическим операциям.

Сочетание интерпретирование сигналов или данных может использоваться лишь в контексте их представления, сравнения, толкования как носителей информации, но не в контексте интерпретации результатов их обработки и анализа.

Под *применением результатов* обработки и анализа Данных и информации будем понимать действия, связанные с использованием этих результатов для решения теоретических и практических задач, с реализацией технологии, с достижением конкретных целей, в которых сбор, обработка, анализ Данных и информации и интерпретация их результатов – лишь промежуточные, вспомогательные и необходимые технологические этапы, операции.

Действия, операции, связанные с применением результатов, относятся к прагматическим, а любые используемые при этом операнды назовем *прагмандами*. К ним относятся польза, вред, нейтральность от применения результатов сбора, обработки, анализа, интерпретации Данных и информации.

Наконец, еще один важный момент, касающийся понятия *носители информации* – наличие их разноуровневости: из одних (*финальных*) мы непосредственно выявляем, добываем, извлекаем информацию, другие являются промежуточными носителями *финальных* носителей. Поясним изложенное.

Прежде всего отметим, что рассмотренные носители информации – сигналы, данные, знания-1, контент относятся к уровню *финальных*. Когда мы говорим, что сигнал (например, температура тела, пульс, артериальное давление, электрические потенциалы, отражающие состояние сердца, и т. п.) – носитель (*стартовый источник*) информации об объекте, мы в это вкладываем тот смысл, что его мгновенные текущие значения и/или значения его параметров, вид и значения его характеристик, отражающих те или иные состояния объекта, несут необходимые, полезные для диагностики текущего состояния объекта (здоровья человека) сведения. Аналогично данные (например, отсчеты сигнала) содержат семантические сведения о состоянии объекта только на момент их получения с учетом погрешностей, достоверности. Но эти данные мы записываем, представляем, храним, передаем с помощью их промежуточных носителей (символов, кодов, образов, графиков, звуков, электромагнитных колебаний и т. п.), внутри которых можно

ввести свою иерархию в процессе оперирования с ними. Носителями этих промежуточных носителей являются другие промежуточные носители – бумага, диск, флэшка, экран, ... А что служит носителем генной, наследственной информации? Понятно, что можно все их обобщенно считать и называть просто носителями информации (как это часто делается). Однако это вряд ли желательно и целесообразно делать. Почему? Когда мы рассматриваем принципы построения устройств хранения Данных, а пишем «устройств хранения информации», не делаем ли мы тем самым «медвежью услугу» читателю, неосознанно заставляя его отождествлять понятия *Данные* и *информация*, т. е. понятие *Данные* как сырье, руду, из которых получается (добывается, извлекается, ...) информация, и понятие *информация* как продукт, полезное ископаемое, которые получаются из сырья, руды? Ведь для извлечения информации из Данных необходимо выполнить определенную последовательность технологических операций (особенно если Данные закодированы, заархивированы, защищены, сознательно «искажены»), без чего информация не может быть вообще или правильно получена потребителем. Почему же тогда это не рассматривается в принципах построения *носителей информации*, отождествляемых с *носителями Данных*? Может, принципы построения именно информации, той ее части, которая нужна конкретному пользователю, будут другими, отличными или не полностью совпадающими с принципами построения носителей Данных?

Наконец, коснемся важных для алгоритмов понятий – *метод* и *алгоритм*.

Обычно под *методом* (от гр. *methodos* – путь к чему-либо) понимают правило или систему стандартных правил, подходов, способов, путей, приемов, действий, направленных на достижение цели или решение конкретной задачи. *Алгоритм* же решения какой-либо задачи понимается двояко. Во-первых, как формальное представление, точное предписание (в виде конечного набора правил), однозначно определяющее содержание и последовательность чисто механически выполняемых действий (алгоритмических операций), переводящих исходные данные задачи в искомый результат. Во-вторых, как сам процесс решения поставленной задачи в виде такой совокупности действий (операций), предпринимаемых по строго определенным правилам, которая после последовательного (пошагового) их выполнения приводит к такому ее решению, когда результат однозначно определяется исходными данными.

Таким образом, для алгоритмов характерны следующие особенности:

- *дискретность процедур* – расчлененность определяемого алгоритмом процесса решения на отдельные элементарные акты (действия, операции), возможность выполнения которых не вызывает сомнения и выполнение каждого из которых допустимо только после завершения всех операций на предыдущем этапе (шаге);
- *определенность (общепонятность, детерминированность) действий* – жесткая четко определенная последовательность элементарных операций, выполнения шагов, когда, во-первых, совокупность промежуточных результатов-операндов на любом шаге однозначно определяется операндами, имевшимися на предыдущем шаге; во-вторых, когда ни у кого не возникает возможность различно толковать путь решения задачи;
- *направленность* – если способ получения последующих операндов из предыдущих не приводит к результату, то указывается, что следует считать результатом (применения) алгоритма;
- *элементарность* – когда правило получения последующих операндов из предыдущих должно быть простым и локальным;
- *результативность (сходимость)* – нахождение искомого результата после выполнения конечного числа шагов;

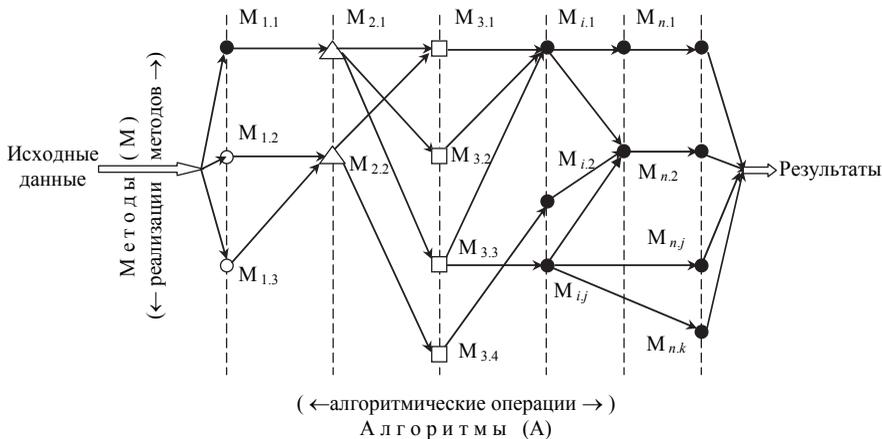


Рис. 1.4. Условное изображение связи методов и алгоритмов:

- – отдельные элементарные алгоритмические операции; \circ , \triangle , \square – набор последовательности элементарных операций, реализующих метод (подалгоритм)

- *однозначность* – единственность результата процесса при заданных исходных данных;
- *массовость* – применимость для различных исходных данных и для классов задач.

Из изложенного следует, что условно соотношение метода и алгоритма решения конкретной задачи в виде последовательности операций можно представить в виде рис. 1.4.

На рисунке $M_{i,j}$ – j -я разновидность метода, связанного с реализацией i -й алгоритмической операции. Например, разновидность метода параметрического или непараметрического оценивания, реализуемая своим подалгоритмом как составной частью всего алгоритма статистического измерения или оценивания.

Понятно, что если названия метода и алгоритма в точках ветвления совпадают, то полное наименование всего алгоритма решения задачи будет составным из n названий, отражающих каждую операцию, если, конечно, это длинное наименование не будет заменено эквивалентным коротким.

§ 1.4. ИНФОРМАТИКА И ЕЕ СТРУКТУРЫ

Центральным понятием настоящей дисциплины является термин *информатика*. Хотя его разные определения и история трансформации понятия рассмотрены в разных источниках (см. об этом, например, в [1, 3]), общепринятого толкования термин не имеет до настоящего времени, несмотря на то что он не относится к исследовательской (познавательной) модели понятия, т. е. может быть введен в качестве договорной или созидательной (прагматической) модели. Поэтому рассмотрим его в трактовке, которой в такой или близкой форме придерживается все большее число специалистов. Это определение и связанные с ним понятия, рассматриваемые в дисциплине «Введение в специальность», приведены в концентрированном виде и на рис. 1.5, 1.6, и в табл. 1.1 (более подробно о них см. в [1, 3, 4]). Здесь уместно обратить внимание на следующие моменты:

1) на триединство *информатики* как состоящей из фундаментальной и прикладной научных дисциплин и соответствующей области практической деятельности человечества, имеющих одинаковые объемы, но разные предметы исследования (см. рис. 1.5);

ОБЪЕКТЫ ИНФОРМАТИКИ

• собственно информация, информационные модели, процессы, технологии, средства, структуры, ресурсы, услуги, отношения с другими информационными и «неинформационными» объектами (включая субъекты) Вселенной

ПРЕДМЕТЫ (ВНИМАНИЯ) ИНФОРМАТИКИ

Фундаментальная наука • сущность, состав, свойства, правила и закономерности строения, функционирования, поведения, развития, ... объектов информатики; познание, изучение, описание, исследование происхождения, объяснение и предсказание их строения, свойств и особенностей

Прикладная наука • теоретические основы создания и эффективного применения объектов информатики

Область практической деятельности • разработка, изготовление, выпуск и разностороннее эффективное использование объектов информатики, а также сбор, передача, обработка, анализ, интерпретация, применение информации и ее носителей в нужном объеме заданного качества в требуемые сроки при минимальной себестоимости

МЕТОДЫ ИНФОРМАТИКИ

Общенаучные • анализ, синтез; аналогия, сравнение; индукция, дедукция, традиция, абдукция; наблюдение, экспериментирование; формализация, абстрагирование; аксиоматические, гипотетические, эвристические, аналитические; естественно-научные, эмпирические, аксиоматические (математические); инженерные; изобретательские; моделирующие; информационные и т.п.

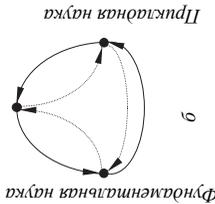
Специфичные • алгоритмизация, моделирование, программирование (планирование); системные; имитационные; виртуальная реальность; поиска, анализа, выбора вариантов, решений; обнаружения и исправления ошибок; распознавания образов; постановки и решения задач, доказательств, творчества, ...

ПОНЯТИЕ ИСТИНЫ (ИСТИННОСТИ)

Истинным (верным) считается то, что получено с помощью корректных логических выводов и доказательств, подтверждено корректным физическим и/или машинным экспериментом, а также принципиально другими вариантами решений

Рис. 1.5. Объекты, предметы, методы исследования и понятие истины в информатике

а
Область практической
деятельности



Аппаратные (технические)
Математические (модельные + алгоритмические)
Программные
Информационные
Лингвистические
Технологические
Метрологические
Логистические
Правовые
Организационно-экономические
Эргономические

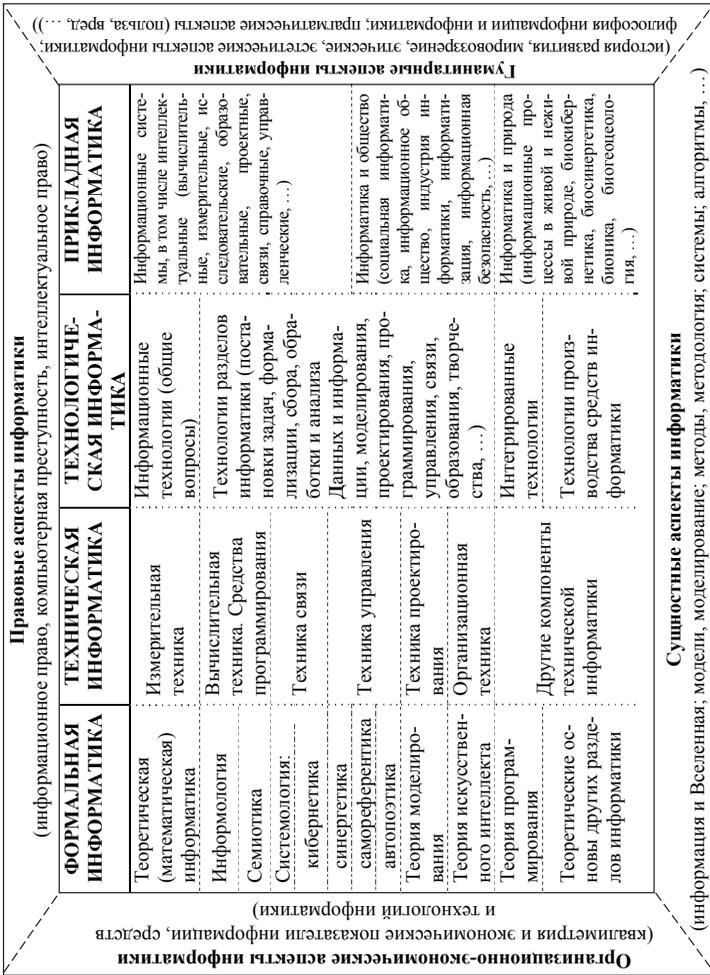


Рис. 1.6. Пояснение структуры информатики: а – схемное изображение триединства информатики; б – средства информатики – подсистемы информационных систем различного назначения; в – состав информатики как научной дисциплины

«Рабочие» определения базовых терминов информатики

Термин	Определение термина
Информатика	Научная дисциплина, охватывающая совокупность фундаментальных и прикладных направлений, и область практической деятельности, занимающиеся исследованием сущности, свойств, закономерностей информации и решением проблем создания, внедрения и исследования технологий, процессов, средств сбора, передачи, обработки, изучения, анализа, интерпретации и применения Данных и информации, а также решением разнообразных задач использования Данных и информации, информационных технологий, средств, ресурсов и структур в естественных и искусственных материальных объектах, живых и неживых организмах и сообществах
Информационные ресурсы	Различные формализованные знания (теории, идеи, изобретения), данные (в том числе документы), технологии и средства их сбора, обработки, анализа, интерпретации и применения, а также обмена между источниками и потребителями информации
Информационная технология	1. Совокупность научных дисциплин, занимающихся изучением, созданием и применением методов, способов, действий, процессов, средств, правил, навыков, используемых для получения новой информации (сведений, знаний), сбора, обработки, анализа, интерпретации, выделения и применения Данных и информации с целью удовлетворения информационных потребностей национального хозяйства и общества в требуемом объеме и заданного качества 2. Совокупность самих этих методов, способов, действий и т. д.
Информационный процесс	Последовательность действий по сбору, передаче, обработке, анализу, выделению и использованию с различной целью информации (и/или ее носителей) в ходе функционирования и взаимодействия материальных объектов
Информационный технологический процесс	Компонент информационной технологии как практического инструмента рецептурной деятельности, часть производственного процесса, состоящая из последовательности согласованных технологических операций, связанных со сбором и с обработкой Данных как носителей информации, выделением из них необходимых сведений, новостей, знаний, их накоплением, исследованием, анализом, интерпретацией и применением

2) на соборность (агрегированность) информатики (см. рис. 1.6), наличие разных структур информатики как моделей ее состава, т. е. совокупности элементов, частей, разделов, образующих ее, а также связей между составляющими ее элементами. Структура объекта есть целевой образ, его куомодная модель, отвечающая на вопрос «Как устроен, из чего состоит объект?». Поэтому степень детализации структуры и ее элементов, включение той или иной части объекта в его модельное представление в виде структуры, отражение в структуре соответствующих связей, их назначения, особенностей, ... определяются целью построения структуры. Примеры разных структур информатики представлены на рис. 1.6. В частности, на рис. 1.6, в представлена одна из структур формальной информатики, которую мы условно в пособии назвали теоретической (см. введение).

§ 1.5. СТРУКТУРА ТЕОРЕТИЧЕСКОЙ (ФОРМАЛЬНОЙ) ИНФОРМАТИКИ

Один из вариантов более детального перечня разделов (дисциплин) формальной (теоретической) информатики представлен ниже. Заметим, что в указанный далее перечень разделов входят как собственно развиваемые в информатике, так и те, что создавались в других дисциплинах (прежде всего в математике) задолго до появления информатики как научной дисциплины, составляют ее теоретический каркас, аппарат формализации. Ситуация здесь аналогична той, что имеет место в теоретической физике, математической физике, химии, биологии, геологии.

Выделим следующие разделы теоретической (математической, формальной) информатики, составляющие ее теоретические начала, формализованные основы.

1. Базовые математические дисциплины, используемые в информатике

- 1.1. Дискретная и вычислительная математика и геометрия.
- 1.2. Компьютерная алгебра и теория графов.
- 1.3. Математическая логика и исчисления высказываний.
- 1.4. Математические методы системного анализа.
- 1.5. Теория алгоритмов и формальных грамматик.
- 1.6. Теория автоматов.
- 1.7. Математическая и структурная лингвистика.
- 1.8. Теория сигналов.

1.9. Теория экспериментов, компьютерного анализа Данных, их идентификации, имитации прогнозирования и интерпретации результатов.

1.10. Теории оптимизации, исследования операций, игр, систем массового обслуживания и т. п.

1.11. Теория Данных, их сбора, хранения, кодирования и передачи.

1.12. Нетрадиционные методы формального описания систем и Данных (фрактальные, нечеткие, экспертные, интервальные, нелинейной динамики и динамического хаоса, квантовые и пр.).

2. Системология (теоретические основы нижеперечисленных разделов)

2.1. Общая теория систем.

2.2. Теория выбора и решения системных проблем.

2.3. Теория управления. Кибернетика.

2.4. Имитационное моделирование систем.

2.5. Компьютерные игры.

2.6. Техническая диагностика и тестирование систем.

2.7. Теории обучения, классификации, распознавания образов.

2.8. Надежность аппаратного и программного обеспечения.

2.9. Синергетика, самореферентика, автопоэтика.

3. Теоретические основы искусственного интеллекта

3.1. Когнитология и когнитивная психология.

3.2. Инженерия знаний. Базы знаний и экспертные системы.

3.3. Управление знаниями.

3.4. Компьютерная лингвистика и психолингвистика.

3.5. Машинный перевод, машинное моделирование рассуждений, доказательств и проверки гипотез.

3.6. Концептуальное проектирование искусственных объектов. Синтез текстов, речи, литературных и музыкальных произведений.

3.7. Теории решения изобретательских задач (ТРИЗ), развития технических систем (ТРТС), развития творческой личности (ТРТЛ), сильного мышления (ТСМ).

3.8. Теория интерпретации и выводов.

3.9. Теория моделей и моделирующих систем.

3.10. Нейрокомпьютеры и нейрокомпьютерные сети и системы.

3.11. Автоматы и автоматные сети.

3.12. Робототехника и интеллектуальные агенты.

3.13. Интеллектуальный и когнитивный анализы Данных.

3.14. Другие разделы искусственного интеллекта.

4. Теория программирования (теоретические основы нижеперечисленных разделов)

4.1. Теоретические основы программирования.

4.2. Классификация средств программирования и программных систем.

4.3. Архитектура системного программного обеспечения.

4.4. Операционные системы.

4.5. Языки программирования.

4.6. Трансляторы, компиляторы, редакторы.

4.7. Специальные виды программирования.

4.8. Прикладное программное обеспечение.

4.9. Программирование и структуры данных. Базы и банки данных. Электронные таблицы.

4.10. Параллельное, системное, сетевое программирование.

4.11. Интегрированные программные системы.

4.12. Технологии программирования.

5. Информология

5.1. Предмет, объект и методы информологии.

5.2. Информация и эволюция.

5.3. Информация, образование и формирование личности.

5.4. Другие разделы информологии.

ЗАКЛЮЧЕНИЕ

Итак, в первой главе приведены и пояснены первичные понятия информатики. Прежде всего рассмотрены понятия *модель* и *термин*. Обращено внимание на недопустимость отрывания модели от отображаемого ею объекта, цели отображения, субъекта, оперирующего с моделью, и среды, в которой осуществляется это отображение. Рассмотрены базовые свойства модели. Дано пояснение науки как области человеческой деятельности и ее составной дисциплины – информатики, их объектов, предметов, методов и понятия истинности результатов. Особый упор сделан на базовые понятия, связанные с информацией и ее носителями.

Обращено внимание на возможность взаимного контекстного использования (при категорической недопустимости отождествления понятий!) термина *информация* и терминов, касающихся носителей информации: *сигналы*, *данные*, *знания-1*, *контент*, *модель*, *кодированное представление* и т. п. Для дополнительного акцентирования внимания на недопустимость отождествления понятий, стоящих за терминами, касающимися информации и ее составных частей, с одной стороны, и

их носителями, с другой, введены следующие новые термины: *надные, анзня, Данные, синанды, семанды, прагманды*. Первые из них являются конкретизирующими (анзня) и обобщающими (надные, Данные) для носителей информации как альтернатива самой информации, которую они несут. Вторые (синанды, семанды, прагманды) – отражают функции, назначение Данных и содержащейся в них информации как операндов информационных технологических процессов. А именно, являются ли Данные и информация объектами несмысловых синтаксических операций (сбора, накопления, передачи, ...), семантических (связанных с извлечением информации, преобразованием ее смыслового содержания, ...) или практического применения, когда важны вопросы ценности, пользы, затрат на хранение, получение, передачу Данных или информации. Заметим, что именно в первых (синтаксических) и третьих (прагматических) операциях допустимо эквивалентное дизъюнктивное применение терминов *Данные* или *информация*. Например, сбор Данных об объекте означает синтаксический сбор информации об объекте через эти Данные, т. е. сбор информации, которую несут Данные об объекте. Именно в этом контексте чаще всего можно считать эквивалентными словосочетания средства сбора, хранения, передачи Данных и средства сбора, хранения, передачи информации.

ВОПРОСЫ ДЛЯ САМОПОДГОТОВКИ

1. Согласны ли вы с наименованием образовательного направления «Информатика и вычислительная техника»? Обоснуйте свой ответ.
2. Что такое структура объекта? Что такое структура информатики? Может ли быть несколько структур исследуемого объекта? Почему?
3. Поясните термин *теоретическая информатика*. Как он соотносится с терминами *фундаментальные основы информатики* и *формальная информатика*?
4. Что такое модель? Приведите поясняющие примеры. Может ли быть модель модели или бывают только модели оригинала? Может ли оригинал быть моделью? Если да, то чего?
5. Поясните словосочетание *термины как модели определяемых ими понятий*.
6. Поясните термины теории моделирования: *объект, субъект, среда, цель*.
7. Перечислите и охарактеризуйте основные свойства модели. Приведите примеры.

8. Что понимается под термином *наука*? Укажите необходимые условия, при которых область человеческой деятельности можно отнести к научной.

9. Что такое состав науки? Перечислите и поясните критерии научности познания.

10. Что представляют собой результаты науки? В чем их сходство и отличие от результатов других видов человеческой деятельности?

11. На какие виды делятся научные знания?

12. Приведите и поясните типы научных дисциплин.

13. Дайте определения терминам *сигналы, данные, знания, контент, надные, анзния, Данные*. В чем сходство и различие этих понятий?

14. Что такое *информация, новости, протознания, информационный мусор, релевантная информация*?

15. Чем различаются понятия *сбор, обработка, анализ, исследование, применение Данных*?

16. Что такое *интерпретация данных, интерпретация результатов работы с Данными* (сбора, обработки, анализа, ...)?

17. Что такое *технологический процесс? Информационный процесс? Информационный технологический процесс?*

18. Что такое *информационные ресурсы*?

19. Что такое *информатика*? В чем суть ее триединства?

20. Перечислите и поясните объекты информатики.

21. Перечислите и поясните предметы информатики как фундаментальной и прикладной наук и области практической деятельности человечества.

22. Перечислите и поясните методы информатики.

23. Как трактуется понятие истины в информатике?

24. Приведите примеры разных структур информатики. Почему их несколько?

25. Укажите основные средства информатики.

26. Приведите пример структуры теоретической информатики.

Глава вторая

МОДЕЛИРОВАНИЕ КАК МЕТОД ИССЛЕДОВАНИЯ ОБЪЕКТОВ

§ 2.1. МОДЕЛИРОВАНИЕ ОБЪЕКТОВ

2.1.1. Определение моделирования

Прежде чем переходить к описанию моделирования объектов, обратим внимание на оговоренную в п. 1.1 условность трактовки понятия *исследование объекта*. Под исследованием объекта имеется в виду не только его изучение, но и его описание, проектирование, управление и другие сходные действия с объектом.

Как следует из изложенного в первой главе, во взаимодействии с объектом создание его модели – не самоцель. Это лишь первый этап его моделирования.

Моделирование объекта есть метод его исследования, основанный:

- а) на замене исследуемого объекта-оригинала его моделью;
- б) на работе с моделью вместо объекта, т. е. исследовании объекта по (на) его модели (е);
- в) переносе полученных по (на) модели (е) результатов на объект.

Обратим внимание на трехстадийность моделирования и понимания под ним именно совокупности всех трех стадий, а не какой-то одной из них. При этом, как указывалось ранее, важно помнить о целевой направленности моделирования, а также о том, что оно выполняется в определенной среде с учетом условий функционирования объекта и реализации каждой стадии моделирования. В частности, необходимо иметь в виду, как, в каких условиях, с какой достоверностью и информативностью (см. далее) получены Данные об объекте, используемые для построения модели и проверки ее адекватности (см. далее), какие задачи будут решаться на основе результатов моделирования с учетом

многоуровневости исследования объекта. Пример многоуровневости исследования объекта и, как следствие, согласования целей получения результатов на выходе каждого уровня приведен на рис. 2.1 [5] и рис. 2.2 [6, 7].



Рис. 2.1. Условное изображение разных уровней моделирования объекта

Прокомментируем приведенные определения моделирования.

Во-первых, обратим внимание на двойственность выделения трехстадийности моделирования. Она должна иметь место (т. е. мы должны иметь ее в виду) как на каждом этапе построения и использования

модели, т. е. относиться к стадиям *a–в*, так и по отношению ко всему процессу моделирования (см. рис. 2.1). Первое означает, что все три стадии должны выполняться (учитываться) при построении модели, т. е. на стадии замены объекта моделью, исследования объекта по модели, и при возвращении к объекту для проверки верности результатов построения модели объекта и полученных по модели результатов решения прикладной задачи, их правильности, пригодности именно для данного исследуемого объекта. Если это иметь в виду, то можно было бы моделирование в целом определить как метод исследования, основанный на замене объекта его моделью. Именно в такой трактовке чаще всего и дается определение моделирования.

Во-вторых, учитывая опыт изучения студентами данного понятия моделирования, считаем необходимым включить в дефиницию трехстадийности моделирования как метода исследования. Иначе у студентов иногда складывается представление о моделировании как только об одной какой-то его стадии, например только построении модели либо замене объекта моделью и т. п., как самоцели, а не отдельного этапа (стадии) моделирования.

Наличие многостадийности подчеркивает рис. 2.1, а также рис. 2.2–2.4.

Все незнакомые термины рисунков приведены в глоссарии (он будет дан во второй части пособия; см. также [1–5]).

2.1.2. Требования к модели

Хотя модель есть целевой образ объекта-оригинала, не каждая модель может быть пригодной для решения конкретной задачи. С одной стороны, это следствие конкретной модели, в частности того, как проявляется свойство ее финитности при решении задачи, с другой стороны, помимо цели, назначения модели (см. рис. 1.1, [4, 8]) необходимо иметь в виду функцию, которую будет выполнять модель при достижении этой цели [8].

Это такие функции, как описательная, измерительная, интерпретаторская, предсказательная, критериальная [8].

Для того чтобы модель была пригодна для решения конкретной задачи, при ее построении и/или выборе необходимо проверить модель на удовлетворение ею совокупности требований, определенных этой задачей. Это следующие требования [8].

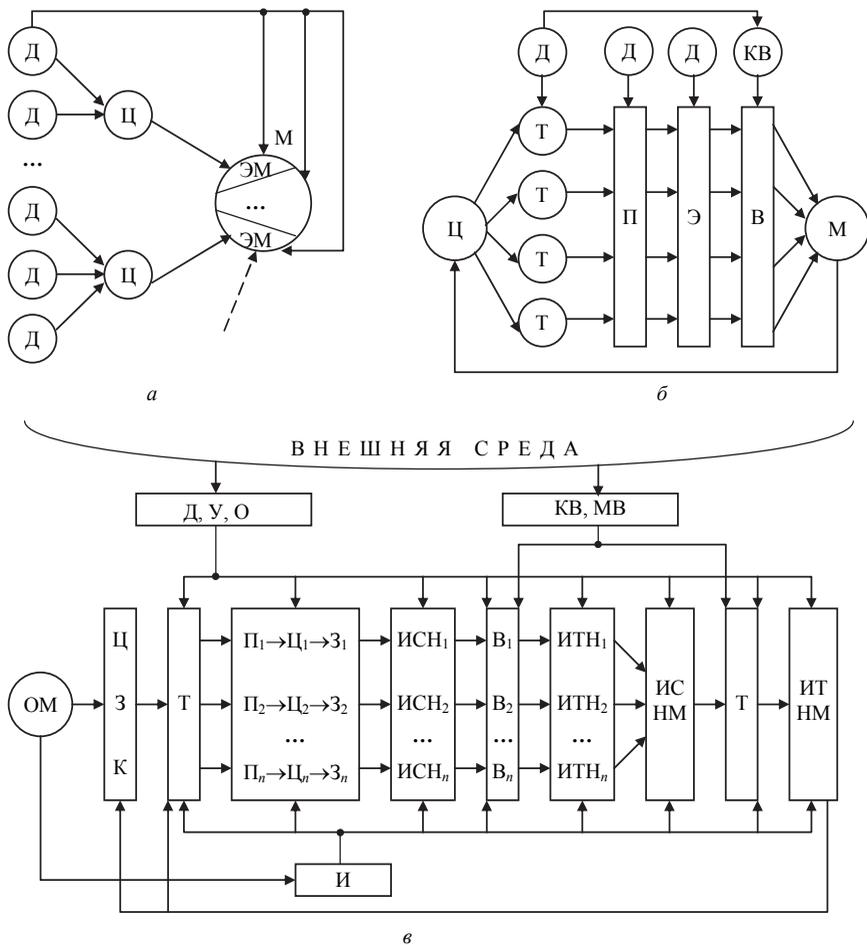


Рис. 2.2. Процесс синтеза модели на основе классического (а), системного (б) подходов и его реализации в рамках вариативного моделирования (е):

V_i, B – выбор моделей i -й подсистемы, объекта; D – данные; I – информация; $ИСН_i, ИТН_i$ – исходный и итоговый наборы моделей i -й подсистемы; $ИСНМ, ИТНМ$ – исходный и итоговый наборы моделей объекта; K – критерии достижения цели; $КВ, МВ$ – критерии и методы выбора; O – ограничения; $ОМ$ – объект моделирования; $Ц, З$ – глобальные цель и задачи моделирования; $П_i$ – i -я подсистема; T – требования; $У$ – условия; $Ц_i, З_i$ – цель и задачи моделирования i -й подсистемы; $Э$ – элемент; $ЭМ$ – элемент (компонента) модели [6, 7]



Рис. 2.3. Укрупненная схема технологического процесса моделирования физического объекта: стадия идентификации



Рис. 2.4. Стадия исследования объекта по модели: работа с моделью и перенос полученных по ней результатов на объект

1. Требование адекватности (соответствия чему-либо, кому-либо в заданном смысле)

1.1. **Целевая адекватность** (требование соответствия цели)

Модель должна быть согласована с целью ее введения (изучение, анализ и осмысление объекта, обобщение, синтез, управление, включая прогнозирование, измерение и постановка эксперимента, обучение и тренаж, имитация и пр.).

1.2. **Физическая адекватность** (требование интерпретируемости)

Модель и ее параметры должны иметь четкую физическую интерпретацию в приложении к описываемому ею материальному объекту, приводить к частным случаям, соответствующим частным состояниям и значениям параметров объекта, служить связующим звеном между теоретическими знаниями об объекте и экспериментальными данными.

1.3. **Адекватность инженерного соответствия** Данным и условиям их получения (информационная, формальная)

Модель должна согласовываться в смысле принятых критериев с эмпирическими данными, получаемыми при натурном эксперименте с объектом, и условиями их получения. Требования соответствия формальным показателям.

1.4. **Ресурсная адекватность**

Модель должна быть согласована с имеющимися у ее создателя, исследователя или пользователя ресурсами, в частности с затратами на ее получение и работу с ней, а также с пользой (вредом) от моделирования объекта именно с этой моделью при данных ресурсах.

2. Требование идентифицируемости

Должна существовать некоторая комбинация априорной и апостериорной информации, позволяющая отличить вид и параметры модели от любой другой модели, записанной, быть может, в той же самой форме.

3. Требование точности

Отличие (в принятой метрике) воспроизводимых (в частности предсказываемых) с помощью модели значений параметров объекта от истинных значений должно быть как можно меньше и не должно превышать допустимых пределов при оговоренных условиях «жизни» (функционирования) объекта.

4. Требование экономичности (простоты)

Модель должна быть настолько простой, насколько это возможно при заданных требованиях к решению практической задачи, т. е. иметь как можно меньшее число параметров, быть простой при идентифика-

ции и в применении. Затраты (денежные, ресурсные) на получение, проверку и применение модели должны быть как можно меньше.

5. Требование устойчивости (корректности)

Модель должна быть устойчивой к поведению объекта, внешним условиям, давать стабильные результаты решения конкретной задачи. С одной стороны, модель должна отслеживать изменения параметров объекта, но быть инвариантной по отношению к варьированию параметров объекта в пределах погрешностей их измерения. С другой стороны, незначительные погрешности в определении параметров модели, незначительное варьирование модели не должны приводить к большим погрешностям результатов решения практических задач с использованием модели.

В связи с рассмотрением требования устойчивости модели рассмотрим два важных понятия.

Первое понятие связано с **корректностью**¹ постановки задачи построения математической модели объекта по исходным данным χ , U (где χ – количественные значения, U – условия) и нахождением затем по этой модели искомого решения Z конкретной задачи. Если Z и χ связаны функциональной зависимостью $Z = f(\chi)$ при заданных U , то задача называется *корректной*, или *корректно поставленной*, если выполняются следующие три **условия корректности** (по Ж.С. Адамару, Франция, 1923 г.): существования, единственности и устойчивости ее решения.

Условие существования означает, что задача имеет решение при любых допустимых исходных данных. Иными словами, среди исходных данных нет противоречащих друг другу значений χ и условий U , в рамках которых формулируется (ставится) формализованно задача и которые исключили бы возможность ее решения.

Условие единственности означает, что каждым исходным данным соответствует только одно решение (однозначность задачи), т. е. исходных данных достаточно для однозначной определенности решений задачи.

Условия существования и единственности называют **условиями определенности задачи**.

Условие устойчивости означает, что значение меры отклонения решений $Z_1 = f(\chi_1, U_1)$ и $Z_2 = f(\chi_2, U_2)$ не превышает допустимой

¹ От лат. correctus – исправленный.

приемлемой приложениями величины для всех критически допустимых значений меры отклонения исходных данных (χ_1, U_1) от (χ_2, U_2) . Это условие **физической реализуемости** (детерминированности) *задачи*. Иными словами, как угодно малые возмущения исходных данных не должны вызывать больших отклонений в решениях.

Итак, корректными называются классы математических задач, которые удовлетворяют этим трем условиям определенности их решений. Задачи, не удовлетворяющие хотя бы одному условию корректности, называются *некорректными*, или *некорректно поставленными*. Заметим, что сам Ж.С. Адамар [1] корректность задач связывал с решением краевых задач для уравнений с частными производными.

Второе – понятие **робастности**¹. Обычно под робастностью процедур решения задачи измерения (оценивания, проверки гипотез, ...) понимается устойчивость результата решения задачи к небольшим отклонениям от номинальных условий, в рамках которых она решалась. Идейную основу робастности составляют принципы релевантности, непрерывности и уже упоминавшейся устойчивости. **Принцип релевантности** означает, что «все относящиеся к делу (к постановке и решению задачи, релевантные) Данные должны быть использованы при этом, и, наоборот, все не имеющие к нему отношения (иррелевантные) и ложные данные должны быть исключены из рассмотрения». **Принцип непрерывности** звучит так: «малые возмущения Данных должны вызывать соответствующие малые изменения в вытекающих из них выводах». Наконец, **принцип устойчивости** следующий: «малая ошибка в математической модели не должна приводить к существенной ошибке окончательных выводов» [5]. Более подробно робастность рассмотрим в § 4.3 (математическая статистика). Здесь лишь отметим, что робастность связана прежде всего с отклонениями от предположений, касающихся вида используемой для решения задачи модели Данных, а именно, с одной стороны, какие отклонения оказывают решающее влияние на конечный вывод, с другой, – нарушения каких допущений являются относительно безвредными. Усилия *специалистов-робастников* (строителей моделей, методов, алгоритмов) и *модельмастеров* (специалистов по исследованию объектов по моделям) направлены на разработку таких робастных моделей, методов, алгоритмов, которые будут наилучшими в каком-то смысле

¹ От *англ.* robust – крепкий, устойчивый.

не на «узких» условиях, нарушение которых способно привести к резкой потере преимуществ таких моделей (методов, алгоритмов), а на классе близких условий. Такой подход может привести к некоторым потерям, если на самом деле Данные соответствуют «узкой» модели, но к существенному выигрышу, если модель Данных отличается от приписываемой ей «узкой» модели. Иными словами, робастность ориентирована на преодоление неадекватности модели Данным и построение моделей, не чувствительных или мало чувствительных к отклонениям от предложений о принятой модели.

6. Требование универсальности

Модель должна допускать возможные обобщения, быть применимой к многочисленной группе однотипных объектов, функционирующих в одном или многих режимах.

7. Требование многоцелевости, многофункциональности

Модель должна позволять решать несколько задач (см. требование адекватности, а также [3, 4] и далее табл. 2.2).

8. Специфические требования

Эти требования определяются спецификой приложений. Например, в теории управления они касаются свойств модели отражать управляемость и наблюдаемость системы.

Укажем еще два важных требования, которые вместе с требованием адекватности иногда называют основными *законами математического моделирования* природных объектов. Это требования (законы) **ориентации на Данные и независимости** параметров моделей от искоемых переменных. Закон ориентации: модель должна соответствовать уровню достоверности Данных и реальности ее достижения в конкретных условиях. Если это не так, модель должна быть изменена, упрощена. Закон независимости очевиден и означает, что искоемые переменные не должны входить в параметры модели.

2.1.3. Виды моделей

Еще один важный момент в моделировании – выявление вида (класса, типа) моделей. Рассмотрим некоторые важные виды моделей (см. [4, 8]).

Прежде всего обратим внимание на деление моделей по их направленности (табл. 2.1, см. также [4, 8]). Другие варианты деления моделей по видам приведены в табл. 2.2 и 2.3 (см. [4, 8]).

Т а б л и ц а 2.1

Виды моделей по их целевой направленности

Направленность (применение) модели	Наименование модели	Содержание направленности
1. Отношение объект–модель	1.1. Исследовательская (познавательная)	Модель приближается к объекту, настраивается на него, подгоняется к нему
	1.2. Созидательная (прагматическая)	Объект подгоняется к модели, приближается к ней
	1.3. Договорная (конкордовая, принятая по соглашению)	Соответствие объекта и модели постулируется, принимается по соглашению между субъектами
2. Предметная	2.1. Куáлизная (терминологическая)	Отвечает на вопросы: «Что за объект?», «Какой объект?», «Как его отличать от других?»
	2.2. Куомóдная	Отвечает на вопросы: «Как устроен и функционирует (живет, работает) объект?», «Каков механизм работы?»
	2.3. Каузáльная	Отвечает на вопрос «Почему так устроен и функционирует объект?»
	2.4. Телóсная	Отвечает на вопрос «Зачем, для чего существует, именно так устроен и функционирует объект?»
	2.5. Эволюционная (историческая)	Отвечает на вопрос «Как зародились и изменялись объект, его устройство, функционирование?»
3. Временнáя	3.1. Гéнезисная	Отражает прошлые состояния и функционирование объекта
	3.2. Диагностическая (мониторинговая)	Отражает настоящее (текущее) состояние и функционирование объекта
	3.3. Проектная, прогностическая	Отражает будущие состояния и функционирование объекта
4. Методологическая	4.1. Элементная	Ориентируется на исследование элементов объекта
	4.2. Системная	Направлена на познание объекта как целого, его частей, а также объекта как части более сложной системы (надсистемы)

Т а б л и ц а 2.2

**Фрагмент морфологической таблицы классов моделей
и методов моделирования**

Главные определяющие факторы	Отличительный (сравнительный) признак	Возможные реализации признака (наименование моделей, методов моделирования)		
Назначение модели	Вид отражаемого объекта	Существующий (исследовательские, познавательные)	Желаемый (созидательные, прагматические)	
	Цели моделирования	Теоретические, гносеологические (познавательные), практические (прагматические), коммуникабельные, логостические и др.		
	Выполняемые функции	Передаточная, измерительная, описательная, интерпретаторская, объяснительная, критериальная, ...		
Физический носитель	Вид физического носителя модели	Материя: вещество, энергия (материальные)	Элементы формального аппарата: (идеальные) описания, представления объекта моделью, разные символы (буквы, цифры, линии, цвета, коды, ...)	
	Способ установления подобия для материальных моделей	Через физическое взаимодействие с оригиналом (прямого подобия)	Через аналогию, т. е. путем обнаружения объективно существующего сходства (косвенного подобия)	Через соглашения (условного подобия)
Возможность физической реализации	Реализованные (физические, реальные)	Допускающие физическую реализацию (виртуальные)	Нереализуемые физически (абстрактные)	

Окончание табл. 2.2

Главные определяющие факторы	Отличительный (сравнительный) признак	Возможные реализации признака (наименование моделей, методов моделирования)		
Отношение к априорным, апостериорным и промежуточным данным, знаниям, результатам	Назначение эксперимента при построении модели	Поверочное (априорные)	Исходное (апостериорные)	
	Сходство физических носителей модели и оригинала	Одной природы (идентичные)		Разной природы (аналоговые)
		Один и тот же носитель (натурные)	Однородные по природе носители (физические, биологические, химические, геометрические, прочие)	
	Характер исходных данных	Единственный источник знаний – экспериментальные данные (эмпирические)	Единственный источник знаний – разум (рационалистические)	
	Использование модели в эксперименте	Пассивное (отражательные)	Активное (имитационные)	
	Информационная направленность	Отражающие форму (синтактические)	Отражающие содержание (семантические)	Отражающие пользу (прагматические)
Вид результата	Характер результата моделирования	Произвольный из определенного множества (неоптимальные)	Наилучший в определенных условиях (оптимизационные)	
			Наилучшие в конкретных условиях (оптимальные)	Наилучшие (устойчивые) на классе условий (робастные)

**Фрагмент морфологической таблицы математических моделей
(в частности аналоговых)**

Главные определяющие факторы	Отличительный (сравнительный) признак	Возможные реализации признака (наименование моделей, методов моделирования)			
Метод, положенный в основу модели	Средство, используемое для отражения оригинала	Визуальные образы (образные)		Условные знаки, символы (знаковые)	
	Уровень описания объекта	Нутро (элементы) объекта (микромодели)	Объект как целое (макромодели)	Часть более общего целого (мегамоделей)	
	Способ возникновения модели	Эвристические приемы (логические)			Формальные приемы (аналитические)
		Наблюдение объекта (феноменологические)	Построения построения (логические)		
	Предмет описания в оригинале	Строение оригинала (структурные)		Поведение оригинала (функциональные)	
	Поведение объекта во времени	Статическое		Динамическое	
	Вид аргумента, переменных	Непрерывный (е)		Дискретный(е)	
	Число выходных переменных	Одна переменная (скалярные)		Более одной переменной (векторные)	
	Вид исходных, промежуточных и итоговых результатов, представление операндов (количество значений)	Одно число (точные)		Интервал значений (интервальные)	

Главные определяющие факторы	Отличительный (сравнительный) признак	Возможные реализации признака (наименование моделей, методов моделирования)		
	Число разрядов операндов	Бесконечное (непрерывнозначные)	Конечное (цифровые)	Смешанное (гибридные)
Тип математического аппарата	Вид модели по отношению к переменным и параметрам	Линейная(е)		Нелинейная(е)
	Геометрическая (топологическая) размерность	Цельно-размерные		Дробно-размерные (фрактальные)
	Учет повторяемости результатов при многократных экспериментах	Наличие повторяемости (детерминированные, детерминистские)		Отсутствие повторяемости или ее невозможность (индетерминированные, недетерминистские)
	Наличие некоторой регулярности в экспериментах	Регулярность имеет место (регулярные)		Регулярности нет или она бессмысленна (иррегулярные)
	Использование закономерностей регулярного поведения	Учитывающие объективные закономерности (стохастические: в частности априорные – вероятностные, апостериорные – статистические; хаотические, в частности иррегулярные детерминированные)		Учитывающие субъективное восприятие закономерностей (астохастические: регулярные – нечеткие, размытые; иррегулярные – экспертные)

§ 2.2. НЕОБХОДИМЫЕ ПОНЯТИЯ СЕМИОТИКИ

Одной из важных дисциплин формальной информатики является семиотика. Согласно ФГОС она не обязательна для подробного изучения ее студентами бакалавриата. Поэтому рассмотрим только те базовые термины, которые широко используются в разных дисциплинах

учебного плана УГС «Информатика и вычислительные технологии» и будут применяться дальше в тексте пособия.

Семиотика – наука о знаках, свойствах знаков, знаковых системах и языках.

Синтактика – раздел семиотики, изучающий структуру сочетаний знаков и правил их образования и преобразования безотносительно к их значениям и функциям (план выражения).

Семантика – раздел, изучающий знаковые системы как средства выражения смысла, правила интерпретации знаков и составленных из них выражений (план значения).

Прагматика – раздел, изучающий отношения между знаковыми системами и теми, кто ими пользуется, отношения субъектов, воспринимающих и использующих какую-либо знаковую систему, к самой знаковой системе (план использования).

В ряде приложений (например, при рассмотрении структуры языка как символической системы) раздел синтактика делится на сигматику и синтаксис (см. рис. 2.4 в [4]).

Сигматика – это раздел семиотики, изучающий словарные аспекты языка, отношения между знаками, словами и объектами, отражаемыми ими.

Синтаксис – раздел семиотики, изучающий коммутативные аспекты языка, отношения между разными знаками и словами, порядок их следования, связи между ними.

К вопросу связи понятий семиотики с уже введенными и вводимыми далее понятиями мы обратимся позже, во второй части пособия.

§ 2.3. ПОНЯТИЕ О СИСТЕМОЛОГИИ

2.3.1. Системология как наука и область практической деятельности.

Системные дисциплины. Свойства систем

Одним из важнейших разделов теоретической информатики является системология – наука о системах, или общая теория систем, а также составляющие ее дисциплины. Ее состав как раздел науки и области практической деятельности компактно представлен на рис. 2.5 [1].

Центральным понятием системологии является термин *система*. Памятуя о том, что термин – это модель определяемого им понятия, и учитывая многообразие определений и интерпретаций этого термина

разными авторами, справочниками, словарями, приведем то, которое ориентировано на материал настоящего пособия.

Система – это модель объекта как **цельной целой** совокупности выделенных из объекта его взаимосвязанных элементов, объединенных для реализации (обеспечивающих достижение) общей цели¹, обособленных от окружающей среды, взаимодействующих с ней как *целостное целое* и обуславливающих при этом проявление основных рассматриваемых важных для исследования («системных») свойств объекта.

Прокомментируем это определение системы (см. [5, 8]). Прежде всего поясним используемые в нем выделенные светлым курсивом понятия.

Целое означает единое полное, внутренне сохраняемое без каких-либо изъятий неделимое на независимые части без потери его существенных свойств и функций.

Цельное – состоящее из одного куска, не составное, но модельно делимое на части при внутреннем рассмотрении.

Целостное – неделимое при наружном рассмотрении, внутренне единое, автономное, независимое от среды, не исчерпываемое конечным набором своих представлений, не предполагающее взаимодействия элементов при наружном рассмотрении.

Цель жизни объекта (системы как его модели) – его определенное желаемое состояние или желаемые значения его параметров, характеристик, свойств, особенностей – модель желаемого будущего.

Далее обратим внимание на трактовку понятия *система* как модель некоторого объекта, который мы называем системой. Модельность системы проявляется в следующем. Во-первых, в относительности этого понятия: нечто (исследуемый объект) может быть системой относительно данного системообразующего (в частности, приводящего к эмерджентности) отношения или свойства его, при данном целевом рассмотрении, но не быть системой относительно другого отношения или свойства, другой сути. Во-вторых, применяя понятие *система* к какому-то объекту, мы выражаем не только сущность объекта, но и наше отношение к сущности, строению, свойствам и поведению объекта. В-третьих, понятие *система* (поскольку это модель) может относиться как к объектам реального мира, так и к их моделям. Наконец, в-четвертых. Не следует смешивать рассматриваемое модельное понятие

¹ Жизни, функционирования объекта с точки зрения цели моделирования.

СИСТЕМОЛОГИЯ, ТЕОРИЯ СИСТЕМ, наука о системах
ОБЪЕКТЫ СИСТЕМОЛОГИИ: понятие системы, виды, структуры
и свойства систем

ПРЕДМЕТЫ ВНИМАНИЯ СИСТЕМОЛОГИИ¹ == сущность, принципы,
закономерности функционирования, существования, поведения, эволюции
систем; описание, объяснение и предсказание их свойств и особенностей
(их познание, изучение, синтез)

КИБЕРНЕТИКА – наука об управлении, связи и передаче информации
в живых (биологических и социальных) и неживых объектах
(системах) + Д²

ОБЪЕКТЫ КИБЕРНЕТИКИ: устойчивые управляемые системы; общие
принципы, цели и закономерности управления, связи и передачи информации;
информационный подход, механизмы целенаправленного и самоконтролируе-
мого поведения в естественной и искусственной природе, живых организмах
и обществе; проблемы понимания разума и его роли
в управлении

ПРЕДМЕТЫ ФУНДАМЕНТАЛЬНОЙ КИБЕРНЕТИКИ == общие для всех
целевых устойчивых управляемых систем особенности
и законы функционирования; ограничения, свойственные управляемым систе-
мам, их выявление и познание происхождения; обратные связи, гомеостаз
(равновесие), устойчивость систем, их выявление и исследование; пути прак-
тического использования основ кибернетики; линейная равновесная термоди-
намика систем; отрицательная обратная связь

СИНЕРГЕТИКА³ – наука о самоорганизации в живых (биологических
и социальных) и неживых объектах (как системах) + Д

ОБЪЕКТЫ СИНЕРГЕТИКИ: нелинейные неустойчивые неравновесные (нахо-
дящиеся вдали от состояний равновесия), открытые (способные обмениваться
с окружающей средой веществом, энергией и информацией), когерентные (имею-
щие согласованное протекание процессов), диссипативные (необратимые во вре-
мени, рассеивающие материю) системы, их состояния, условия существования;
динамический хаос; проблемы рождения
и развития материи, разума, мышления

ПРЕДМЕТЫ ФУНДАМЕНТАЛЬНОЙ СИНЕРГЕТИКИ == принципы
и механизмы организации и самоорганизации, возникновения, развития, само-
усложнения и разрушения систем, появления новых
качеств у целого, которым не обладают его части; общие закономерности
и граничные условия, характеризующие возникновение самоорганизующихся
систем, их структур, функций, качеств; эволюция, развитие, переходы
от хаоса к порядку и, наоборот, от порядка к хаосу; точки бифуркации
(состояния системы, после наступления которых возможно множество вари-
антов ее дальнейшего развития), их установление и анализ, нахождение
и исследование свойств; аттракторы (наиболее реальные траектории разви-
тия системы в фазовом пространстве после точек бифуркации, относительно
устойчивые состояния, «притягивающие» к себе множество траекторий
развития); нелинейная неравновесная термодинамика систем; положительная
обратная связь

РЕФЕРЕНТИКА – наука о рефлексии, в том числе саморефлексии (самоосознании), в живых системах + Д

ОБЪЕКТЫ РЕФЕРЕНТИКИ: осознающие себя, ссылающиеся на себя, самообщающиеся, контингентные⁴, самообучающиеся системы (самореферентные системы); их выявление, познание, ...

ПРЕДМЕТЫ ФУНДАМЕНТАЛЬНОЙ РЕФЕРЕНТИКИ == принципы, механизмы, законы операционной замкнутости и контингентности: модели систем, среды, взаимодействия с ней системы, их единение

АВТОПОЭТИКА – наука о самовоспроизводстве объектов (как систем) + Д

ОБЪЕКТЫ АВТОПОЭТИКИ: автономные, операционно замкнутые («не зависящие» от среды), самосотворяющиеся, самовоспроизводящиеся системы с неизменной организацией (автопоэтические системы)

ПРЕДМЕТЫ ФУНДАМЕНТАЛЬНОЙ АВТОПОЭТИКИ == принципы, механизмы, законы автономности, сохранения своей организации, производства и потери собственных элементов в сети собственных элементов

СИСТЕМОТЕХНИКА – часть прикладной системологии – наука о разработке (проектировании, создании, испытании) и эксплуатации искусственных сложных технических и организационных систем

ТЕКТОЛОГИЯ – наука об организации в системах разной природы

¹ == означает: идет перечисление объектов, далее добавляется ..., их познание, описание, изучение, анализ, синтез, ...

² Д – область практической деятельности человечества с объектами и предметами данной науки.

³ От synergos – вместе действующий, буквально: энергия совместных действий.

⁴ Контингентность – рефлексивность познающего наблюдателя и наблюдения за наблюдателями. Рефлексия – главные отношения в системе, определяющие ее идентичность, назначение, смысл. Операционная замкнутость – избирательное восприятие нужных сигналов из среды для запуска рекурсий. Автономность – отсутствие влияния входных воздействий. Самореферентность – способность системы ссылаться на себя, когда среда становится внутренним миром системы.

Рис. 2.5. Некоторые понятия формальной информатики

система с широко используемыми терминами в другой их трактовке, таких как солнечная система, система счисления, вычислительная или программная система, простейшая система управления и т. п. Подобные объекты, называемые системами, могут не иметь тех системных свойств (в частности, свойства эмерджентности), которые и отличают эти объекты с точки зрения их моделирования как системы в нашем понимании от комплексов, наборов подсистем, множества элементов и т. п.

Помимо рассмотрения свойств целевости, целого, цельности и целостности, объект, рассматриваемый как система, должен иметь важнейшее (системное) свойство *эмерджентности* (эмергентности, синергии) – несводимости свойств системы к совокупности свойств частей (элементов), из которых она состоит, и невыводимости из них. Наличие этого свойства означает невозможность познания системы (объекта, адекватно представляемого как система) только аналитическими методами.

Альтернативой свойству эмерджентности является *аддитивность* – суммируемость, выводимость свойств системы из совокупности свойств образующих ее элементов. Аддитивные системы – это простые системы, познаваемые аналитическими методами. В отличие от них эмерджентные системы считаются сложными. Чем сильнее проявляется свойство эмерджентности по сравнению с аддитивной совокупностью свойств элементов системы, тем труднее познавать систему, выявлять связи (особенно нелинейные, непостоянно проявляющиеся, неустойчивые) между ее элементами. Заметим, что понятие сложности объекта (системы) относительное: для одного исследователя, при одних условиях, для одних целей объект может представляться как сложная система, для других – как простая.

Укажем еще два системных свойства: пороговости и иерархичности.

Пороговость означает, что некоторый объект становится системой, т. е. приобретает другие основные системные свойства и особенности, переходит в существенно отличные новые («системные») состояния, только если его внутренние или внешние ресурсы превышают некоторый порог.

Иерархичность – свойство, характеризующее тем, что отдельные части системы могут рассматриваться как локальные системы – подсистемы всей системы.

Со множеством других свойств систем можно ознакомиться по специальной литературе (см., в частности, [9–11]).

2.3.2. Системные принципы, законы и закономерности

Помимо видов и свойств систем, их описания, классификации и сравнения общая теория систем рассматривает системные принципы, законы и закономерности, основные характеристики систем, задачи их исследования и т. п. Желающие подробнее ознакомиться с ними самостоятельно могут начать это с работы [8]. Чтобы иметь представление об этих понятиях, рассмотрим несколько их примеров, важных для информатиков.

Прежде всего остановимся на системных методологических принципах. Под ними понимают основные исходные положения используемого в исследовании подхода, в данном случае системного (§ 2.6). Из изложенного в первой главе ясно, что принципов, как моделей основ подхода, может быть много в силу разных целей и назначений их использования, исследуемых объектов и т. д. Действительно в литературе и Интернете можно найти десятки принципов, как, впрочем, и свойств, законов, закономерностей.

Вот некоторые из них.

Принцип обратной связи – без наличия обратной связи (отрицательной или положительной) между входом и выходом (а иногда и между взаимосвязанными и взаимодействующими элементами, частями и подсистемами) системы невозможно долгосрочно минимизировать отклонение траектории ее движения к цели от эффективной в непредвиденных меняющихся условиях. Этот принцип широко используется в биологических и социальных системах, в средствах управления.

Принцип функционального подобия (принцип «черного ящика») – при фиксированной структуре объекта существенным является лишь то внешнее поведение каждой из его подсистем при внешнем воздействии, которое определяет их взаимодействие, причем не существенно, каким способом, за счет какой внутренней структуры поведение достигается. Использование данного принципа позволяет строить модель объектов типа «черного ящика», в которых учитываются и описываются только входы и выходы объекта и не учитывается, не описывается внутреннее устройство его.

Принцип ограниченности ресурсов – результаты требуемого качества необходимо получать при минимальных затратах.

Принцип внешнего дополнения (в формулировке регуляризации С.Т. Бира, А.Н. Тихонова) – при решении взаимосвязанных задач,

особенно оптимизационных, число критериев качества результатов должно быть не меньше числа решаемых задач, достигаемых целей.

Принцип декомпозиции (У.Р. Эшби, Г. Клаус) – управляемый объект всегда можно рассматривать как состоящий из относительно независимых друг от друга частей (подсистем).

Принцип взаимной дополнителности – все системообразующие аспекты природных объектов взаимно дополняют друг друга.

Иногда методологические системные принципы оформляются в виде законов с нечетко определенными условиями их применения. Вот их примеры.

Закон (принцип) необходимого разнообразия У.Р. Эшби – разнообразие системы, способной справиться с решением проблемы, должно быть больше разнообразия проблемы («только разнообразие может победить (уничтожить) разнообразие»).

Закон (принцип) иерархических компенсаций Е.А. Седова – действительный рост разнообразия на высшем уровне обеспечивается его эффективным ограничением на предыдущих уровнях.

В отличие от подобных законов-принципов существуют так называемые предельные системные законы. Например, *предельный закон устойчивости* – оптимального «обмена» элементов системы на время ее жизни; *помехоустойчивости* – оптимального «обмена» времени существования системы на важные для этого внешние сигналы в условиях внутренних и внешних помех и др. [9, 8]. Это законы, выводимые в виде математических соотношений, которые ограничивают предельные возможности систем. Например, следствием предельного закона устойчивости является ограничительное утверждение: «для неограниченно долгого существования (бессмертия) системы необходим (но не достаточен!) логарифмический со временем рост числа ее элементов». Другими словами «системы с фиксированным числом невозобновляемых элементов заведомо смертны». Примеры других предельных законов приведены в [9].

Прежде чем переходить к рассмотрению системных закономерностей, необходимо отметить разницу между понятиями *закон* и *закономерность*.

Закон – это совокупность правил, нормативных актов, особенностей, объективно существующих свойств, зависимостей, ..., которым подчиняются (или должны подчиняться) все объекты (субъекты), попадающие под действие закона в оговоренных условиях его применимости. Это, например, химические, физические, юридические, предельные системные и подобные им законы.

Закономерность – совокупность правил, норм, ..., которым подчиняются в тех же условиях не все, но подавляющая или бóльшая часть попадающих под их действие объектов (субъектов).

Теперь отметим отличие физических законов от предельных системных законов-принципов.

Физические законы представляют собой исследовательские (познавательные) модели объективно существующей действительности.

В качестве **системных законов** могут выступать: физические законы как познавательные модели действительности, отражающие системные особенности природных объектов, в частности законы-принципы; созидательные модели виртуальной (ожидаемой, возможной, желаемой) реальности (предельные системные законы); исследовательские как модели моделей (модемоделей) (в приложениях).

Принципиальное отличие физических законов – их познавательный характер. Это означает, что критерием их истинности является подтверждение эмпирическими данными, экспериментальными результатами.

В отличие от них системные законы, полученные только с использованием логических, аналитико-математических и других формализованных методов на базе аксиоматических посылок и исходных утверждений (как, например, предельные системные законы), не требуют для доказательства их истинности экспериментальных подтверждений. Их истинность определяется соответствием полученных результатов, используемых для их доказательства, выводов и интерпретации логико-математического аппарата и технологий поставленным целям и задачам, исходным посылкам и условиям.

Другое дело – определение пригодности таких системных законов для исследования природных объектов. Здесь, безусловно, необходима их эмпирическая, экспериментальная апробация, проверка на соответствие действительности, реальности. Подчеркнем, *проверка не истинности, а пригодности, полезности* для решения конкретной практической задачи.

В отличие от законов системные закономерности выявляются, выводятся, формализуются, исходя из наблюдений и экспериментирования с природными объектами, рассматриваемыми как системы. Их примеры приведены в [8–10]. Для получения представления о том, что это такое, рассмотрим качественные системные закономерности, представленные на рис. 2.6 [9, 8]. Это следующие закономерности.

1. **Закономерность разнообразия** – количество видов систем данного класса возрастает с увеличением уровня иерархии системы.

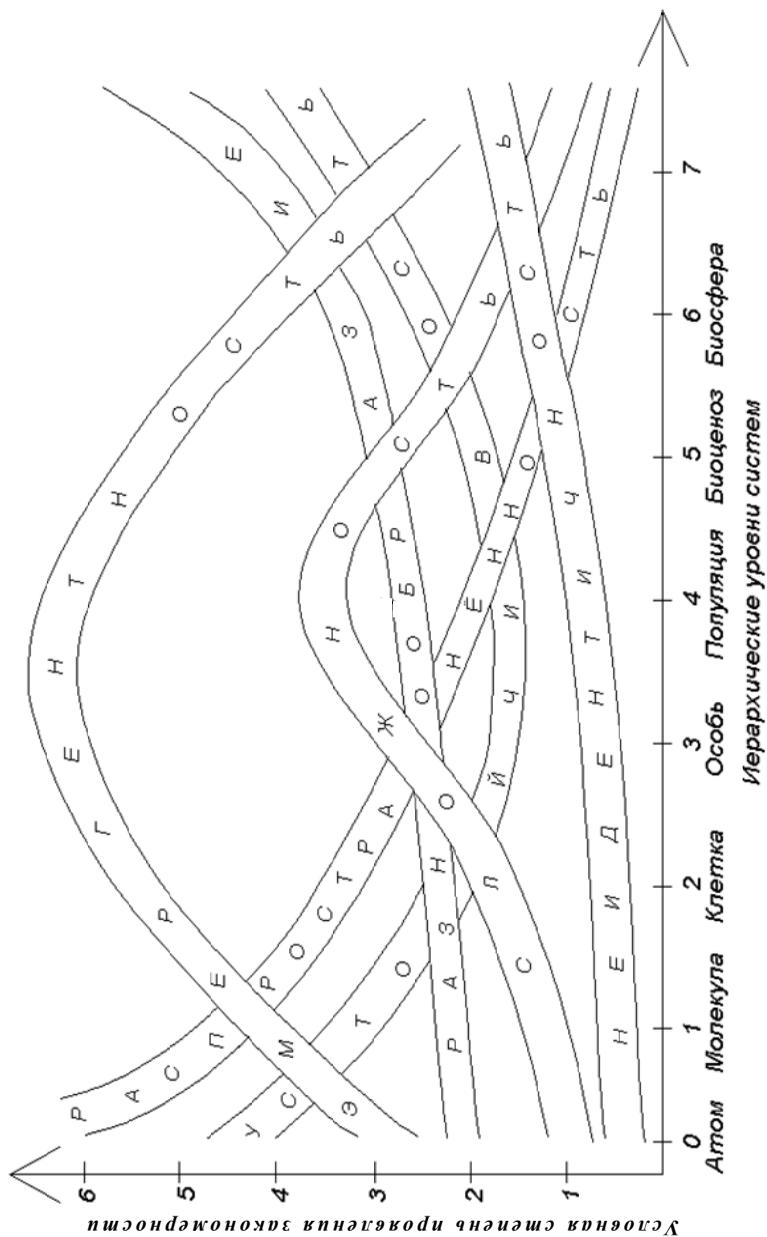


Рис. 2.6. Условное изображение связи системных закономерностей с уровнем иерархии систем

2. *Закономерность распространенности* – количество однотипных систем в конкретном пространстве убывает с увеличением их размеров.

3. *Закономерность сложности и варибельности* – варибельность структур по типам систем, составляющим класс, возрастает быстрее, чем возрастает средняя сложность класса.

4. *Закономерность устойчивости* – средняя способность системы противостоять внешним воздействиям для самосохранения у физических систем уменьшается, а у биологических – увеличивается с ростом их сложности (феномен Фёрстера) и имеет явную тенденцию к снижению при переходе от физико-биологической иерархии к социально-технической.

5. *Закономерность эмерджентности* (эмергентности) – степень несводимости свойств системы к свойствам ее отдельных элементов для физико-биологических систем возрастает до уровня особи и убывает для более высоких иерархий (ср. с закономерностью сложности).

6. *Закономерность неидентичности* – степень различия систем одного и того же типа возрастает с повышением их уровня иерархии.

§ 2.5. ВИДЫ ОБЪЕКТОВ С ТОЧКИ ЗРЕНИЯ ИХ МОДЕЛЬНОГО ПРЕДСТАВЛЕНИЯ

Одной из общих задач системологии и моделирования объектов является выделение видов объектов с точки зрения их системного модельного представления. Фрагмент разнообразия видов представлен в табл. 2.4.

Таблица 2.4

Виды объектов с точки зрения их системного модельного представления

1. По возможности управления извне	1.1. Управляемые	1.2. Неуправляемые
2. По возможности связи с окружающей средой, обмена с ней веществом, энергией, информацией	2.1. Открытые	2.2. Закрытые
3. По способу появления (создания) и усложнения	3.1. Составные (конструируемые, создаваемые по частям)	3.2. Организовые (рождаемые, выращиваемые, растущие и эволюционирующие как цельное целое)

Продолжение табл. 2.4

<p>4. По приспособляемости к окружающей среде – по степени самоприспособления</p>	<p>4.1. Приспособляемые</p> <p>4.1.1. Частично приспособляющиеся (слабоприспособляющиеся)</p> <p>4.1.1.1. Самонастраиваемые (саморегулируемые (...щиеся), перестраивающие значения параметров)</p>	<p>4.2. Неприспособляемые, неприспособляющиеся</p> <p>4.1.2. Полностью приспособляющие (адаптивные)</p> <p>4.1.2.1. Самосознаваемые (самообучаемые, самореферентные, перестраивающие алгоритмы, правила, цели)</p>
	<p>4.1.1.2. Самоорганизующие (перестраивающие структуру)</p>	<p>4.1.2.2. Самосотворяемые, самовоспроизводимые (авто-(ауто-)поэтические*, живые)</p>
<p>5. По сходству элементов</p>	<p>5.1. Однородные</p>	<p>5.2. Неоднородные</p>
<p>6. По применимости принципа суперпозиции: реакция на сумму воздействий равна сумме реакций на каждое воздействие (по виду математического аппарата)</p>	<p>6.1. Линейные, аддитивные (принцип применим)</p>	<p>6.2. Нелинейные, неаддитивные (принцип не применим)</p>
<p>7. По характеру изменения, разрывности состояний во времени</p>	<p>7.1. Континуальные (непрерывно изменяемые)</p>	<p>7.2. Квантовые (изменяемые скачками)</p>
<p>8. По поведению (изменению) состояний во времени</p>	<p>8.1. Статические, статичные (не изменяющие своих состояний, значений параметров)</p>	<p>8.2. Динамические, динамичные (изменяющиеся, переходные)</p> <p>8.2.1. Развивающиеся (улучшающие показатели своих состояний, особенностей)</p> <p>8.2.2. Деградирующие (ухудшающие показатели своих состояний, особенностей)</p>

Продолжение табл. 2.4

9. По зависимости состояний, параметров от начала отсчета времени	9.1. Стационарные (со стабильными во времени состояниями/параметрами)	9.2. Нестационарные
10. По возвращаемости в исходное состояние после прекращения действия отклоняющих сил	10.1. Устойчивые (возвращающиеся в то же состояние)	10.2. Неустойчивые (переходящие в другое состояние)
11. По чувствительности состояний к внешним и внутренним воздействиям (силам)	11.1. Равновесные (мало чувствительные к воздействиям)	11.2. Неравновесные (чувствительные даже к незначительным воздействиям)
12. По способности рассеивать (терять со временем) вещество, энергию, информацию, по обратимости во времени	12.1. Диссипативные (рассеивающие), необратимые во времени	12.2. Консервативные (стабильные, не рассеивающие)
13. По согласованности происходящих в объекте процессов, переходов, состояний	13.1. Когерентные (согласующие процессы, ...)	13.2. Инкогерентные (не согласующие процессы, ...)
14. По проявлению своей способности к изменениям, действиям	14.1. Активные	14.2. Пассивные
15. По типу множества моментов изменения состояния, переходов, воздействий, выходов во времени	15.1. Непрерывные (континуальное множество)	15.2. Дискретные (конечное, счетное множество)
16. По множеству уровней состояния (значений входных и выходных переменных)	16.1. Непрерывнозначные (аналоговые, континуальное множество)	16.2. Цифровые (конечное множество)
17. По распределению базовых элементов в пространстве	17.1. Распределенные	17.2. Сосредоточенные
18. По распределению параметров элементов в пространстве	18.1. С распределенными параметрами	18.2. С сосредоточенными параметрами

19. По необходимости обязательного общения с другими объектами и окружающей средой	19.1. Неавтономные (несамостоятельные)	19.2. Автономные (могущие обходиться самостоятельно)
20. По наличию обратной связи (отрицательной, положительной; выход–вход или поэлементной)	20.1. Замкнутые	20.2. Разомкнутые

* Способные поддерживать свою целостность в процессе функционирования и взаимодействия с окружающей средой.

Другие виды систем можно найти в литературе по теории управления, системологии, системному анализу, кибернетике, синергетике.

§ 2.6. СИСТЕМНЫЙ ПОДХОД К ИССЛЕДОВАНИЮ ОБЪЕКТОВ

Одними из практических приложений системологии являются разработка и применение в исследовании объектов самой различной природы системного подхода. Системный подход к исследованию объектов есть методологическая концепция, основанная на стремлении построить целостную картину изучаемого объекта как системы. Системный подход – это особая, внутренне единая исследовательская позиция, направленная на ясное развернутое выражение процедур системного представления изучаемых объектов и способов их исследования. Он отличается от элементного (элементаристского) подхода к представлению объектов, ориентированного на исследование отдельных элементов объекта и через результаты этих исследований – всего объекта. По охвату изучаемого объекта системный подход может реализоваться в рамках единичного или комплексного исследования. В отличие от единичного комплексность означает охват как можно большего для исследования множества особенностей строения и «жизни» (функционирования) объекта в среде. Системность же базируется на рассмотрении объекта со всех сторон как системы.

Отличительные особенности системного подхода:

1) изучаемый объект рассматривается при модельном представлении его как система, описание и исследование отдельных элементов

которой не выступают как самоцель, а выполняются с учетом их места в «целом»;

2) исследование объекта не отделяется от условий его существования, функционирования. Объект сам рассматривается как часть некоего «целого»;

3) один и тот же исследуемый элемент рассматривается как обладающий разными характеристиками, функциями и даже принципами (по)строения. При исследовании особое внимание уделяется иерархичности строения объекта;

4) при системном подходе на первое место выступают не только механизмовые (конструктивные) и причинные объяснения функционирования объекта, но и целесообразность включения в его состав отдельных элементов (пример подхода, при котором выявляются прежде всего конструктивные, реже причинные объяснения, – ранний подход при разработке физических законов);

5) допускается возможность внутренней самоорганизуемости исследуемого объекта (т. е. наличия у него некоторого множества индивидуальных характеристик и степеней свободы и, как следствие, альтернатив поведения), рассмотрения объекта как источника преобразования самого себя;

6) выявление целей исследований и определение подлежащих решению задач, проблем¹ производится на основании анализа общей цели, исходя из общей идеи решения проблемы, когда альтернативы сравниваются в первую очередь по критерию затрат (стоимость, эффективность).

Иными словами, системный подход – это методология познания частей и целостности объекта на основании рассмотрения объекта как целого в отличие от классического элементного подхода, ориентированного на познание объекта как целого через его части (элементы). При системном подходе стремятся к познанию сложного через простые системные модели, в отличие от описания объекта переходом от простых моделей элементов объекта к сложным моделям всего объекта. В отличие от элементного подхода в системных методах исследования используется не только модель объекта, но и модели среды и

¹ Обратим внимание на отличие понятий *задача* и *проблема* (от гр. *próblēma* – задача, сложный вопрос, требующий изучения и разрешения, противоречивая ситуация). В отличие от проблемы, когда мы говорим *задача*, мы знаем, как ее поставить и решить.

ситуации, в которых находится исследуемый объект, изучаются прежде всего не столько свойства, особенности строения, правила функционирования и другие элементы объекта, сколько его системные особенности (свойства, строение, функционирование, ...) как «целого», то, что делает объект целым, обеспечивает его эмерджентность. Сам объект считается тем более сложным, чем сильнее проявляется его свойство эмерджентности, т. е. чем меньше исследуемые особенности строения объекта, его свойства, правила и условия функционирования выводятся, следуют, обуславливаются особенностями, свойствами строения и функционирования его элементов. При этом чаще речь идет не о сложности объекта как такового, а о сложности его поведения.

Обратим внимание на то, что важно не только объявить о применении системного подхода, но и определить (указать) при практическом применении его, как он реализуется, в частности определить, к какому классу относится системное представление (как модель) исследуемого объекта (например, какая это система – простая или сложная; кибернетическая, синергетическая или самореферентная; статическая или динамическая; линейная или нелинейная; открытая или закрытая и т. п.) и какие системные свойства объекта наиболее значимы для конкретного исследования.

ЗАКЛЮЧЕНИЕ

В настоящей главе рассмотрены важнейшие базовые понятия информатики: модель, моделирование, система и связанные с ними минимально необходимые для дальнейшей ориентации и работы информатика сведения. В заключение главы приведем дополнительные (факультативные) сведения, которые необходимо иметь в виду при практическом применении полученных знаний.

1. Природные законы и закономерности могут быть представлены в виде их математического описания, поскольку нет иного способа их количественного выражения. Но согласно моделированию любой объект допускает множественное представление его моделями в зависимости от цели моделирования, требований к модели, имеющихся ресурсов, условий функционирования объекта и т. д. Поэтому может быть несколько моделей законов и закономерностей объектов природы.

2. Любые модели вводятся в условиях искусственного упрощения объектов, явлений, процессов, условий их «жизни» и окружения, сохранения главного для цели моделирования и опускания, неучета

несущественного. Поэтому степень полезности модели определяется тем перечнем задач, для решения которых она пригодна с учетом детальности получаемых при этом результатов решения.

3. Степень общности применения модели и детализация решения конкретных задач являются противоречивыми требованиями к модели. Возможна замена одной модели другой путем огрубления одних требований для повышения общности или, наоборот, конкретизации, детализации, упрощения других.

4. Главная цель построения любой модели:

а) понимание с точки зрения постановки задачи, цели исследования «механизма» изучаемого явления, процесса, строения или функционирования объекта, влияния внешних и внутренних факторов на него;

б) использование ее для решения генезисных, диагностических, прогностических и других задач с ее помощью.

Прогностическое назначение модели – это очень важное ее назначение. Вторым важным назначением является использование одной модели для построения на ее основе других моделей, например, для имитации поведения объекта, уменьшения объема хранимых данных, а также замены Данных их идентификационной моделью для удобства организации исследований (упрощения операций хранения и представления с последующим восстановлением Данных по модели при необходимости), улучшения качества Данных (устранения пропусков, выбросов, трендов и т. п.) и решения иных «внутренних» или «внешних» по отношению к цели моделирования задач.

5. Процесс создания и/или выбора модели является неформализуемой до конца задачей, решение которой зависит от ряда факторов, в частности от квалификации исследователя и имеющихся знаний о моделируемом объекте. При этом надо иметь в виду, что любая модель – это лишь определенное приближение действительности. Критерий правильности модели – качество решения задач согласно назначению модели. Главную проверку (верификацию) модели обеспечивает ее практическое использование в соответствии с той целью, с которой она создавалась. Например, степень совпадения прогноза ситуации, получаемой по модели, с фактически наступившей ситуацией, определяемой данными, которые не использовались при построении модели.

6. Выбор модели определяется природой реальности. Так, в ряде случаев между X и Y может существовать функциональная детерминированная связь, но субъект вместо детерминированной вводит вероятностную модель (см. далее), чтобы учесть погрешности измерений,

случайно воздействующие неучитываемые факторы и т. д. Другая ситуация, когда природа действительности (объекта, условий) такова, что конкретному X обязательно соответствует такое множество значений Y , которое можно описать только языком теории вероятностей.

7. Различные оптимальные модели позволяют находить наилучшие или предельно возможные решения. В этом их несомненная теоретическая и практическая значимость. Однако следует весьма аккуратно подходить к их практическому применению. Во-первых, модели (методы, алгоритмы) оптимальные по одному критерию, как правило, не являются таковыми по другим критериям. Во-вторых, нарушение условий, при которых существует оптимальная модель (метод, алгоритм), может привести к существенным потерям ею своих достоинств. В-третьих, необходимо всегда проверять наличие и поддерживать условия оптимальности как самой модели (метода, алгоритма), так и условий реализации всех операций технологического процесса (алгоритма) ее получения, хранения, применения.

ВОПРОСЫ ДЛЯ САМОПОДГОТОВКИ

1. Что такое моделирование объекта?
2. Почему при определении понятия *моделирование* необходимо оговаривать все три стадии моделирования?
3. Исходя из чего определяется цель и формулируется постановка задачи моделирования исследуемого объекта?
4. Охарактеризуйте возможные уровни моделирования объектов различной природы.
5. Перечислите и охарактеризуйте этапы моделирования объектов на разных стадиях: идентификация объекта, его исследование по модели, перенос на объект результатов исследования объекта по модели и применение модели.
6. Перечислите и охарактеризуйте основные требования к модели объекта и сопоставьте их со свойствами модели. Есть ли среди требований противоречивые? Если да, то назовите их и поясните причину противоречивости.
7. Представьте и охарактеризуйте разные классификации моделей.
8. Перечислите и охарактеризуйте различные альтернативные виды моделей.
9. Что такое семиотика, синтактика, семантика, прагматика, сигматика, синтаксис?

10. Что такое система? Почему вы придерживаетесь именно этого определения (дефиниции) термина *система*?

11. Что такое системология?

12. Перечислите и охарактеризуйте основные свойства систем.

13. Связана ли эмерджентность системы с ее сложностью? Можно ли утверждать, что эмерджентность системы есть следствие ее сложности или наоборот?

14. Что такое системные принципы, законы, закономерности? В чем сходство и отличие между ними?

15. Приведите примеры системных принципов, законов, закономерностей. С какими из них вам уже приходилось встречаться?

16. Перечислите и охарактеризуйте виды объектов с точки зрения их системного модельного представления.

17. Что понимается под системным подходом к исследованию объектов? В чем его специфика? Какие еще вы знаете подходы? Сопоставьте их с системным подходом.

Глава третья

ЭКСПЕРИМЕНТИРОВАНИЕ КАК МЕТОД ИССЛЕДОВАНИЯ И ЭТАП МОДЕЛИРОВАНИЯ ОБЪЕКТОВ

§ 3.1. ФИЗИЧЕСКИЕ И МАТЕМАТИЧЕСКИЕ ВЕЛИЧИНЫ

3.1.1. Необходимые определения

Начнем настоящую главу с рассмотрения разницы между понятиями *физическая величина* и *математическая величина*. Прежде всего определимся с понятием величина. Согласно различным словарям *величина* – то, что можно измерить (физическая величина) или вычислить (математическая величина).

Физическая величина – это:

а) особенность, свойство, общие в качественном отношении многим физическим объектам (масса, плотность, сила, вязкость, объем, площадь, длина, ... физического объекта);

б) характеристика физических объектов или явлений материального мира.

Математическая величина – обобщение конкретных понятий: длины, площади, объема и иных величин математического объекта, фигуры.

Рассмотрим еще два связанных с величиной понятия.

Аналоговая величина – величина, принимающая значения в некоторой непрерывной области значений.

Цифровая величина – величина, принимающая значения из некоторого фиксированного конечного множества.

Обратим внимание на двоякость термина *величина*. С одной стороны, термин *величина* (как математическая, так и физическая) может обозначать величину в общем смысле (длина, площадь или температура, масса, концентрация вещества), с другой – конкретную, определенную величину (длина данного отрезка или стержня, концентрация точек в некотором участке рисунка или сахара в данной пробе напитка).

Согласно приведенным определениям физическая величина отличается от математической тем, что первую (физическую величину) мы измеряем, чтобы перевести ее значение в число, а вторую (математическую величину) вычисляем или исчисляем. Что же это за операции?

Они и другие важные понятия, касающиеся этих операций, представлены в табл. 3.1.

3.1.2. Виды физических величин

С точки зрения измерения значений, математических операций с результатами измерений физические величины принято делить (см., например, [12]) на два класса: экстенсивные и интенсивные.

Экстенсивные (аддитивные) физические величины – это такие, для которых справедлив принцип аддитивности (суммируемости): если новый объект получен объединением (суммированием) нескольких объектов, то значение его экстенсивных величин получается как сумма значений этой величины для первоначальных объектов. Это, например, длина, масса, время. Значения экстенсивных величин допускают суммирование, умножение на числовой коэффициент, деление друг на друга.

Т а б л и ц а 3.1

Исходные термины

Термин (английский эквивалент)	Толкование термина
Измерение (<i>количественное, численное, метрологическое</i>) (measurement, measuring <i>/quantitative, numerical; metrological/)</i>	<ol style="list-style-type: none"> 1. Опытное нахождение количественного значения (эквивалента) физической величины с помощью специальных средств путем сравнения измеряемой величины с некоторой мерой, значение которой принято за единицу измерения 2. Действия, связанные с нахождением числового значения измеряемой физической величины в принятых единицах измерения 3. Преобразование множества значений физических величин в числовое множество или множества физических величин в множество математических величин
Измерение (<i>обобщенное понятие, алгоритмическое, символическое</i>) (measurement)	<ol style="list-style-type: none"> 1. Алгоритмическая операция, в результате выполнения которой наблюдаемому состоянию (материального) объекта (процесса, явления) ставится в соответствие определенное обозначение: число, номер или символ 2. Действие по установлению соответствия определенному состоянию наблюдаемого (исследуемого) объекта

Продолжение табл. 3.1

Термин (английский эквивалент)	Толкование термина
	(явления, процесса) символа (числа, номера) из выбранной для регистрации этого состояния измерительной шкалы 3. Преобразование множества состояний физического объекта в множество символов измерительной шкалы
Измерительная шкала (measuring scale)	Множество обозначений (чисел, номеров, символов), используемых для регистрации состояния наблюдаемого объекта
Сила измерительной шкалы (measuring scale strength)	Допустимое количество операций (действий) над данными, выраженными в этой шкале
Вычисление (calculation, computation, computing)	1. Выполнение арифметических и логических операций 2. Произведение арифметических действий над числами 3. Преобразование одного множества чисел в другое множество чисел
Исчисление (calculus)	1. Выражение в каком-либо числе, количестве 2. Формализованный вывод логических утверждений 3. Процесс образования синтаксически правильных символических выражений из букв алфавита системы, языка, которые ставятся в соответствие предметам определенной области (например, объективного мира) 4. Преобразование одного множества выражений, утверждений, знаний в другое
Идентификация (объекта) (identification)	Часть (этап) моделирования объекта, связанная с построением математической модели объекта по экспериментальным Данным
Оценивание (статистическое) (оценка – как процедура) (estimation) Оценка – функция наблюдений (estimator) Оценка – численное значение (estimate)	1. Получение приближенного значения характеристик или параметров случайных элементов (величин, векторов, функций) по их реализациям (выборочным данным, выборке из генеральной совокупности) 2. Математические операции определения значений параметров и характеристик случайных элементов путем вычисления их по выборке из генеральной совокупности

Термин (английский эквивалент)	Толкование термина
Имитация (<i>объекта</i>) (imitation)	1. Подражание, более или менее точное воспроизведение свойств, строения, функционирования, поведения объекта, воспроизведения его в разнообразных вариантах 2. Часть (этап) моделирования, связанная с искусственным воспроизведением всех или части рабочих функций или свойств моделируемого объекта с помощью модели (на модели)
Симуляция (simulation)	Сознательно организованная ложная имитация объекта
Генезис (<i>объекта</i>) (genesis)	Определение состояния (объекта) в прошлом
Диагностика (<i>объекта</i>) (diagnostics)	Определение состояния (объекта) в настоящем
Прогнозирование (<i>объекта</i>) (prediction, forecasting)	Определение (предсказание) состояния (объекта) в будущем
Система (system)	Модельное представление чего-либо или кого-либо (объекта) в виде целого, состоящего из совокупности взаимосвязанных элементов, объединенных для реализации общей цели, обособленного от окружающей среды, взаимодействующего с ней как цельное целостное и проявляющего при этом основные (системные) свойства объекта
Метод (method)	Правило или система стандартных правил, подходов, способов, путей, приемов, действий, направленных на достижение цели или решение конкретной задачи
Алгоритм (algorithm)	1. Формальное представление, точное предписание (в виде конечного набора правил), однозначно определяющее содержание и последовательность чисто механически выполняемых действий (алгоритмических операций), переводящих исходные данные задачи в искомый результат 2. Сам процесс решения поставленной задачи в виде такой совокупности действий (операций), предпринимаемых по строго определенным правилам, которая после последовательного (пошагового) их выполнения приводит к такому ее решению, когда результат однозначно определяется исходными данными

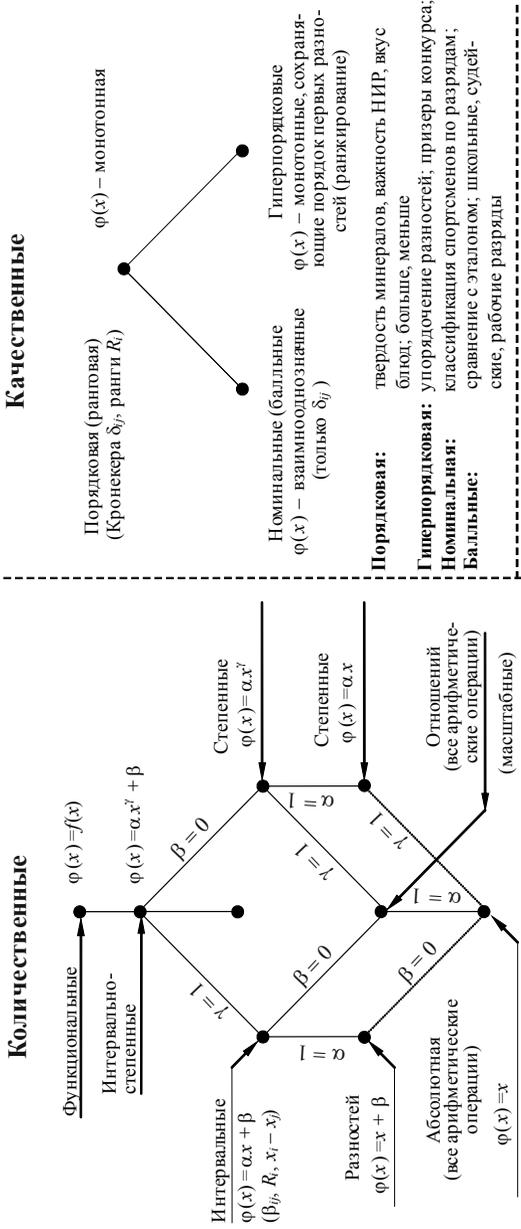
Интенсивные физические величины не удовлетворяют принципу аддитивности. Пример интенсивных величин – давление, температура. Так, нельзя предложить никакую простую операцию, с помощью которой тело с температурой 50 °С (либо по шкале Реомюра, Фаренгейта) можно объединить с телом с температурой 30 °С и получить тело с температурой 80 °С. Для введения шкалы и единицы измерения таких величин определяются две крайние точки шкалы, соответствующие началу и концу отсчета (например, точки таяния льда и кипения воды по температурной шкале Цельсия). Разделив всю шкалу на заданное число частей, за единицу измерения принимается отношение разности значений, соответствующих концу и началу точек шкалы на заранее заданное число частей (единиц измерения). В отличие от экстенсивных величин для интенсивной величины, измеренной в таких шкалах, умножение на число или деление друг на друга ее значений не имеет физического смысла. Например, физически неверно утверждать, что при 30 °С в 2 раза теплее, чем при температуре 15 °С. Однако если в качестве температурной шкалы выбрать шкалу Кельвина, где нулевая точка соответствует абсолютному нулю температур, то в такой шкале утверждение, что температура 500 К в пять раз выше, чем 100 К, физически корректно, хотя температура как физическая величина все равно осталась интенсивной. Речь здесь идет лишь о целевом выборе шкалы измерения. Остановимся на этом очень важном вопросе более подробно.

3.1.3. Измерительные шкалы

Как следует из только что рассмотренного, при измерении физических величин, т. е. при переводе значений физических величин в их количественные (числовые) эквиваленты, весьма важным является то, с использованием какой шкалы это измерение производится. Поскольку, как указывалось ранее, измерения могут быть количественными (числовыми) и категориальными (качественными, символьными¹), измерительные шкалы делятся на два класса: количественные и качественные (категориальные). Их разнообразие демонстрируется рис. 3.1, 3.2 и табл. 3.2 [4, 12–16].

¹ Обобщенное понятие измерения объединяет в себе те и другие.

Измерительные шкалы



- Абсолютная (натуральная):** количество стульев, студентов
- Отношений:** отношение масс $\varphi[f(a_1)]/\varphi[f(a_2)] = f(a_1)/f(a_2)$
- Интервальная:** температурные $(\varphi[f(a_1)] - \varphi[f(a_2)]) / (\varphi[f(a_3)] - \varphi[f(a_4)]) = \frac{f(a_1) - f(a_2)}{f(a_3) - f(a_4)}$
- Разностей:** летоисчисление

Рис. 3.1. Измерительные шкалы

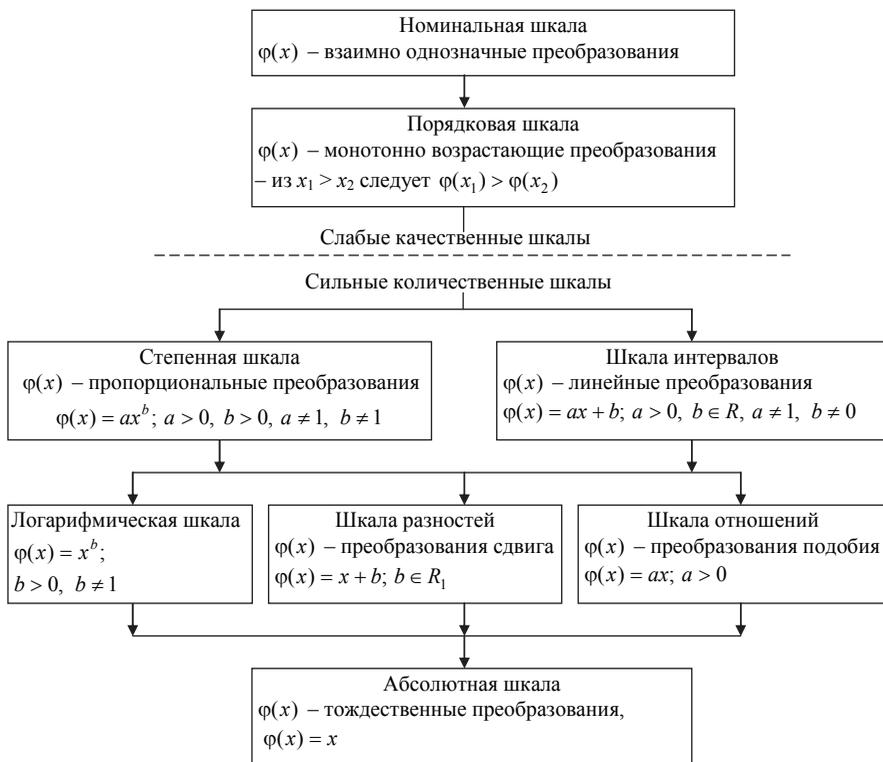


Рис. 3.2. Иерархическая структура измерительных шкал [11]

Как видно из рис. 3.2, самая слабая шкала – качественная номинальная шкала (шкала наименований, классификационная шкала). Примеры измерений в номинальной шкале – номера автомашин, телефонов, кодов городов, лиц и т. д. Такая шкала может использоваться только для различения объектов. Обратим внимание на две особенности этой шкалы [12]. Первая: если двум различным объектам А и В поставить в соответствие одно и то же значение номинальной шкалы, то это означает, что при измерении эти элементы не различаются. Вторая особенность: если при измерении в номинальной шкале используются символы $1, 2, \dots, n$, то они являются не числами, а цифрами, служащими только для обозначения и различия объектов.

Измерительные шкалы [15]

Тип шкалы	Определяющие отношения	Эквивалентное преобразование шкалы	Допустимые операции над данными (первичная обработка)	Вторичная обработка данных	Примеры
Номинальная (наименований, классификаций)	Эквивалентность =	Перестановка классов, их переименования	Вычисление символа Кронекера $\delta_{ij} = \{1 : x_i = x_j;$ $0 : x_i \neq x_j\}$	Вычисление относительных частот и операции над ними	Имена, названия, номера домов и автомобилей, знаки, символы
Порядковая (ординальная, ранговая)	То же и предпочтение =, >	Не меняющее порядка («монотонное»)	δ_{ij} и вычисление функции сравнения $C_{ij} = \{1 : x_i \geq x_j;$ $0 : x_i < x_j\}$	То же и вычисление рангов $R_j = \sum_{i=1}^N C_{ij}$, квантилей, операции над ними	Всевозможные упорядочения, балльные оценки
Интервальная	То же и постоянство отношения интервалов =, > $\Delta y / \Delta x = \text{const}$	$Y = ax + b$ $a > 0, b \in R$	То же и вычисление интервалов $\Delta_{ij} = x_i - x_j$	Арифметические действия над интервалами	Температура, летоисчисление, высота местности, географическая широта

Окончание табл. 3.2

Тип шкалы	Определяющие отношения	Эквивалентное преобразование шкалы	Допустимые операции над данными (первичная обработка)	Вторичная обработка данных	Примеры
Циклическая (периодическая)	То же и периодичность $=, >$, $y = x + b + n\tau$, $\tau = \text{const}$, $n = 1, 2, \dots$	$y = x + b + n\tau$	То же	То же	Направления на страны света, время суток, фазы колебаний, долгота, времена года
Отношений	То же и постоянство отношения замеров $=, >$, $y/x = \text{const}$	$Y = ax + 0$	Все арифметические операции	Все арифметические операции	Длина, вес, объем, масса, площадь, деньги
Абсолютная	То же и абсолютность нуля и единицы	Шкала уникальная (числовая ось)	Любые арифметические и трансцендентные операции	Любые операции над числами	Счет единиц чего-либо, определение долей, основа других шкал

Качественная (категориальная) шкала порядков (ранговая, порядковая) допускает только любые монотонные (возрастающие или убывающие) преобразования, т. е. она допускает не только различие объектов (как номинальная шкала), но и упорядочение объектов по их измеряемым свойствам (характеристикам, показателям). Однако в отличие от сильных шкал порядковая шкала определяет измеряемое свойство объекта лишь с точностью до класса, когда множество всех классов упорядочивается по степени появления в них наблюдаемого свойства. При этом структура порядковой шкалы не меняется при любой взаимно однозначной подстановке, сохраняющей порядок. Следовательно, порядковая шкала позволяет заключить, что одно значение измеряемого свойства равно, ниже или выше другого его значения. Однако она не позволяет сказать, на сколько больше (или меньше) одно значение другого. Заметим, что, если в качестве алфавита шкалы для записи в ней результатов измерения используются цифры натурального ряда (а можно использовать любые символы, имеющие общепринятую упорядоченность), то получаемые при этом числа лишь указывают на усиление (или ослабление) изучаемого свойства объекта и не являются объектами арифметических операций. Например, при упорядочении объектов во времени или в пространстве; в соответствии с каким-либо качеством, когда не предвидится его точное измерение; когда качество измеримо, но в настоящий момент не может быть почему-то измеренным. Примеры таких шкал: твердости минералов Мооса (1811 г.), силы ветра (шкала Бофорта, 1806 г.), землетрясений (шкала Рихтера), сортности товара в торговле, социологические шкалы, измерительные шкалы (табл. 3.2).

Следует обратить внимание на то, что для данных, измеренных в ранговых шкалах, в ряде случаев разрабатывают специальные средства их обработки для ответа на вопросы, которые ставятся и решаются обычно в числовых шкалах. Например, для балльных ранговых шкал не применима операция среднего арифметического, обычно используемая в числовых шкалах как оценка математического ожидания (генерального среднего). Вместо среднего арифметического в ранговых шкалах может быть применима операция нахождения медианы или квантилей других порядков; для определения размаха – интерквантильная ширина; для оценивания связей – коэффициенты корреляции Спирмена и Кендалла (числовой теоретический аналог – конкоры [12, 17]) и т. д.

Более сильной, чем порядковая, является качественная гиперпорядковая шкала, которая помимо операции монотонных преобразова-

ний упорядочения допускает операции гипермонотонных преобразований, т. е. упорядочения как объектов, так и разностей численных значений этих объектов в шкале. Это, например, призерные шкалы конкурсов.

Допустимые операции в количественных (метрических) шкалах видны из рис. 3.1 и табл. 3.2. Сделаем по этому поводу несколько замечаний.

1. Нельзя отождествлять замеренные в интервальных шкалах переменные с изучаемым свойством объектов, когда свойства, измеренные в шкале интервалов, принимаются в качестве показателей других свойств, монотонно связанных с рассматриваемыми свойствами. *Например, физическое время измеряется в шкале интервалов, а отрезки времени, используемые для ранжирования умственных способностей человека, оформляются в порядковой шкале.* Второй важный аспект, связанный с интервальными шкалами, это то, что в них только интервалы имеют смысл настоящих чисел и только над интервалами следует выполнять арифметические операции, а не над самими отсчетами (числами) этой шкалы. Это является следствием того, что в интервальных шкалах нет абсолютного начала отсчета, а в качестве начала отсчета принимается некоторое условное значение. Поэтому такая шкала позволяет определить, какая из двух величин больше или меньше другой и на сколько, но не позволяет сказать, во сколько раз. Например, как уже указывалось, говорить, что вода с температурой $36\text{ }^{\circ}\text{C}$ вдвое теплее, чем с температурой в $18\text{ }^{\circ}\text{C}$, некорректно, так как при измерениях по Фаренгейту это будет странным, неверным, поскольку $^{\circ}\text{F} = 32 + 9\text{ }^{\circ}\text{C}/5$. В противоположность этому температура, измеренная в кельвинах, образует шкалу отношений, и здесь уже можно утверждать, что температура 300 K не только выше, чем 100 K , но и что она втрое выше.

2. Часто применяемые для измерения связанных свойств исходные шкалы интервалов становятся всего лишь шкалами порядка.

3. Шкалы отношений (подобия) являются подмножеством шкал интервалов. При выборе нулевой точки отсчета они сохраняют не только отношения свойств объектов, но и отношения расстояний между парами объектов.

4. Шкалы разностей как частный случай шкал интервалов – это единственные шкалы с точностью до преобразований сдвига $\varphi(x) = x + b$, где x – шкальное значение из области определения шкалы; b – действительное число. В этой шкале измеряется прирост про-

дукции предприятия (в абсолютных единицах) в текущий период по сравнению с прошлым, увеличение доходов, числа студентов, количество приобретений за какой-то период времени и т. д., а также летоисчисление (в годах), когда переход от одного летоисчисления к другому осуществляется изменением начала отсчета. Шкалы разностей, как и шкалы интервалов, сохраняют отношения интервалов между значениями пар объектов, но в отличие от шкал отношений не сохраняют отношения значений свойств объектов.

5. В самых сильных абсолютных шкалах единственным допустимым преобразованием является тождественное преобразование $\varphi(x) = x$. Абсолютные шкалы сохраняют любые отношения между числами значений измеряемых свойств объектов: различие, порядок, отношение интервалов, отношение и разность значений и т. д. Абсолютная шкала имеет абсолютный нуль и абсолютную единицу. Она единственная, уникальная. Ее главные особенности по сравнению со всеми остальными шкалами – безразмерность (отвлеченность) и абсолютность используемых единиц, что и допускает любые операции в этой шкале. Такой шкалой является числовая ось.

6. Шкала считается тем *сильнее*, чем меньше свободы в выборе преобразования $\varphi(x)$. Иерархия шкал по силе видна из рис. 3.1 и 3.2. Опасность применения недопустимых в этой шкале операций с данными хорошо иллюстрирует пример с непригодностью операции среднего арифметического в балльных шкалах. Пусть на экспертизу представлена подготовка студентов, оцениваемая по предметам x_1, x_2, x_3, x_4 . Качество подготовки каждого студента оценивается по пятибалльной системе, когда балл представляет собой округленное значение оценки по более дифференцируемой («истинной») системе оценок. В качестве примера приведем табл. 3.3.

Т а б л и ц а 3.3

Пример балльной экспертизы подготовленности студента

Предмет	Оценка подготовленности студента А		Оценка подготовленности студента В	
	«истинная»	в баллах	«истинная»	в баллах
x_1	4,3	4	3,6	4
x_2	4,4	4	3,6	4
x_3	3,4	3	3,7	4
x_4	4,3	4	3,8	4
Суммарная оценка	16,4	15	14,7	16
Средняя оценка	4,1	3,75	3,68	4,0

Из таблицы видно, что «истинно» лучше подготовлен студент А, а по среднебалльной системе эксперты вынуждены предпочесть ему хуже подготовленного студента В. Среднеарифметическое применимо для результатов, измеренных в шкалах интервалов, разностей, отношений и абсолютной, но недопустимо для порядковой шкалы. Близкой к среднеарифметической, но более устойчивой является медиана (квантиль порядка 0,5), рекомендуемая для шкал порядка, интервалов, разностей, отношений и абсолютной. Отсюда следует вывод о недопустимости других операций, связанных со средним (дисперсия, корреляция, регрессия и т. п.) в шкалах, где недопустимы операции типа среднеарифметического.

Расширением качественных измерительных шкал являются шкалы, построенные с использованием вероятностных либо нечетких и лингвистических переменных. Их отличие – принадлежность значений измеряемых величин и результатов измерения некоторому множеству шкальных значений с какой-то вероятностью (байесовский подход) или нечетким переменным с функцией принадлежности, принимающей любое значение из диапазона $[0, 1]$, а не только 0 или 1, как в классических («четких») измерительных шкалах (см. подпараграфы 4.4.2, 4.4.3).

§ 3.2. НАБЛЮДЕНИЕ И ЭКСПЕРИМЕНТ

Центральными для настоящей главы являются понятия *наблюдение* и *эксперимент*.

Наблюдение (*observation*) – метод исследования, основанный на целевом пассивном однократном восприятии объекта в реальных условиях (пассивный опыт).

Эксперимент (*experiment*) – научно поставленный в учитываемых и/или регулируемых естественных или искусственных условиях активный опыт с исследуемым объектом или его моделью, допускающий многократное повторение.

Сделаем несколько пояснений к этим понятиям.

1. Как следует из определений, наблюдение и эксперимент сходны в том, что они ориентированы на выполнение тех же двух функций: опытная проверка моделей объекта и того, что с ними связано (в частности гипотез и теорий), и формирование новых моделей (научных концепций, гипотез, теорий).

2. Наблюдения связаны с пассивными, а эксперимент – с активными действиями субъекта-исследователя. Активность означает, что

исследователь (наблюдатель) может вмешиваться в ход, «жизнь» наблюдаемого явления, процесса, объекта, их окружения, как-то изменять их или условия, в которых они наблюдаются. Назначение наблюдения объекта (за объектом) – сбор, накопление, систематизация первичных эмпирических (исходных) данных, необходимых для моделирования объекта. Основные требования к наблюдению – их объективность (непредвзятость) с точки зрения как особенностей объекта и его поведения, так и фиксации условий, в которых наблюдения проводились, а также сохранность данных наблюдения в неизменном виде. Ведь наблюдения всегда проводятся конкретным субъектом в конкретном месте, времени, при определенных условиях с помощью конкретных средств, включая естественные (органы чувств человека) и искусственные (технические). Необходимыми условиями объективности (но не достаточными, не гарантирующими истинность данных) наблюдения являются независимость его результатов (данных) от индивидуальных особенностей субъекта-наблюдателя и существование возможности его выполнения другими субъектами в тех же условиях. В любом случае данные наблюдений требуют всесторонней проверки на достоверность и предварительной обработки, в частности очищения от ошибок, выбросов, субъективных наслоений и т. п. Поскольку первичные данные могут содержать не соответствующие исследуемой действительности сведения, применять утонченные методы обработки и анализа первичных данных и моделирования по ним наблюдаемого объекта, как правило, не только напрасная трата времени, но и опасная возможность убедить себя и других в повышенной точности, достоверности и объективности получаемых с использованием таких методов закономерностей, выводов, искомым результатов, которые очищенные («истинные») данные могут не содержать. Во многом это относится к эксперименту.

3. Что касается эксперимента, то, как следует из его определения, в отличие от наблюдения он:

- а) является активным;
- б) может проводиться в естественных и искусственных (даже в таких, которые в реальности нежелательны или невозможны) условиях;
- в) допускает многократное повторение;
- г) может осуществляться с исследуемым объектом или его моделью.

Из этого следуют многие достоинства эксперимента по сравнению с наблюдением, часть из которых будет понятна из излагаемого далее

материала, а часть будет рассмотрена на практических занятиях. Здесь лишь перечислим некоторые из них. Это возможности обнаружения таких свойств объекта и связанных с ним явлений, которые не наблюдаются в естественных условиях; рассмотрения объекта (явления) в «чистом виде» путем его изоляции от ненужных обстоятельств и условий, контроля за поведением объекта.

§ 3.3. РАЗНОВИДНОСТИ ЭКСПЕРИМЕНТОВ

Вначале рассмотрим разновидности экспериментов, увязав их с используемыми для представления их результатов моделями (см. также табл. 2.1–2.4).

По назначению эксперименты с неким объектом, модель которого мы строим, можно разделить на *проверочные* (им соответствуют *априорные* (доопытные) модели объекта) и *исходные* (апостериорные модели объекта). В первом случае модель объекта (например, гипотеза) задается (известна, выбирается, строится) до эксперимента и назначение эксперимента – проверить или опровергнуть ее, т. е. выявить, соответствует ли она объекту и насколько адекватно, точно, а также, при необходимости, определить и/или уточнить ее параметры. Во втором случае именно Данные, получаемые в ходе эксперимента, являются исходными для построения модели и выполнения с нею последующих действий. Отсюда и ее название – *апостериорная*, послеопытная, полученная на базе опытных Данных. Здесь лишь перечислим некоторые из них. Это действия, связанные с обнаружением не наблюдаемых в естественных условиях свойств объекта и связанных с ним явлений; с рассмотрением объекта (явления) в «чистом виде» путем его изоляции от ненужных обстоятельств и условий, контролем за поведением объекта и т. п. В свою очередь исходные эксперименты можно детализировать на поисковые, конструктивные, контрольные и т. п. (см. рис. 1.1 и табл. 2.1 – назначение и функции моделей). *Поисковые* – это эксперименты, нацеленные на обнаружение новых особенностей, свойств объекта, связанных с ним новых явлений или ранее неизвестных связей между ними. *Конструктивные эксперименты* – предназначены для создания новых объектов (устройств, веществ, ...), а *контрольные* – ориентированы на поверку и отладку измерительных приборов и других средств. Понятно, каждый из подобных экспериментов может быть качественным (категорийным) либо количественным. *Качественный* эксперимент имеет цель установить, подтвердить или опровергнуть

что-то (гипотезу, теорию, свойство, явление, ...), т. е. дает бинарные ответы на поставленные вопросы типа «да» или «нет», «противоречит» – «не противоречит» (модели, гипотезе, теории, факту, ...), либо сделать вывод о недостатке данных для категоричного ответа. Цель же **количественного** эксперимента – выявление числового (количественного) свойства объекта (предмета, явления, процесса, сигнала, ...).

В связи с изложенным уместно заметить, что именно несогласие с результатом корректно поставленного и проведенного эксперимента является критерием ошибочности исследовательской модели (гипотезы, физической теории, картины мира). Особенно это касается физических гипотез и теорий, для которых главным критерием их истинности служит подтверждение экспериментом. Однако обратное утверждение не всегда верно: согласие с экспериментом не может быть (единственным) доказательством «правильности» и применимости модели (гипотезы, теории). Необходимо учитывать еще массу других аспектов, в частности условия применимости моделей.

Если физические носители оригинала и модели одной природы, эксперименты и модели называют **идентичными**, иначе **аналоговыми**, **машинными**. В том случае, когда экспериментальные Данные являются единственным источником исходных Данных, эксперименты и построенные по ним модели называются **эмпирическими**, если же единственным источником исходных данных является разум, – **рационалистическими**.

Если модели строятся по результатам эксперимента, т. е. являются вторичными, пассивными по отношению к нему, то такие модели и эксперименты называются **отражательными** в отличие от ситуации, когда исходные модели являются имитационными, ориентированными на подражание оригиналу и используются как исходные для постановки соответствующего эксперимента, называемого **имитационным**.

Реальные – это эксперименты, проводимые на самом деле, в реальном действии, в отличие от **мысленных**:

а) придуманных, воображаемых, не имеющих реальной реализации или не встречаемых в действительности;

б) связанные с мысленным извлечением выборки из генеральной совокупности (см. понятия математической статистики).

По виду объектов, с которыми проводится эксперимент, их можно разделить на **натурные** (физические), проводимые с самими существующими в действительности объектами, и **машинными** (в частности, **компьютерными**, проводимыми на ЭВМ), выполняемыми

с моделями объектов. Последние в свою очередь можно разделить на вычислительные (численные) и имитационные. *Вычислительные* проводятся на ЭВМ над математической моделью объекта и состоят в том, что по параметрической модели вычисляются ее другие параметры и решения, их исследуют на устойчивость, выявляют предельные случаи и т. д. На основе полученных результатов делаются выводы об объекте, его особенностях, свойствах, возможностях. В *имитационных* экспериментах по модели или по ранее полученным данным наблюдений или экспериментов имитируют особенности объекта, его строение, поведение, свойства, механизмы функционирования, предельные свойства, ограничения и т. д. Заметим, что зачастую результаты вычислительных и имитационных экспериментов могут быть несовпадающими, а в ряде случаев (условий, ситуаций, объектов) расходиться с результатами натуральных экспериментов.

В свою очередь натуральный эксперимент как инструментарий познания природы по своей направленности может быть *физическим* (его объекты и средства есть объекты физики), *химическим, биологическим, психологическим, социальным* и т. п. Заметим, что зачастую для упрощения термины *натурный* и *физический* используются как синонимы, т. е. термин *физический* обобщает понятия собственно физический, а также химический, биологический и им подобные эксперименты, чтобы отличить их от модельных. Добавим к изложенному, что иногда в физике, философии и других науках (см., например, теорию относительности) используется *мысленный* эксперимент, в котором структура, план, ход, условия, ... натурального (реально реализуемого) эксперимента воспроизводятся в воображении.

В заключение выделим еще три разновидности проверочных экспериментов, называемых критериальными, контрольными и критическими.

Критериальные – это эксперименты, выступающие как эталонные, образцовые для других (например, эксперименты при проверке измерительных приборов). *Контрольные* – параллельно проводимые с неизменными объектами и/или в неизменных условиях, когда исследуемые объекты и/или условия их функционирования и экспериментирования изменены и исследуются именно эти изменения, их наличие, вид, ход, динамика, влияние на объект и результаты измерения и т. п. Они часто используются, например, в медицине, педагогике. *Критические* – это эксперименты, исходы которых заведомо однозначно определяют, является ли конкретная модель (гипотеза, теория,

вывод, рекомендация) верной. Такие эксперименты должны однозначно давать предсказательные результаты, которые не могут быть выведены из других общепринятых моделей.

Наконец, **однородные** эксперименты (см. далее однородные Данные) – это проводимые с одним и тем же объектом в тех же условиях для реальных экспериментов или связанные с одной и той же генеральной совокупностью для мысленных экспериментов. **Неоднородные** – противоположность однородным, не относящиеся к ним.

До сих пор мы эксперимент увязывали в основном лишь с исследуемым объектом и частично – со средой через условия экспериментирования и «жизни» объекта. Теперь увяжем между собой три системных элемента: моделируемый объект-оригинал, эксперимент с ним и модель объекта и/или экспериментальных Данных. Для этого вернемся к рис. 2.1. Чтобы увязать эти три системных элемента, необходимо в рис. 2.1 добавить снизу между объектом и средствами измерений блок «средства планирования и организации эксперимента (наблюдения), в частности измерения» или считать их включенными в средства измерений.

Вернемся к рис. 1.1 и, используя системный подход, увяжем между собой все четыре системных элемента: моделируемый объект–модель–субъект–среда с учетом того, что между объектом-оригиналом и моделью необходимо поставить условия функционирования объекта, эксперимент и условия его проведения, а также первичные результаты эксперимента – *экспериментальные Данные*. В первом рассмотрении из элементов среды будем учитывать только условия функционирования и проведения эксперимента и получения экспериментальных Данных, а затем – полученные при этом разновидности экспериментальных Данных.

Необходимость такой увязки объекта-оригинала, модели, субъекта и среды связана с тем, что тип используемой для моделирования объекта-оригинала модели зависит не только от цели, задач, требований и других атрибутов моделирования, определяемых (назначаемых) субъектом, но и от природы объекта-оригинала, условий его функционирования («жизни»), назначения эксперимента, условий его проведения, типа и качества средств, используемых при экспериментировании для получения экспериментальных Данных, а также для их обработки, использования, анализа и других факторов (см. понятие *среда* в § 1.1). Например, в квантовом мире наличие наблюдателя (субъекта или прибора) приводит к изменению состояния объекта и условий его «жизни».

Все это означает, что между объектом-оригиналом и моделью на рис. 1.1 необходимо условно поместить подсистему (блок) эксперимент–экспериментальные Данные–среда, а именно ту часть среды, которая влияет на объект-оригинал и результаты экспериментирования. Ведь именно по экспериментальным Данным будет моделироваться объект-оригинал, т. е. строиться модель и проверяться соответствие ее и полученных по ней модельных результатов объекту-оригиналу. Иными словами, именно Данные влияют на результат моделирования и, что наиболее важно, на его интерпретацию, выявления по нему особенностей именно объекта-оригинала и его поведения за вычетом тех нюансов, которые внесет в моделирование эксперимент. Вполне может оказаться, что тип адекватной модели объекта-оригинала будет отличаться от типа той модели, которая окажется более адекватной для **совокупного объекта** моделирования, представляющего собой объект-оригинал + эксперимент + экспериментальные Данные + среда. Это означает, что на рис. 2.1 после средств экспериментирования и измерения также надо поставить блок «средства получения, хранения, передачи, ... экспериментальных Данных», а все элементы рисунка окружить «средой» и включить субъекта.

Внимание! Попробуйте сделать эту работу самостоятельно.

Пам'ятую об этом, рассмотрим возможные с точки зрения модельного представления разновидности объектов как совокупных объектов либо отдельно объектов-оригиналов и экспериментов.

В компактном виде их разновидность, а также применяемые для их представления классы моделей представлены на рис. 3.3.

Что касается экспериментальных данных, то они будут отражать не только вид объекта и/или эксперимента, но и те элементы среды, которые сопутствуют эксперименту, в частности используемые измерительные шкалы, предшествующие эксперименту модельные представления, включая и априорные знания (см. рис. 2.1) об объекте, условиях его функционирования и экспериментирования.

Фрагментарные представления о разновидностях экспериментальных материалов дает рис. 3.4. Поясним его.

Согласно рис. 2.1 **точечные** и **интервальные** данные отражают вариант представления результатов измерения, считывания, съема, использования значений физических или математических величин в следующем виде: один результат – одно значение (точка) величины (точечные данные) или два (нижнее /левое/ и верхнее /правое/) интервала (диапазона) значений величины (интервальные данные).

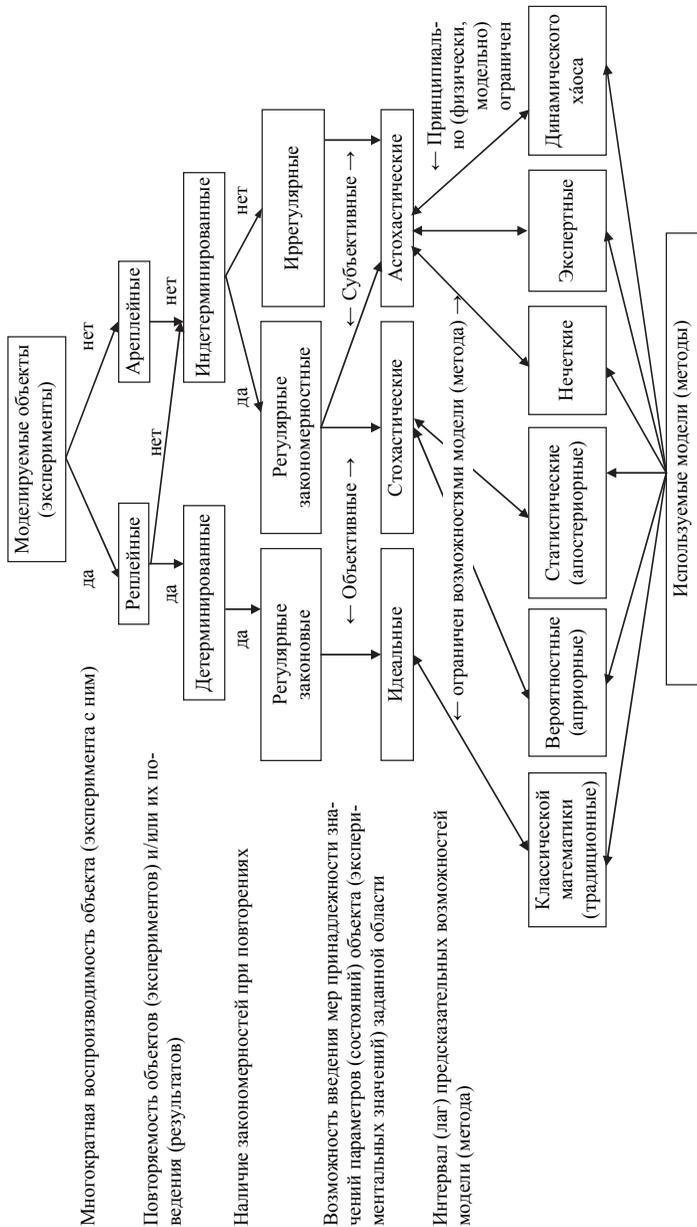


Рис. 3.3. Разновидность моделируемых объектов, экспериментов и моделей (методов)

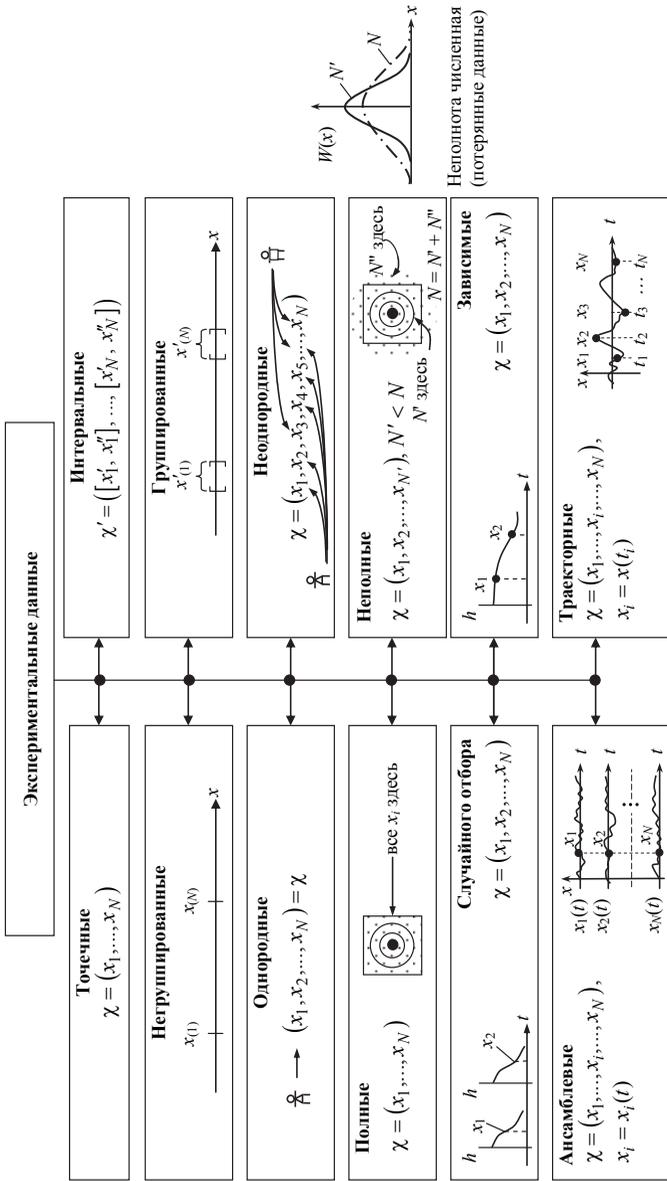


Рис. 3.4. Виды экспериментальных данных

Примеры точечных данных: длина стола – 1,2 м, вес гири – 16 кг и т. д. Примеры интервальных данных: длина стола от 1,1 до 1,3 м, т. е. $(1,1 \dots 1,3)$ м или $[1,1 \text{ м}; 1,3 \text{ м}]$, рост человека $(1,7 \dots 1,8)$ м или $[1,7 \text{ м}; 1,8 \text{ м}]$. Из множества причин введения интервального представления данных рассмотрим две. Первая – учет погрешности измерения физических величин. Например, результат d измерения длины стола 1,2 м получен с погрешностью в пределах ± 1 см. Тогда его можно представить интервально в виде $d = (1,2 \pm 0,01)$ м, $d = (1,19 \dots 1,21)$ м или $d \in [1,19 \dots 1,21]$ м. Вторая причина. Нам надо знать рост конкретного человека, чтобы сведения о его росте были в анкетных данных. Измеряем его и получаем 1,78 м. Это точечный результат. Учитывая условия измерения и класс используемых средств измерения, мы можем учесть пределы погрешности измерения и представить результат интервально, например, в виде $[1,76 \text{ м}; 1,80 \text{ м}]$. Для уменьшения статистических погрешностей измерения можно сразу же произвести несколько измерений и усреднить их результаты, получив тот же или близкий к нему средний точечный результат 1,78 м и интервальный, например, $[1,77 \text{ м}; 1,79 \text{ м}]$. Однако если мы хотим учесть, что физическая величина «рост человека» (как и пульс, давление крови и другие показатели) колеблется в течение суток, зависит от нагрузки и других факторов, нам следует сразу представить его интервально в виде двух граничных (минимального и максимального при исследуемых условиях) значений, которые с учетом погрешностей измерения равны $1,75 \pm 0,01$ и $1,8 \pm 0,01$ м. Тогда мы рост этого человека можем представить интервально в виде $[1,75 \text{ м}; 1,8 \text{ м}]$ с погрешностью $\pm 0,01$ на обоих концах интервала либо как $[1,74 \text{ м}; 1,81 \text{ м}]$. Какое из этих интервальных значений выбрать, определяется целью измерения, назначением их результатов и используемыми при этом операциями с ними (см. прил. 1).

Негруппированные данные – это точечные числовые значения, фактически полученные в данном конкретном эксперименте, например, $\chi = (x_1, x_2, \dots, x_N)$. Напомним, что их ранжированное по возрастанию или убыванию значений расположение называется *вариационным рядом*. Пример такого ряда $x_{(1)}, x_{(2)}, \dots, x_{(N)}$, где $x_{(1)}$ – 1-й элемент вариационного ряда, имеющий наименьшее значение из x_1, x_2, \dots, x_N ; $x_{(2)}$ – следующее значение и т. д. **Группированные** – это данные, полученные разбиением всего диапазона возможных или полученных

значений x_1, \dots, x_N на заранее заданное число s одинаковых или разных по ширине (*шагу группирования*) интервалов группирования (рис. 3.4). Например, если $x_{\min} = x_{(1)}$ и $x_{\max} = x_{(N)}$ – наименьшее и наибольшее значения из x_1, \dots, x_N , то при равномерном шаге группирования $\Delta x = (x_{\max} - x_{\min})/s = (x_{(N)} - x_{(1)})/s$ группированные данные – это набор полуинтервалов $x'_{[k]} = [x'_{(k)}; x'_{(k)} + \Delta x)$, где $x'_{(k)} = x_{\min} + (k-1)\Delta x = x'_{(k-1)} + \Delta x$, $k = 1, 2, \dots, s = \overline{1, s}$. Отсюда следует, что каждому x_i , $i = \overline{1, N}$, из множества χ , т. е. из x_1, x_2, \dots, x_N , в группированных данных соответствует свой k -й полуинтервал $x'_{[k_i]} = [x'_{(k_i)}; x'_{(k_i)} + \Delta x)$, $k = \overline{1, s}$, в который попадает x_i , т. е. для которого справедливо неравенство $x'_{(k_i)} \leq x_i < x'_{(k_i+1)} = x'_{(k_i)} + \Delta x$. Заметим, что группированные данные в операциях с ними представляются либо своими интервалами, как, например, при построении гистограмм, либо наименьшим (левым) $x'_{(k)}$, средним $\bar{x}'_k = x'_{(k)} + \Delta x/2$ или наибольшим (правым) $x'_{(k)} + \Delta x$ значением каждого k -го интервала. Зачастую представление интервала его средним значением $\bar{x}'_{[k_i]}$ при этом предпочтительнее с точки зрения влияния группирования на дополнительную погрешность результатов, получаемых при замене x_1, \dots, x_N на их группированные аналоги. Это означает, что группированные данные, соответствующие набору (выборке) точечных данных $\chi = (x_1, x_2, \dots, x_N)$, представляют собой либо интервальный набор $\chi' = (x'_{[k_1]}, x'_{[k_2]}, \dots, x'_{[k_N]})$, либо набор типа $\bar{x}'_{[k_1]}, \bar{x}'_{[k_2]}, \dots, \bar{x}'_{[k_N]}$, где $\bar{x}'_{[k_i]}$ – середина k_i -го интервала, $k_i = \overline{1, s}$.

Заметим, что группирование может быть как специально организованной процедурой при обработке и анализе данных (например, при построении гистограмм), так и следствием алгоритмических операций, выполняемых при их сборе, обработке, анализе (например, при аналого-цифровом преобразовании отсчетов сигналов в цифровые данные, при округлении результатов вычислений).

Как уже упоминалось, **однородные** – это данные, полученные для одного объекта в тех же или приемлемо для решаемой задачи сходных условиях. На языке математической статистики – это выборочные данные, полученные из одной и той же генеральной совокупности (ГС). **Неоднородные** данные отличаются от однородных либо тем, что среди них есть такие, которые относятся к другим объектам, либо такие, которые получены в других условиях, не относящихся к условиям решаемой задачи. Например, это выборочные данные из ГС либо с разными законами распределения, либо с одним и тем же по наименованию распределением, но с разными значениями его параметров положения, масштаба, формы, связи, или с разными другими характеристиками. Это иллюстрирует соответствующий фрагмент рис. 3.4.

Рассматривая экспериментальные Данные, весьма важно быть уверенным в их *полноте* или, наоборот, в наличии неполноты в конкретном для решаемой задачи смысле. Понятие о полноте данных будет рассмотрено в § 3.3.

Следующим очень важным качеством экспериментальных данных является наличие в них *зависимости* между их элементами. Это связано с тем, что при решении одних теоретических и практических задач усилия исследователя направляются именно на обнаружение, идентификацию и использование зависимостей в данных, отражающих соответствующие зависимости в моделируемом (исследуемом) объекте-оригинале или влияние на него либо зависимость от него факторов, показателей среды. В других же задачах стараются избавиться от зависимости в Данные, поскольку ее наличие уменьшает количество полезной для пользователя информации по сравнению с независимыми (случайными) Данными о том же объекте. Использование Данные, полученных в условиях, когда их элементы зависимы, может привести к существенному увеличению статистических погрешностей результатов измерений и статистического анализа. Не останавливаясь на этом подробнее, заметим, что в каждом классе моделей объектов, построенных по экспериментальным данным (см. рис. 3.3), вводятся свои понятия и меры зависимостей (например, функциональная, статистическая, экспертная). Некоторые из них будут рассмотрены при описании подобных моделей (см. главу 4). Здесь лишь отметим два обстоятельства. Первое касается понятия *случайного отбора* данных из ГС (не всегда удачно их порою называют случайными данными, см. подпараграфы 4.3.7 и 4.3.8). Это такой рандомизированный (ослучаенный) отбор

данных в выборку из ГС, когда каждый элемент ГС может быть включен в выборку с одинаковой вероятностью, т. е. имеет одинаковый с другими элементами ГС шанс быть отобранным в данном опыте эксперимента. Если это имеет место быть, то тем самым обеспечивается независимость (в вероятностно-статистическом смысле) элементов подобной случайной выборки x_1, x_2, \dots, x_N (см. рис. 3.4) в отличие от зависимой выборки, когда x_1, x_2, \dots, x_N есть реализации зависимых случайных величин X_1, X_2, \dots, X_N (см. § 4.3).

Второе обстоятельство следующее. Обратим внимание на то, как на рис. 3.4 качественно представлены варианты случайного и зависимого отбора значений $x_1 = h(t_1)$, $x_2 = h(t_2)$, ..., $x_N = h(t_N)$ некоторого развивающегося во времени t показателя h состояния некоего объекта. Предположим, что $h(t)$ отражает реакцию человека на какое-то воздействие (например, на лекарство). Исследователь хочет построить модель (аналитическую или графическую) такой зависимости путем измерения N значений $x_i = h(t_i)$, $i = \overline{1, N}$, для разных t_1, \dots, t_N из диапазона от начала приема лекарства до фактического вывода его из организма. *Случайный отбор* означает, что исследователь получает каждое x_i для того же пациента в сходных условиях (чтобы выборка была однородной), после того как лекарство предыдущего приема уже выведено из его организма, а *зависимый* – когда N измерений x_i проводятся по ходу усвоения и вывода лекарства при однократном приеме. Речь идет о качественной, а не количественной иллюстрации, т. е. без учета особенностей накопления лекарства в организме, опосредованного его влияния со временем и т. д. Рисунок призван лишь подчеркнуть, что если $h(t)$ имеет детерминированное описание, то x_1, x_2, \dots, x_N , полученные по зависимой выборке (без учета погрешностей измерения), конечно, окажутся связанными этой зависимостью в отличие от x_1, x_2, \dots, x_N , в которых зависимость не проявляется.

Наконец, поясним отличие между ансамблевыми и траекторными данными (рис. 3.4). *Ансамблевые* данные получаются по ансамблю (набору, совокупности) однородных объектов. Например, в виде кардиограмм пациентов со сходными патологиями; изменяющихся во времени одинаковых показателей x_1, \dots, x_N ; поставленных на исследование N процессоров в один и тот же момент времени t , т. е. в виде

$x_i = x_i(t)$, $i = \overline{1, N}$ (см. левую нижнюю картинку на рис. 3.4). *Траекторные* же данные получаются путем длительного наблюдения или экспериментирования с одним и тем же объектом как отсчеты показателя $x(t)$, получаемые в последовательности временных отсчетов, т. е. как $x_i = x(t_i)$ (см. правую нижнюю картинку рис. 3.4).

§ 3.4. ПОНЯТИЕ О НЕПОЛНОТЕ АПРИОРНЫХ И АПОСТЕРИОРНЫХ (ЭКСПЕРИМЕНТАЛЬНЫХ) ДАННЫХ О МОДЕЛИРУЕМОМ ОБЪЕКТЕ

Одним из неоднозначно понимаемых, трактуемых, интерпретируемых в информатике является термин *неполнота данных* об исследуемом, моделируемом, проектируемом, описываемом... субъекте и объекте. Приведем один из вариантов многоуровневой трактовки термина.

Неточность Данных. Этот вид неполноты данных об объекте¹ связан с погрешностями измерения значений элементов данных, ошибками в значениях, выборе, включении данных в выборку $\chi = (x_1, \dots, x_N)$.

Неопределенность – неуверенность субъекта в истинности Данных, обусловленной, в частности, отсутствием сопровождения Данных описанием технологий, средств, условий их получения, компетентностью и предвзятостью лиц, участвующих в этом.

Нечеткость (расплывчатость) – неполнота, обусловленная субъективным восприятием принадлежности элементов Данных к соответствующим элементам измерительной, прежде всего категорической, шкалы. Например, субъективностью восприятия понятий «холодно», «тепло», «большой», «малый».

Незнание – вид неполноты, являющейся следствием отсутствия априорных сведений, модельных представлений об объекте, о его показателях, характеристиках и/или о технологиях и об алгоритмах обработки и анализа Данных о нем. Например, отсутствие представлений о геометрическом представлении объекта, площадь или объем которого субъект хочет измерить, отсутствие сведений о сильной сердечной аритмии пациента при измерении его пульса и/или давления большин-

¹ В дальнейшем слова «об объекте» будем опускать.

ством автоматических измерителей артериального давления, основанных на усреднении.

Нерепрезентативность – вид неполноты, означающий непредставительность, недостаточность Данных для верного (адекватного) целям, условиям, задачам, технологиям и пр.) представления ими объекта при решении по ним субъектом конкретной задачи. Причинами нерепрезентативности (непредставительности) Данных могут быть их малый объем, частичная потеря; пропуски элементов Данных; наличие в них искажающих истинное положение дел выбросов, помех, промахов. На рис 3.4 показан пример нерепрезентативности данных, отражающих исследование качества некоторого оружия по расстоянию от центра попадания пули до центра мишени без соответствующих априорных сведений о стрельбе. На рисунке представлен случай, когда по мишени было произведено N выстрелов, а в ней оказалось только N' отметок от попаданий. Остальные $N - N'$ это «попадание в молоко», т. е. вне мишени. К чему это может привести? Если субъект – исследователь качества оружия – не знает, что было произведено N выстрелов, то он по мишени построит график плотности распределения расстояния между местом попадания пули и центром мишени, отмеченный сплошной линией и символом N' , в то время как реальный график должен иметь вид, отмеченный штрихпунктирной линией и символом N . В то же время, если бы субъект знал, что $N'' = N - N'$ пуль прошли мимо мишени, он построил бы более адекватно отражающую ситуацию плотность усеченного распределения, график которой был бы представлен точками и символом N'' , повторяющими по виду часть графика N , попадающую в зону мишени, с ординатами в N/N' раз большими, чем в кривой N . Нарисуйте график самостоятельно.

Наконец, еще один вид неполноты – это **неучтенность**. Он связан с тем, что субъект не учитывает при моделировании и экспериментировании влияние на объект внешнего окружения, его внутренней среды, условий функционирования объекта и экспериментирования с ним, важные внешние и внутренние влияющие и мешающие факторы и пр.

Обратим внимание на неоднородность толкования некоторых из приведенных терминов. Рассмотрим, например, как термин «неопределенность» трактуется в метрологии – науке об измерениях.

§ 3.5. ИЗМЕРЕНИЯ КАК ЭТАП НАБЛЮДЕНИЯ, НАТУРНОГО ЭКСПЕРИМЕНТИРОВАНИЯ И МОДЕЛИРОВАНИЯ

При работе с материальными объектами, в том числе при их наблюдениях, экспериментировании с ними и их моделировании (см. рис. 2.1), в тех случаях, когда субъекту приходится иметь дело с физическими величинами, он часто вынужден прибегать к количественным измерениям. Их цель – получение первичных данных как носителей информации об объекте в виде количества единиц измерения, укладываемых в интересующей субъекта физической величине – измеряемой величине. Приведем здесь нужные для дисциплины элементарные понятия из метрологии как науке об измерениях, из специальной литературы, в частности международного руководства [17].

Основные нужные нам термины из руководства приведены в табл. 3.4. Прокомментируем их и одновременно некоторые ранее изложенные понятия, в том числе понятие полноты.

1. Прежде всего отметим, что из множества проблем измерения самые главные следующие три из них. Все они связаны с назначением измерения – получить и представить данные результаты измерения физической (измеряемой) величины так, чтобы не потерять и не исказить информацию, которую измеряемая величина несет об объекте и о влияющих на него условиях существования в момент и в месте измерения. Ведь результат измерения не истина, а всего лишь модель, оценка, аппроксимация настоящего значения измеряемой величины.

Т а б л и ц а 3.4

Основные метрологические термины [17]

Термин	Толкование термина
Измерение	Совокупность операций, имеющих целью определение значения (физической, измеряемой) величины
Измеряемая величина (физическая величина)¹	Свойство явления, объекта или вещества, которое может выделяться качественно и определяться количественно
Однородные физические величины	Величины, которые могут быть сгруппированы по категориям: а) работа, теплота, энергия; б) толщина, длина окружности, длина волны

Продолжение табл. 3.4

Термин	Толкование термина
Значение (физической, измеряемой)² величины	Значение конкретной величины, выражаемое, как правило, произведением единицы измерения на число (длина стержня 5,34 м)
Истинное значение величины	Значение, соответствующее определению данной конкретной величины ³
Действительное значение величины (приписанное, номинальное, наилучшая оценка величины)	Значение, <i>приписываемое</i> конкретной величине и <i>принимаемое</i> , часто по соглашению, как имеющее неопределенность, приемлемую для данной цели
Принцип измерения	Научная основа измерения (например, применения для измерений некоторых физических эффектов)
Метод измерения	Логическая последовательность операций, описанная в общем виде, которая выполняется при измерениях: метод измерения замещением, дифференциальный метод, нулевой метод
Модель измерения	Правило преобразования численных значений влияющих X_1, X_2, \dots, X_n и дополнительных величин ⁴ в значение измеряемой величины Y . Данные для нее – значения входных X_1, X_2, \dots, X_n и выходной величины Y в ней – обычно представляют в виде детерминированной функциональной связи $Y = f(X_1, X_2, \dots, X_n)$
Измерительная процедура (методика выполнения измерений)	Специально описанная совокупность операций, используемая при выполнении конкретных измерений в соответствии с данным методом
Измеряемая физическая величина	Конкретная величина, подвергаемая измерению (иногда требуется указывать конкретные условия, при которых происходит измерение: температура, давление, влажность и т. п.)
Влияющая величина	Величина, которая не является предметом измерения, но влияет на результат измерения (например, температура микрометра, применяемого для измерения длины)
Результат измерения	Полученное путем измерения значение, приписываемое измеряемой величине ⁵
Неисправленный результат измерения	Результат измерения до введения поправки на систематическую погрешность

Продолжение табл. 3.4

Термин	Толкование термина
Исправленный результат измерения	Результат после введения поправки на систематическую погрешность
Точность измерения	Качественное понятие, характеризующее близость результата измерения к истинному значению измеряемой величины
Сходимость результатов измерения	Близость результатов последовательных измерений одной и той же измеряемой величины, выполненных в одинаковых условиях измерений – условиях сходимости: одни и те же измерительные процедуры, наблюдатель, измерительный прибор, применяемый в одних и тех же условиях, место, а также повторение измерений в течение короткого (по сравнению с динамикой изменения величины, условий) периода времени
Воспроизводимость результатов измерений	Близость результатов измерений одной и той же измеряемой величины при проведении измерений в измененных условиях (изменение принципа измерения, метода, наблюдателя, прибора, эталона, места, условий применения, времени)
Экспериментальное стандартное отклонение	<p>Величина S_i, $i = \overline{1, n}$ для ряда N измерений одной и той же измеряемой величины ξ_i, характеризующая рассеяние результатов и определяемая по формуле</p> $S_i = \sqrt{\sum_{j=1}^N (x_{i,j} - \bar{x}_i)^2 / (N - 1)},$ <p>где $x_{i,j}$ – результат j-го измерения ξ_i;</p> $\bar{x}_i = \frac{1}{N} \sum_{j=1}^N x_{i,j}$ <p>– среднее арифметическое из N рассматриваемых результатов $x_{i,j}$, $j = \overline{1, N}$, измерения ξ_i</p>
Неопределенность измерения	Параметр, связанный с результатом измерений и характеризующий рассеяние значений, которые достаточно обоснованно могли бы быть приписаны измеряемой величине (например, стандартное отклонение; число, кратное ему; половина доверительного интервала)

Продолжение табл. 3.4

Термин	Толкование термина
Погрешность⁶ измерения	Отклонение результата измерения от истинного значения измеряемой величины ⁷
Промаях (грубая погрешность)	Погрешность, возникающая вследствие недосмотра экспериментатора или неисправности аппаратуры
Относительная погрешность измерения	Отношение (абсолютной) погрешности измерения к истинному (действительному) значению измеряемой величины
Приведенная погрешность	Отношение абсолютной погрешности к условно принятому из диапазона измерений значению
Случайная погрешность⁸	Разность результата измерения и среднего значения, которое могло бы быть получено при бесконечно большом числе повторных измерений одной и той же измеряемой величины, проводимых в условиях сходимости
Систематическая погрешность измерения⁹	Разность между средним значением, полученным при бесконечном числе измерений одной и той же измеряемой величины в условиях сходимости, и истинным значением измеряемой величины
Поправка	Значение величины, которое алгебраически суммируется с неисправленным результатом измерения для компенсации (аддитивной) систематической погрешности
Поправочный коэффициент	Числовой коэффициент, на который умножают неисправленный результат измерения для компенсации (мультипликативной) систематической погрешности
Вероятность случайного события¹⁰ (P, P, Pr)	Действительное число в интервале от 0 до 1, приписываемое случайному событию (и используемое как численная мера возможности появления события в опыте, эксперименте)
Случайная переменная (физическая) величина	Переменная величина, которая может принимать любое значение из указанного ряда величин и с которой связано распределение вероятностей
Распределение вероятностей (случайной переменной величины X)	Функция, определяющая вероятность того, что случайная величина X принимает любое заданное значение x или принадлежит к заданному ряду (или области) значений

Продолжение табл. 3.4

Термин	Толкование термина
Функция распределения (вероятностей)	Функция $F(x)$, определяющая для каждого значения x вероятность того, что случайная величина X меньше или равна x : $F(x) = \mathbf{P}\{X \leq x\} = \mathbf{P}(X \leq x) = \Pr(X \leq x)$
Функция плотности вероятности (плотность распределения вероятностей) для непрерывной случайной величины	Производная (если она существует) функции распределения $W(x) = dF(x)/dx$; $W(x)dx = \mathbf{P}\{x < X \leq x + dx\}$ ($dx > 0$)
Частота	Число случаев N_i данного типа событий или число наблюдений, попадающих в i -ю определенную группу
Относительная частота (частость)	Отношение числа случаев (наблюдений) N_s попадания в s -ю группу к общему числу N наблюдений: $p_s^* = N_s/N$
Статистика	Функция $f(x_1, x_2, \dots, x_N)$ элементов x_1, \dots, x_N выборки (из генеральной совокупности) случайной переменной (X)
Оценивание^П	Операция приписывания, на основании наблюдений в выборке, числовых значений параметрам распределения, выбранного в качестве статистической модели совокупности, из которой взята эта выборка
Оценка^П	Статистика $\hat{\theta}$, используемая для оценивания параметра θ совокупности
Значение оценки^П	(Конкретное) значение статистики $\hat{\theta}$, полученное в результате оценивания
Доверительный интервал	Интервал $(\hat{\theta}_1, \hat{\theta}_2)$, к которому принадлежит значение оцениваемого параметра θ с вероятностью $(1 - \alpha)$, где α – фиксированное неотрицательное число меньше единицы, называемое уровнем значимости
Коэффициент доверия; доверительный уровень	Значение $(1 - \alpha)$ вероятности, связанное с доверительным интервалом или статистическим интервалом охвата

Термин	Толкование термина
Статистический интервал охвата	Интервал, для которого можно с заданным доверительным уровнем констатировать, что он включает по крайней мере определенную часть совокупности
Стандартная неопределенность	Неопределенность, выражаемая с использованием стандартного отклонения
Расширенная неопределенность	Неопределенность, выражаемая в виде интервала вокруг результата x_i измерения ξ_i , в котором, как ожидается, находится большая часть распределения значений, которые с соответствующим заданным уровнем доверия могут быть приписаны измеряемой величине

¹ По всей таблице заключенное в скобки, как правило, добавлено автором.

² Далее слово *физическая*, измеряемая в приложении к величине, будет опускаться.

³ В теоретической метрологии под истинным понимают значение (измеряемой) величины, которое могло бы быть получено при идеальном измерении. В силу природных причин истинное значение неопределимо. Поэтому в Руководстве [17] этот термин не используется, а термины «истинное значение (измеряемой) величины» и «значение (измеряемой) величины» рассматриваются как эквивалентные.

⁴ Дополнительные – это, например, данные о физических константах, каждая из которых известна недостаточной точно (модуль упругости, удельная теплоемкость и др.), и иные, взятые из справочников.

⁵ При приведении результата следует ясно указать:

1) к чему он относится: к показанию прибора без учета поправки, с учетом поправки, к среднему нескольких значений;

2) сведения о неопределенности (либо погрешности) измерения.

⁶ Иногда ее называют абсолютной погрешностью измерения, что не следует путать с абсолютным значением (модулем) погрешности.

⁷ Так как истинное значение не может быть определено, на практике вместо него применяется действительное значение.

⁸ Рекомендуем сравнить эти термины с терминами *случайная ошибка (погрешность) оценивания* и *смещение (систематическая ошибка)* оценки, используемых в статистике.

⁹ В силу необходимости проведения бесконечного числа измерений это понятие является лишь теоретическим. Практически систематическая погрешность не может быть точно известна и вместо нее пользуются ее оценкой.

¹⁰ Рекомендуем нижеприводимые термины сопоставить с определениями из теории вероятностей.

¹¹ См. табл. 3.1, а также [5] и п. 4.3.

Первая проблема как раз и связана с тем, как узнать и понять, что такое истинное значение измеряемой физической величины: существует ли оно (например, измеряемое время (момент времени), рост или вес человека и т. п.), единственно ли (т. е. представляет собой детерминированную величину, константу в данных условиях измерения) или само является индетерминированным объектом модельного представления (см. рис. 3.3). В подавляющем числе случаев при измерениях полагается, что истинное значение хотя и не известно, но существует, является единственным определимым, т. е. допускает его количественное выражение с помощью подходящих единиц измерения. А так ли это или всегда ли такие допущения приемлемы? Что если эта величина индетерминирована (см. § 4.4)?

Вторая проблема связана со сходимостью и воспроизводимостью результатов измерения, возможного четкого определения и сохранения либо изменения заданным образом условий измерения. Ведь все в мире динамично, нестабильно, изменяется, в том числе в процессе и с учетом процесса измерения, вмешательства его в существование и поведение объекта и условий его «жизни».

Наконец, третья важнейшая проблема: как наилучшим образом выразить то, что мы получаем в виде данных об измеряемой величине, т. е. как представить результаты измерения с учетом их качества, достоверности, доверия к ним. Как можно понять по табл. 3.4, в настоящее время применяются и обсуждаются два подхода к решению этой проблемы.

Первый подход давать наилучшую оценку измеряемой величины вместе со значениями систематической и случайной **погрешностей** результата измерения (см. табл. 3.4), т. е. с итогами анализа, расчета погрешностей, сопровождающих результат. Иными словами, результат измерения как объект моделирования считать стохастическим (см. рис. 3.1 и § 4.3). При этом скалярный результат измерения описывать случайной величиной, а совокупность результатов – случайным вектором (многомерной случайной величиной). Сама измеряемая величина считается постоянной, неслучайной, хотя в стохастическом подходе не только результат измерения, но и сама измеряемая величина может модельно представляться не константой, а случайной величиной. Это байесовские варианты применения стохастических моделей при измерениях.

Погрешностный подход был единственным до последних десятилетий ушедшего века подходом к решению третьей проблемы и, одно-

временно, к пониманию измерения и выражения воспринимаемого субъектом качества (достоверности) результата измерения.

Второй подход решения третьей проблемы – выражения результатов измерения появился в конце XX века и заключается в том, что результат измерения следует представлять как наилучшую оценку (в смысле, указанном в табл. 3.1, 3.4, а также аппроксимации, приближения) измеряемой величины вместе с соответствующей неопределенностью измерения. Что касается понимания и выражения неопределенности измерения, то согласно Руководству [17] допускается для его описания использовать как стохастический аппарат, так и другие модельные представления.

В чем принципиальное различие между первым и вторым подходами? В первом подходе субъект-измеритель признает объективность существования погрешностей, отражающих истинность результата измерения, и допускает возможность повторения и сходимости результата измерений, обеспечивающих обоснованность применения стохастического аппарата для их описания. Во втором подходе признаются:

а) возможность установить, насколько хорошо известно существенно единственное истинное значение измеряемой величины;

б) необходимость попытаться узнать, насколько велико доверие к тому, что субъекту известно об объекте и условиях измерения, о среде.

Тем самым признается, что неопределенность результата измерения отражает определенную неполноту знаний субъекта об измеряемой величине, окружающей среде, условиях ее измерения. При этом рассматриваются многие источники неопределенности, среди которых следующие [17]:

а) неполное определение измеряемой величины;

б) несовершенная реализация определения измеряемой величины;

в) нерепрезентативная выборка – измеренный образец может не представлять определяемую измеряемую величину;

г) неадекватное знание эффектов от условий окружающей среды, влияющих на измерение, или несовершенное измерение условий окружающей среды;

д) субъективная систематическая погрешность оператора при снятии показаний аналоговых приборов;

е) конечная разрешающая способность прибора или порог чувствительности;

ж) неточные значения, приписанные эталонам, используемым для измерения, и стандартным образцам веществ и материалов;

з) неточные значения констант и других параметров, полученных из внешних источников и используемых в алгоритме обработки данных;

и) аппроксимации и предположения, используемые в методе измерения и измерительной процедуре;

к) изменения в повторных наблюдениях измеряемой величины при явно одинаковых условиях.

2. Теперь обратим внимание на несколько трактовок одинаковых понятий, в частности приведенных в табл. 3.1 и 3.4. В качестве примера рассмотрим термин *оценка*.

Во-первых, заметим, что согласно табл. 3.1 в русском языке (в отличие от английского) термин *оценка* в зависимости от контекста понимается как синоним слов *оценивание* (estimation) – отражающее действие, процесс; как *статистика*¹ (estimator) – правило оценивания, функция наблюдений, элементов выборки (табл. 3.1 и 3.4) и как *результат оценивания*, его численное значение (estimate).

Во-вторых, очень часто термины *оценивание* и *оценка* какой-то величины, факта, явления, ... используются как процесс, процедура или как результат получения приближенного, аппроксимирующего представления (в том числе численного) о любой величине, факте, явлении, не обязательно описываемых стохастическими моделями. Следует различать толкования термина, ориентируясь на контекст, существо применения его.

В-третьих, рекомендуем сопоставлять трактовки термина *неопределенность* и иные, приведенные в начале § 3.3 и в табл. 3.4, найти сходство и различие в них с тем, чтобы не путаться с их пониманием в конкретной ситуации.

3. Погрешность есть идеализированное понятие. Она никогда не может быть известна точно.

4. Погрешность – не синоним неопределенности. Их нельзя путать и неправильно использовать. Дело в том, что если погрешность близка к нулю, то мы гарантированно считаем, что результат измерения близок к истинному значению измеряемой величины. В противном случае мы вынуждены констатировать, что результат измерения далек от истинного значения. Что касается неопределенности, то результат изме-

¹ Заметим, что в математической статистике под оценкой какой-то характеристики или параметра случайной величины, вектора, функции понимается не всякая статистика – функция элементов выборки, а лишь такая, которая дает несмещенное, состоятельное и, желательно, эффективное приближенное значение оцениваемой характеристики, параметра (см. § 4.3, а также [5, § 4.1]).

рения может быть близким к значению измеряемой величины, даже если он может иметь большую неопределенность (но не погрешность!), а в качестве значения измеряемой величины принимается ее действительное значение (см. табл. 3.4). Неопределенность измерения означает, что для конкретной измеряемой величины и конкретного результата ее измерения существует не одно, а бесконечное множество значений, рассеянных вокруг того результата, который, во-первых, лучше согласуется со всеми наблюдениями и данными, а также моделями физического мира; во-вторых, может быть приписан измеряемой величине с различной степенью уверенности.

5. Согласно [17] оценивание неопределенности результатов измерений может осуществляться двумя способами с учетом деления их составляющих на две категории: типа А и В. Неопределенность типа А – это неопределенность, оцениваемая по типу (способу) А статистическими методами. Неопределенность типа В – это неопределенность, оцениваемая по типу (способу) В нестатистическими методами. Тем не менее и там, и там в итоге оперируют вероятностными моделями и вероятностно-статистическими понятиями (табл. 3.4). Оба типа оценивания неопределенности основаны на распределении вероятностей. Отличие заключается в способе введения этих распределений. Для неопределенности типа А используют апостериорное статистическое распределение и известные статистические методы оценивания его параметров и характеристик. А именно:

а) среднеарифметическое для оценивания среднего, т. е. \bar{x} (табл. 3.4), или как оценки математического ожидания случайной величины X , используемой для апостериорной модельной замены физической величины ξ по ее выборочным значениям x_j ¹;

б) экспериментального стандартного отклонения S (табл. 3.4) в качестве оценки среднеквадратического (стандартного отклонения) случайной величины X . При этом чаще всего полагают, что X имеет нормальный закон распределения, а x_j рассматривается как элемент случайной выборки из генеральной совокупности объема N величины X . Для неопределенности же типа В используется априорная нестатистическая, т. е. полученная не апостериори, по итогам многократных измерений в условиях сходимости, а на базе априорных, доопытных научных заключений или другой априорной информации, получаемой

¹ Значением x_j модельно представляются результаты j -го, $j = \overline{1, N}$, измерения ξ .

по данным предшествующих или предварительных измерений; обобщения знаний об особенностях и свойствах объекта и измеряемой величины, спецификации изготовителя приборов, данных о поверке, калибровке, сертификатов используемых измерительных приборов, взятых из справочников, и т. п. При этом чаще всего в качестве распределения для x_j принимается равномерное в определенных границах. Значительно реже в обоих случаях применяются другие распределения, в частности асимметричные.

Как показывает специальное сравнение результатов измерений, найденных с помощью классического подхода, связанного с определением погрешностей, и с помощью подхода, основанного на концепции неопределенности, чаще всего они получаются одинаковыми или сходными. И это объяснимо, поскольку в обоих подходах используется вероятностно-статистический аппарат и сходные понятия. А именно погрешность результата измерения (РИ) заменена на неопределенность, случайная погрешность – на неопределенность типа А, неисключенная погрешность – на неопределенность типа В, среднеквадратическое отклонение (СКО) погрешности РИ – на стандартную неопределенность РИ, доверительные границы РИ – на расширенную неопределенность РИ, доверительная вероятность – на вероятность охвата (покрытия) РИ.

Заметим, что иногда термин *погрешность измерения* заменяют на термин *ошибка измерения*, оба понятия используют как синонимы. Однако согласно РМГ 29-99 «Рекомендации по межгосударственной сертификации. Основные термины и определения» термин *ошибка измерения* не рекомендуется применять как менее удачный. Наконец отметим, что согласно ГОСТ Р 50.2.038–2004 допускается использовать как *погрешность*, так и *неопределенность* для документов, использующихся в России.

6. В Руководстве [17] для экономии обозначений один и тот же символ используется как для обозначения самой физической величины (у нас символ греческий, например ξ), так и самой случайной переменной X , используемой для представления ξ и характеризующей множество результатов наблюдения (либо измерения) этой величины. Это широко используемый в литературе прием, который зачастую, к сожалению, приводит обучающихся к отождествлению объекта и его модели. Если обучаемый понимает разницу между этими понятиями и то, что ξ можно описать не только с помощью стохастических, но и других (см. рис. 3.3 и далее) моделей, такое совмещение допустимо. В связи с этим заметим, что когда исследователь переходит от измере-

ния к вычислительным операциям, т. е. к дальнейшей работе с математическими моделями физических величин (объектов), результаты измерений воспринимаются зачастую как «точные», исходные для дальнейших операций числа без их прикладного физического наполнения и лишь редко когда с учетом их погрешностей и неопределенностей, которые опять-таки воспринимаются как «точные» математические объекты без привязки к «точности», адекватности представления ими какого-то реального физического объекта. При этом обычно рассматривается лишь математическая сторона, т. е. не учитываются без необходимости физические размерности единиц измерения. К ним обращаются только при решении прикладных задач. Иными словами, рассматривается распределение либо уравнение математических, а не физических величин.

Ранее мы рассматривали, как представить результаты измерений, имеющих место в натуральных экспериментах. Что касается вычислительных и имитационных экспериментов, то там тоже имеет место подобная проблема. Однако она фактически сводится к задаче, для которой мы, в отличие от проблемы, знаем гарантированные методы решения. Здесь также используют термины *погрешность, абсолютная и относительная, приведенная*. Под погрешностью также понимается разность $\hat{a} - a$, где \hat{a} – найденное приближенное значение некоей величины, точное значение которой равно a и известно, например, из аналитического решения. В этих экспериментах, связанных с численными решениями математических задач, погрешность результата численного решения обуславливается приближенным характером математической модели представления реального объекта (погрешность математической модели), неточностью задания исходных данных (погрешность входных данных), неточностью метода решения и ошибками округления (погрешности метода и вычислительные). Зачастую погрешности модели и входных данных объединяют общим понятием *неустраняемая погрешность*. Напомним, что под *округлением числа* понимается приближенное представление числа в некоторой системе счисления с помощью конечного числа разрядов. Именно возникшую при этом погрешность называют погрешностью, или ошибкой, округления.

Заметим, что округление используется не только в исходных данных, но и при вычислении результатов измерений и вычислений. Общее правило, которым следует руководствоваться, заключается в том, что получаемые числовые данные и результаты надо записывать числом с сохранением одной сомнительной цифры, следующей за верной,

с указанием предельной абсолютной погрешности, выписывая ее с одной значащей цифрой, округляемой в сторону увеличения. Например, результат 3,536 с предельной абсолютной погрешностью 0,046 следует записывать как $3,53 \pm 0,05$. При округлении следует учитывать, что абсолютная погрешность численных данных или результата не должна превосходить половины единицы последнего разряда, сохраняемого при записи [12]. Так, запись 250 означает, что абсолютная погрешность не превосходит 0,5. Если погрешность результата ограничивается значением 0,05, то результат следует записать как 250,0. Записи 25 · 10; 250; 250,0; 250,00 означают, что абсолютные погрешности равны 5; 0,5, 0,05 и 0,005.

Еще один элемент правила округления заключается в том, что если первая из отбрасываемых цифр меньше 5, то последняя оставляемая цифра остается без изменения; если больше 5, то оставляемая цифра увеличивается на единицу. Если же отбрасывается 5 или пятерка с нулями, то последняя оставшаяся цифра остается без изменения, если она четная, или увеличивается на единицу до четной, если она нечетная.

Еще одно замечание. Необходимо всегда помнить, что операции с числами и размерными величинами, состоящими из чисел и физических размерностей, – это различные операции. Для количественных эквивалентов физических или размерных математических величин необходимо отдельно выполнять операции с числами и размерностями.

Наконец, последнее замечание. Наука о природе имеет дело с тем, что повторяется, допускает воспроизводимость. Это касается и измерения. Конкретный научный результат, касающийся исследования природы, получается в определенных начальных, текущих, предельных условиях. А как быть, если это не выполняется? Если субъект пока не в состоянии воспроизвести условия, например не знает, как это сделать, или когда это принципиально невозможно (как в точках бифуркации процессов, при спортивных состязаниях или на экзаменах)?

Внимание! Вернитесь к этому замечанию по окончании изучения дисциплины и попробуйте ответить на поставленный вопрос.

§ 3.6. ВРЕМЕННÓЕ И ЧАСТОТНОЕ ПРЕДСТАВЛЕНИЯ СИГНАЛОВ И ВРЕМЕННЫХ РЯДОВ

Сигналы как физические носители информации могут быть *статическими* и *динамическими*. Для первых характерно их постоянство во времени или вдоль каких-то других физических величин: температуры, давления, расстояния и т. п., для вторых – изменение вдоль других

величин, характеризующих развитие несущего информацию значения сигнала, его параметра, характеристики. Понятно, что это деление условно, поскольку все во Вселенной развивается во времени и в пространстве. Условность сводится к тому, насколько велики изменения сигнала за интересующий субъекта интервал развития сигнала вдоль соответствующей физической величины. Если эти изменения невелики (с точки зрения решаемой субъектом задачи), сигнал считается статичным, в противном случае – динамичным. Примеры статичных сигналов – любые сиюминутные сигналы о состоянии здоровья человека при диагностике: температура тела, пульс, артериальное давление, рисунок сетчатки глаза или узоров на пальцах и т. п. Однако при постоянном мониторинге состояния здоровья по этим сигналам их следует рассматривать как динамичные, изменяющиеся во времени, от количества принятых лекарств, физических параметров окружающей среды и т. д. Для конкретизации, упрощения рассмотрения и интерпретации дальнейших понятий положим, что динамичные сигналы развиваются во времени t . Будем подобные одномерные количественные¹ сигналы обозначать строчными греческими символами $\xi(t)$, $\psi(t)$, $\nu(t)$, $\eta(t)$, ..., а многомерные, состоящие из нескольких меняющихся во времени физических величин $\xi_i(t)$, $i = \overline{1, n}$, – полужирными символами $\xi(t) = (\xi_1(t), \xi_2(t), \dots, \xi_n(t))$. Для простоты в настоящем параграфе рассмотрим только одномерные сигналы. Как уже упоминалось, вместо работы с сигналами очень часто удобнее работать с их математическими моделями, которые мы будем обозначать латинскими буквами². Например, вместо развиваемых в естественном физическом времени t физических сигналов $\xi(t)$, $\psi(t)$, $\zeta(t)$ рассмотрим их модели $x(t)$, $y(t)$, $z(t)$ как математические функции временного аргумента t .

¹ Под количественными сигналами понимают такие, значения которых измеряются в количественных шкалах.

² Заметим, что во многих литературных источниках сигнал как физический носитель информации и его модель отождествляются, одинаково называются и обозначаются. Тем самым производится отождествление объекта-оригинала с его моделью, что вряд ли следует одобрять в связи с изложенными во второй главе сведениями относительно модели и моделируемого объекта, о невозможности, строго говоря, наличия между ними взаимно однозначного соответствия.

Рассмотрение разновидностей математических функций, иначе типов математических моделей, используемых для моделирования реальных (физических) сигналов (см. рис. 3.4), оставим до следующих глав. В настоящем же параграфе будем использовать знакомый всем из школьного курса детерминированный аппарат функций действительного или комплексного аргумента (переменной).

Напомним некоторые математические понятия. Под *функцией* понимают правило преобразования (отображения, установления соответствия) одного множества чисел в другое множество чисел, под *функционалом*¹ – отображение (правило преобразования, соответствия) множества функций в множество чисел (например, определенный интеграл), а под *оператором* – отображение одного множества функций в другие множества функций (операторы дифференцирования, неопределенного интегрирования и т. п.).

Нетрудно убедиться, что термин *оператор* обобщает термин *функционал*, который в свою очередь обобщает понятие *функция*. Поэтому очень часто слова *оператор* и *функционал* применяются в обобщенном понимании, хотя в конкретном контексте их можно заменить более узким термином. Например, при нахождении среднего оператор математического ожидания $\mathbf{M}\{\cdot\}$ (усреднения по вероятностной мере, см. далее) называют функционалом. В то же время для характеристической функции $\mathbf{M}\{\cdot\}$ уже нежелательно называть функционалом.

Зависимая переменная величина y называется (детерминированной) функцией (и записывается в виде $y = f(x)$) переменной величины x (аргумента, независимой переменной), если при заданном значении x величина y принимает только одно определенное значение (однозначная функция f , например, $y = |x|$, $y = x^2$, $y = \sin x$) или несколько определенных (например, n) значений (многозначная (n -значная), например, $y = \pm\sqrt{x}$, $y = \arcsin x$).

Конкретное аналитическое представление $f(x)$ часто имеет общепринятое название. Например, $y = \sin x$ – синусоидальная функция; $y = \exp x$ – экспоненциальная функция, экспонента и т. д.

¹ В более общей современной трактовке функционал есть отображение произвольного множества X в множество R действительных или C комплексных чисел.

Множество значений x , для которых функция $f(x)$ определена, называется **областью задания** (определения, существования) функции. Множество значений y , которым функцией $f(x)$ поставлены в соответствие (отображены) значения x из ее области задания, называются **областью значений функции** $f(x)$.

Функция $f(x)$, приращение которой $\Delta f(x) = f(x_2) - f(x_1)$ при $\Delta x = x_2 - x_1 > 0$ не меняет знака, называется **монотонной**. Функция, являющаяся непрерывной на $[a, b]$ и имеющая на $[a, b]$ непрерывную производную, – это **абсолютно непрерывная функция**. Наконец, если область определения функции ограничена конечным интервалом $[a, b]$, то такая функция называется **финитной** на $[a, b]$. Если за пределами $[a, b]$ функция равна нулю, она называется **конечнозаданной** на $[a, b]$. Заметим, что зачастую понятия *финитная функция* и *конечнозаданная функция* отождествляют, понимая под финитной конечнозаданную функцию.

Из определения функции следует, что если каждому $x \in A$ из области ее задания A соответствует одно единственное значение $y \in B$ из области ее значений B , то функция называется **однозначной**. Если верно обратное утверждение, т. е. каждому $y \in B$ соответствует одно единственное $x \in A$, то функция $f(x)$ называется **взаимно однозначной**.

Итак, в дальнейшем вместо сигналов $\xi(t)$, $\zeta(t)$, ... будем рассматривать их детерминированные модели $x(t)$, $z(t)$, ... Как уже упоминалось, рассматривается при этом только функция временного аргумента t . Такое представление сигналов назовем **временным**. Как следствие, все характеристики сигналов, являющихся физическими оригиналами их математических аналогов – характеристик временных моделей, будем называть временными.

Прежде чем перейти к рассмотрению характеристик сигналов (точнее их математических моделей), введем новые понятия и обозначения.

Изменяющиеся во времени физические величины – сигналы $\xi(t)$, как правило, принимают свои значения ξ из некоторого интервала $[\xi_{\min}, \xi_{\max}]$ – множества значений, которое математически считается непрерывным, непрерывнозначным, относящимся к континуальному –

бесконечному множеству мощности континуум. Одновременно считается, что и время t непрерывно изменяется в интервале $[t_1, t_2]$, которое также математически представляется непрерывным континуальным множеством. Такие физические величины, принимающие непрерывное множество значений, и соответствующие им непрерывные во времени сигналы называются **аналоговыми** (analog) (см. § 3.1). Значения $\xi(t)$, $x(t)$, определенные (рассматриваемые) для текущего момента времени t , называются **мгновенными значениями**. Числовые результаты, представляющие сигнал $\xi(t)$ или функцию $x(t)$ в фиксированные моменты времени t_1, t_2, \dots, t_N , называются их **отсчетами** (samples) и обозначаются в виде $\xi(t_i)$, $x(t_i)$, $i = \overline{1, N}$. Последовательность отсчетов $\xi(t_i)$, $x(t_i)$, $i = \overline{1, N}$, $t_j > t_i$ при $j > i$ называется **дискретным сигналом** (discrete signal) $\xi(t_i)$, $i = \overline{1, N}$, или **дискретным рядом** (discrete series) $x(t_i)$, $i = \overline{1, N}$. Поскольку в рассматриваемом нами случае t_1, t_2, \dots, t_N есть отсчеты времени, дискретный ряд называют **временным рядом** (time series). Дискретные сигналы и ряды могут получаться как естественно, когда мгновенные значения $\xi(t)$, $x(t)$ снимаются (измеряются, фиксируются) в дискретные значения аргумента (у нас моменты времени t_1, t_2, \dots, t_N), так и искусственно, когда аргумент (время t) и, следовательно, $\xi(t)$, $x(t)$ подвергаются искусственно операции **дискретизации** (sampling) их аргумента (в данном случае времени – **временная дискретизация** (time sampling)). Дискретизация может выполняться через неравные или равные промежутки времени Δt . В последнем случае она называется **равномерной**, эквидистантной или **периодической**, а интервал Δt при этом называется **периодом** (или **интервалом, шагом**) дискретизации (sampling time) или **интервалом отсчетов** (samples time). Величина $f_{\text{д}} = 1/\Delta t$, обратная периоду дискретизации Δt , называется **частотой дискретизации** (sampling frequency). Такие последовательности обозначаются в виде $x(0), x(\Delta t), x(2\Delta t), \dots, x[(N-1)\Delta t]$, $x_i = x(i\Delta t)$, $i = \overline{0, N-1}$, либо $x(1), x(2), \dots, x(N)$, $i = \overline{1, N}$. Для конкретности будем рассматривать только этот случай.

Помимо дискретизации по аргументу сигналы $\xi(t)$ и описывающие их функции $x(t)$ могут подвергаться еще одной операции – **квантованию по уровню** (quantization), при котором отдельному интервалу значений $\xi(t)$, $x(t)$ ставится в соответствие одно конкретное значение, как правило, среднее этого интервала. Такая операция имеет обязательно место при группировании результатов измерений, округлении значений сигналов, а также при вводе аналоговых сигналов в цифровые устройства, всегда оперирующие числами с ограниченным числом (двоичных, десятичных) разрядов.

Полученные после квантования по уровню сигналы называют **квантованными**, а погрешности, вызываемые заменой аналоговых сигналов квантованными при выполнении с ними следующих операций, называют **ошибками (шумами) квантования** (quantization error, quantization noise). Сигналы, подвергнутые обеим операциям (дискретизации и квантованию), называют **цифровыми**. На практике они получаются из аналоговых сигналов с помощью аналого-цифровых преобразователей (АЦП, Analog-to-Digital Converter, ADC).

Графическое модельное представление рассмотренных сигналов дано на рис. 3.5.

Прежде чем переходить к изложению дальнейшего материала, обратим внимание на еще одну важную сторону отличия сигналов $\xi(t)$ от их моделей $x(t)$, а также сигналов от данных, полученных по ним¹.

Согласно современной физической картине мира временная продолжительность сигналов ограничивается конечным интервалом $[t_1, t_2]$, не может быть бесконечной, т. е. для $\xi(t)$ величина t не может принадлежать $(-\infty, \infty)$. За пределами интервала $[t_1, t_2]$ нет физического смысла рассматривать $\xi(t)$. В свете этого даже представление $\xi(t)$ конечнозначной функцией является не очень корректным. Например, такое представление массы или температуры тела человека как функции времени равносильно утверждению, что масса или температура

¹ В свете изложенного в § 1.3, 2.1 и 3.4 заметим, что полученные по сигналу $\xi(t)$ данные представляют собой один из вариантов модельного представления сигнала $\xi(t)$.

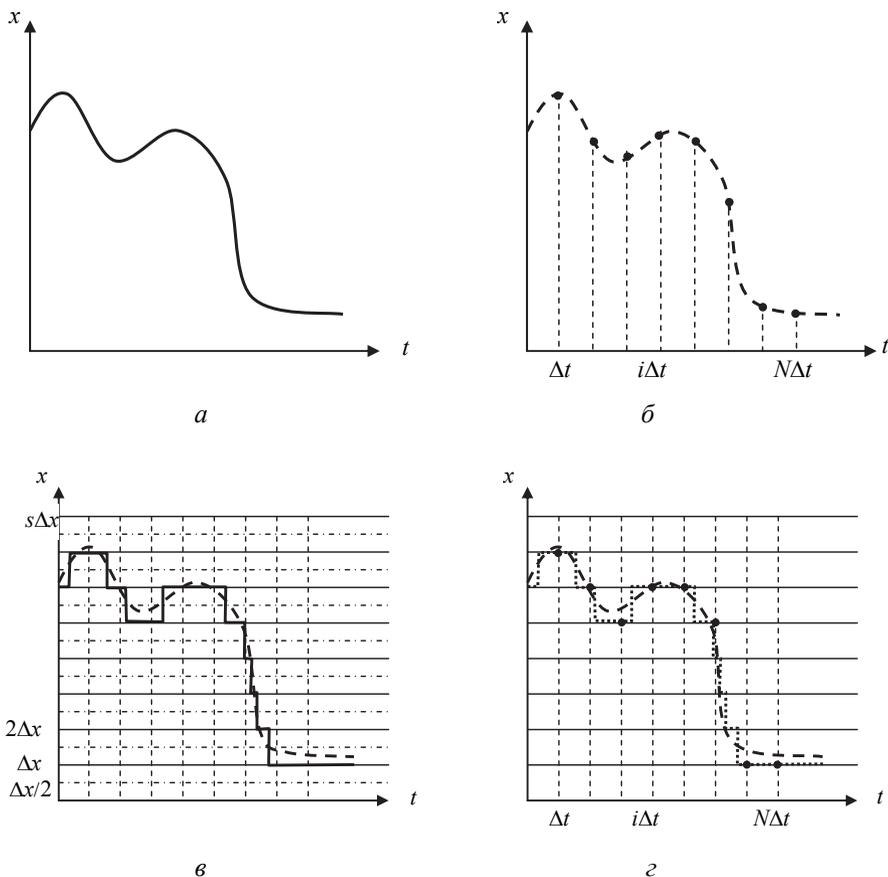


Рис. 3.5. Графическое модельное представление $x(t)$ аналогового (а), дискретного (б), квантованного (в) и цифрового (г) сигналов $\xi(t)$

тела человека до даты (времени) t_1 его рождения, т. е. при $t < t_1$, и после его смерти (время t_2), т. е. при $t > t_2$, равна нулю. Для описания физических сигналов, строго говоря, применимы лишь финитные функции. Широкое использование для них функций, заданных на $(-\infty, \infty)$ или $(0, \infty)$, является лишь удобной модельной аппроксимацией действительности, которая не всегда может быть адекватна решаемой

задаче. Более того, для сигнала, согласно определению термина, t всегда является текущим моментом времени t_0 и, следовательно, другие t могут быть только больше t_0 , т. е. для сигналов $t \geq t_0$. Чаще всего отсчет времени для сигналов начинается с $t_0 = 0$ и поэтому для них полагают $t \geq 0$. Что касается $t < t_0$, в частности $t < 0$, то это означает, что мы имеем дело с уже реализованным, прошедшим сигналом, который в лучшем случае может быть как-то зарегистрирован, записан, запомнен. Например, в виде электрокардиограммы (ЭКГ), электроэнцефалограммы (ЭЭГ), значения роста человека, его артериального давления и иных показателей, характеризующих прошлое состояние человека. А это уже не сигнал, а данные. Иными словами, на практике мы имеем дело не с «прошлым сигналом» и его значениями, а с его модельным представлением в виде данных или знаний-1 (анзний). А для данных, например, для записанной ЭКГ, начало отсчета t_0 может быть выбрано в любом месте ЭКГ, в частности так, что началу графика ЭКГ соответствует $t_1 = -T/2$, середине — $t_0 = 0$ ¹, а концу — $t_2 = T/2$, так что весь интервал имеет длину («длина» ЭКГ), равную T .

Именно поэтому для сигналов $\xi(t)$ обычно рассматривается временной интервал типа $[0, T]$ длиной $\Gamma = T$ или $[0, T + \tau_{\max}]$ длиной $\Gamma = T + \tau_{\max}$. Что касается модельного представления сигналов и данных в виде функции $x(t)$, то для нее будем рассматривать различные интервалы $[0, T]$, $[-T/2, T/2]$, а также $(-\infty, \infty)$, $(0, \infty)$.

Наконец, напомним, что функция $x(t)$ называется периодической с периодом T_0 , если $x(t) = x(t \pm kT_0)$ для всех $t \in (-\infty, \infty)$ и $k = 0, 1, 2, 3, \dots$. Понятно, что все периодические функции являются бесконечно протяженными во времени и не могут, строго говоря, адекватно отражать реальные сигналы, для которых условие периодичности финитно и имеет место для конечного набора $k = 0, 1, 2, \dots, K$.

Теперь введем математический *оператор усреднения* функции $x(t)$ по *аргументу*, а именно по *времени*:

¹ Ясно, что это условное, не физическое, а модельное начало отсчета времени, принятое за нулевое.

$$\mu_{\Gamma}\{f[\mathbf{x}(t)]\}^1 = \begin{cases} \mu_T\{f[\mathbf{x}(t)]\} = \frac{1}{T} \int_{-T/2}^{T/2} f[\mathbf{x}(t)] dt, \\ t - \text{непрерывно, } \Gamma = T; & (3.1) \\ \mu_T\{f[\mathbf{x}(t)]\} = \frac{1}{T} \int_0^T f[\mathbf{x}(t)] dt, \\ t - \text{непрерывно, } \Gamma = T; & (3.1a) \\ \mu_N\{f[\mathbf{x}(t)]\} = \frac{1}{N} \sum_{\substack{i=0 \\ (i=1)}}^{\overline{N-1}} f[\mathbf{x}(i\Delta t)], \\ t - \text{дискретно с шагом } \Delta t, \Gamma = N; & (3.1б) \\ \mu_N\{f[\mathbf{x}(t)]\} = \frac{1}{N} \sum_{i=-N'}^{\overline{N'}} f[\mathbf{x}(i\Delta t)], \\ t - \text{дискретно с шагом } \Delta t, \Gamma = N = 2N' + 1, T = N\Delta t, & (3.1в) \end{cases}$$

где $f(\cdot)$ – некая детерминированная функция, для которой стоящие справа в (3.1) интеграл и сумма математически существуют и имеют конечные значения; $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))$ – векторная функция аргумента t ; $T = (N-1)\Delta t$, если в (3.1б) $i = \overline{0, N-1}$, или $T = N\Delta t$, если $i = \overline{0, N}$ и первый шаг дискретизации (ширина его Δt) от $i = 0$ до $i = 1$ засчитывается в T .

В качестве аналогов характеристик $q(\boldsymbol{\lambda}; \mathbf{a})$ сигналов $\xi(t)$ будем рассматривать характеристики $Q(\boldsymbol{\lambda}; \mathbf{a})$ векторных аргументов $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ и параметра $\mathbf{a} = (\alpha_1, \dots, \alpha_k)$ аналога сигнала – детерминированной функции $\mathbf{x}(t)$.

¹ Поскольку математические операции в (3.1) выполняются с числами, в дальнейшем всюду в подобных случаях будем заменять сигнал $\xi(t)$ и его характеристики $q(\boldsymbol{\lambda})$ математическими образами $x(t)$, $Q(\boldsymbol{\lambda}; \mathbf{a})$, специально не оговаривая такую замену.

В качестве временных характеристик $Q(\lambda; \mathbf{a})$ рассмотрим

$$Q(\lambda; \mathbf{a}) = \mu_{\Gamma} \left\{ f_Q [x(t); \lambda; \mathbf{a}] \right\}, \quad (3.2)$$

где $f_Q[\cdot]$ – преобразование (функция), определяемое видом характеристики Q . Из них рассмотрим *среднее (среднее арифметическое)* $\overline{x(t)}$ функции $x(t)$ на интервале (отрезке) $\Gamma = (0, T]$ или $\Gamma[0, N)$:

$$m_x = \overline{x(t)} = \mu_{\Gamma} \left\{ f[x(t)] \right\}; \quad (3.3)$$

*дисперсию*¹ (среднюю мощность централизованного сигнала)

$$s_x^2 = \mu_{\Gamma} \left\{ [x(t) - m_x]^2 \right\} \quad (3.3a)$$

или среднеквадратическое отклонение (от среднего) $s = s_x = +\sqrt{s_x^2}$; *среднюю мощность* функции (сигнала) на интервале Γ

$$P_{cp} = \mu_{\Gamma} \left\{ x^2(t) \right\}, \quad (3.3b)$$

или энергию сигнала на интервале T : $E = TP_{cp}$. При этом физическая размерность мощности $[P_{cp}] = [\xi^2]$, а размерность энергии $[E] = [T\xi^2]$ или $[E] = [\xi^2][t]$, где $[t]$ – размерность t , в секундах.

Еще один класс временных характеристик сигналов $\xi(t)$ и функций $x(t)$ – это корреляционные функции, характеризующие связь отсчетов $\xi(t)$, $x(t)$, отстоящих друг от друга вдоль временной оси t на интервал τ .

Нормированная автокорреляционная функция $x(t)$ определяется выражением

$$r_{xx}(\tau) = \lim_{\Gamma \rightarrow \infty} \frac{\mu_{\Gamma} \left\{ \overset{\circ}{x}(t) \overset{\circ}{x}(t + \tau) \right\}}{\mu_{\Gamma} \left\{ \left[\overset{\circ}{x}(t) \right]^2 \right\}}, \quad (3.4)$$

¹ Зачастую, если это не приведет к путанице, индексы x будем опускать.

а *нормированная взаимная корреляционная функция* –

$$r_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{\mu_{\Gamma} \left\{ \overset{\circ}{x}(t) \overset{\circ}{y}(t + \tau) \right\}}{\sqrt{\mu_{\Gamma} \left\{ \left[\overset{\circ}{x}(t) \right]^2 \right\} \mu_{\Gamma} \left\{ \left[\overset{\circ}{y}(t) \right]^2 \right\}}}, \quad (3.4a)$$

где $\overset{\circ}{x}(t) = x(t) - \overline{x(t)}$ – центрированная функция $x(t)$. В (3.4) необходимо сократить Γ в числителе и знаменателе до перехода к пределу для непериодических функций $x(t)$, $y(t)$, а для периодических функций можно положить Γ равным интервалу периодичности функции $\Gamma = \Gamma_0$ (или $T = T = T_0$).

Как следует из (3.4) и (3.4a), $|r_{xx}(\tau)| \leq r_{xx}(0) = 1$, $r_{xx}(\tau) = r_{xx}(-\tau)$, $|r_{xy}(\tau)| \leq 1$.

Если $r_{xx}(\tau) = 0$ при $\tau = \tau'$, то говорят, что отсчеты $x(t)$ и $x(t + \tau')$ *некоррелированы*. Если же $r_{xy}(\tau) = 0$ для всех τ , то говорят, что функции $x(t)$ и $y(t)$ *некоррелированы*.

Часто в приложениях вводятся понятия *время (интервал) корреляции* τ_{ε} на уровне $\varepsilon > 0$, $\varepsilon \approx 0$, определяемое неравенством

$$|r_{xx}(\tau)| \leq \varepsilon \text{ при } |\tau| \geq \tau_{\varepsilon}, \quad (3.5)$$

и *радиус корреляции* τ_0

$$\tau_0 = \frac{1}{2} \int_{-\infty}^{\infty} r_{xx}(\tau) d\tau = \int_0^{\infty} r_{xx}(\tau) d\tau. \quad (3.5a)$$

Из (3.5) следует, что $2\tau_0$ характеризует основание прямоугольника высотой, равной единице, с площадью, равной площади корреляционной функции.

Смысл и суть τ_{ε} и τ_0 иллюстрирует рис. 3.6.

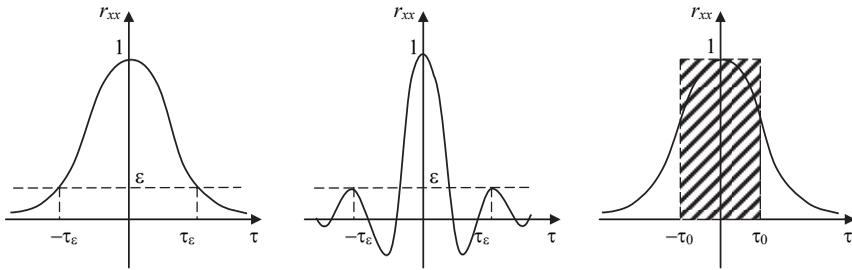


Рис. 3.6. Пояснение терминов *интервал* и *радиус корреляции*

Помимо временного представления непрерывных и дискретных по t сигналов $\xi(t)$ и описывающих их непрерывных и решетчатых (дискретных по t) функций $x(t)$ очень часто рассматривается их частотное представление. Опять-таки, чтобы избежать комментариев по поводу математической «неидеальности» сигналов $\xi(t)$ (например, их периодичности, ограниченности во времени и т. п.), далее будем рассматривать только их модельное частотное представление, т. е. представление функций $x(t)$, оставив рассмотрение вопросов пригодности такого модельного представления сигналов в виде дальнейших комментариев по тексту книги.

Прежде всего положим, что $x(t)$ есть периодическая с периодом T функция, т. е. $x(t) = x(t \pm kT)$, где $k = 0, 1, 2, \dots$. Тогда его частотным представлением является гармоническое разложение, т. е. представление в виде ряда Фурье [18, 19] в синусно-косинусной (3.6), вещественной (3.6, 3.6a) или комплексной экспоненциальной (3.6.б) форме, а именно

$$x(t) = \begin{cases} a_0 + \sum_{k=1}^{\infty} (a_k \cos(k\omega_1 t) + b_k \sin(k\omega_1 t)); & (3.6) \\ a_0 + \sum_{k=1}^{\infty} A_k \cos(k\omega_1 t + \varphi_k); & (3.6a) \\ a_0 + \sum_{k=1}^{\infty} \frac{A_k}{2} \left[e^{j(k\omega_1 t + \varphi_k)} + e^{-j(k\omega_1 t + \varphi_k)} \right], & (3.6б) \end{cases}$$

где $\omega_k = 2\pi f_k$ – круговая частота k -й гармоники в радианах; f_k – линейная частота в герцах, $\Gamma c = 1/c$; ω_1, f_1 – частоты первой гармоники, причем $f_1 = 1/T$;

$$\left. \begin{aligned} a_0 &= \frac{1}{T} \int_0^T x(t) dt, & a_k &= \frac{2}{T} \int_0^T x(t) \cos(k\omega_1 t) dt; \\ b_k &= \frac{2}{T} \int_0^T x(t) \sin(k\omega_1 t) dt; \\ A_k &= \sqrt{a_k^2 + b_k^2}, & \varphi_k &= \arctg(b_k/a_k). \end{aligned} \right\} \quad (3.7)$$

Заметим, что для четной функции $x(t) = x(-t)$ все $b_k = 0$, а для нечетной $x(t) = -x(-t)$ все $a_k = 0$.

Совокупность A_k и φ_k образует дискретный (линейчатый) спектр непрерывной периодической функции $x(t)$. Их изображения в координатах амплитуда A_k –частота ω или f либо фаза φ_k –частота называются амплитудным и фазовым спектрами функции $x(t)$.

Как уже упоминалось, для периодических функций с периодом T , а также для периодизованных с периодом T (склеенных, сплетенных в кольцо) функций $x(t)$ (и сигналов $\xi(t)$) с ограниченной областью определения $[0, T]$ используется понятие средней мощности, которая может быть представлена как во временной, так и в частотной области, а именно

$$P_{\text{ср}} = \frac{1}{T} \int_0^T x^2(t) dt = a_0^2 + \sum_{k=1}^{\infty} A_k^2/2, \quad (3.8)$$

где a_0 и $A_k^2/2$ есть средняя мощность постоянной составляющей (константы a_0) и k -й гармоники функции $x(t)$.

Как видно из правой части (3.8), при переходе от амплитудного спектра к спектру мощности теряется информация о фазе спектральных составляющих функций $x(t)$.

Если $x(t)$ не периодическая функция, то спектральное (частотное) представление определяется в виде

$$X(\lambda) = \mathfrak{F} \left\{ \left[x(t)e^{-j\lambda t} \right]_t \right\}, \quad (3.9)$$

где $j = \sqrt{-1}$ и

$$\mathfrak{F} \left\{ \left[x(t)e^{-j\lambda t} \right]_t \right\} = \begin{cases} \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt, & \lambda = \omega \in (-\infty, \infty), \\ \text{если } t - \text{ непрерывно;} \end{cases} \quad (3.10)$$

$$\mathfrak{F} \left\{ \left[x(t)e^{-j\lambda t} \right]_t \right\} = \begin{cases} \Delta t \sum_{i=-\infty}^{\infty} x(i\Delta t)e^{-j\omega i\Delta t}, & \lambda = \omega \in (-\infty, \infty), \\ \text{если } t - \text{ дискретно;} \end{cases} \quad (3.10a)$$

$$\mathfrak{F} \left\{ \left[x(t)e^{-j\lambda t} \right]_t \right\} = \begin{cases} \Delta t \sum_{i=-\infty}^{\infty} x(i\Delta t)e^{-j\nu i}, & \lambda = \nu = \omega\Delta t, \nu \in [-\pi, \pi], \\ \text{если } t - \text{ дискретно;} \end{cases} \quad (3.10б)$$

есть прямое преобразование Фурье для непрерывной (3.10) или дискретной (3.10a), (3.10б) функции.

Функция $x(t)$ связана с $X(\lambda)$ обратным преобразованием Фурье:

$$x(t) = \frac{1}{2\pi} \int_{\Lambda} X(\lambda)e^{-j\lambda t} d\lambda, \quad (3.11)$$

где $\Lambda = (-\infty, \infty)$, если t – непрерывно, т. е. $\lambda = \omega$, и $\Lambda = (-\pi, \pi)$, если t – дискретно, т. е. $\lambda = \nu$.

Комплексная функция $X(\omega)$ называется **спектральной характеристикой (функцией)** функции $x(t)$ или **спектральной плотностью**. Она может быть представлена в комплексной форме в виде

$$X(\lambda) = R(\lambda) - jI(\lambda) = |X(\lambda)| \exp\{-j\varphi(\lambda)\}; \quad (3.12)$$

$$|X(\lambda)| = \sqrt{R^2(\lambda) + I^2(\lambda)}, \quad \varphi(\lambda) = \text{arctg}(I(\lambda)/R(\lambda)), \quad (3.13)$$

где $|X(\lambda)|$ – модуль спектральной плотности есть четная функция частоты λ ; $\varphi(\lambda)$ – аргумент спектральной плотности – нечеткая функция частоты λ или, иначе, **амплитудно-частотная** (АЧХ $|X(\omega)|$) и **фазо-частотные** (ФЧХ $\varphi(\omega)$) характеристики сплошного спектра функции $x(t)$ (сигнала $\xi(t)$).

Завершим временное и частотное представления функций $x(t)$ некоторыми важными соотношениями (см., например, [18, 19]). Рассмотрим их только для непрерывных функций $x(t)$.

Первое – это **равенство Парсеваля**, касающееся полной энергии E функции $x(t)$:

$$E = \int_{-\infty}^{\infty} x^2(t) dt = \frac{1}{\pi} \int_0^{\infty} |X(\omega)|^2 d\omega. \quad (3.14)$$

Второе – это **эффективная ширина спектра**

$$\Delta\omega_{\text{э}} = \int_0^{\infty} |X(\omega)|^2 d\omega / |X(0)|^2. \quad (3.15)$$

Третье, **соотношение неопределенности**

$$\tau_{\text{эф}} \omega_{\text{эф}} = \mu, \quad (3.16)$$

где μ – небольшое число, $\tau_{\text{эф}}$ – **эффективная длительность** $x(t)$ (или сигнала $\xi(t)$), определяемая как

$$\int_0^{\tau_{\text{эф}}} x^2(t) dt = k_{\text{э}} E, \quad (3.17)$$

где $k_{\text{э}}$ – близкое к единице значение, например, $k_{\text{э}} = 0,9$, а $\omega_{\text{эф}} = 2\pi f_{\text{эф}}$ – **эффективная ширина спектра** $x(t)$ (или сигнала $\xi(t)$), определяемая из соотношения

$$\frac{1}{\pi} \int_0^{\omega_{\text{эф}}} |X(\omega)|^2 d\omega = k_{\text{э}} E. \quad (3.18)$$

Из соотношений (3.15) и (3.16) следует практически важный вывод о том, что чем меньше будет $\tau_{эф}$, тем больше должны быть $\Delta\omega_3$ или $\omega_{эф}$, т. е. чем короче область (интервал) определения $x(t)$, тем шире область (интервал) определения $X(\omega)$, т. е. короткому сигналу соответствует широкий спектр и, наоборот, длинному (во времени) сигналу соответствует узкий (по частоте) спектр, а конечнозаданному или финитному, ограниченному во времени, сигналу соответствует бесконечнопротяженный, нефинитный (неограниченный по частоте) спектр.

Наконец, укажем еще два соотношения. Первое соотношение – связь между спектрами $X(\omega)$ непрерывной $x(t)$ и $X_d(\omega)$ дискретизованной (решетчатой) $x(i\Delta t)$, $i = 0, \pm 1, \pm 2, \dots$ функций [18, 19]:

$$X_d(\omega) = \sum_{k=-\infty}^{\infty} X(\omega - k\omega_d), \quad (3.19)$$

где $\omega_d = 2\pi f_d = 2\pi/(\Delta t)$ – круговая частота дискретизации, Δt – шаг дискретизации. Обратим в первую очередь внимание на то, что спектр дискретизованной (решетчатой) функции $x(i\Delta t)$, $i = 0, \pm 1, \pm 2, \dots$ является периодическим в частотной области с периодом, равным частоте дискретизации ω_d .

Второе соотношение – спектр произведения функций $z(t) = x(t)y(t)$. Спектр $Z(\omega)$ функции $z(t)$ равен

$$Z(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(u)Y(\omega - u)du = X(\omega) \otimes Y(\omega), \quad (3.20)$$

где $X(\omega)$ – спектр $x(t)$; $Y(\omega)$ – спектр $y(t)$, а \otimes – знак интеграла свертки, стоящего в средней части (3.20).

Примеры спектров для разных видов функций $x(t)$ (и соответствующих им сигналов $\xi(t)$) приведены в табл. 3.5.

В подпараграфе 4.3.6 еще раз остановимся на представлении реальных физических сигналов их моделями. В частности, обратим внимание на то, что все физические сигналы ограничены по области определения (во времени).

§ 3.7. ПОНЯТИЕ О ПЛАНИРОВАНИИ НАБЛЮДЕНИЙ И ЭКСПЕРИМЕНТОВ

Прежде всего, рассматривая экспериментирование, введем еще несколько важных понятий.

Опыт – это реализация (осуществление) или воспроизведение исследуемого объекта (явления, процесса) в конкретных (определенных) условиях его наблюдения и проведения эксперимента при возможности регистрации их результатов. Поскольку, согласно определению, эксперимент как действие есть совокупность операций, наблюдений или воздействий, направленных на получение информации об исследуемом моделируемом объекте через экспериментальные Данные, их дальнейшую обработку, анализ и интерпретацию их результатов, ясно, что опыт есть элементарная составная часть наблюдения, эксперимента.

Фактор – переменная величина¹, принимающая в некоторый момент времени (наблюдения или экспериментирования) определенное значение. В зависимости от допустимой и используемой измерительной шкалы они могут быть качественными и количественными. Фиксированное значение фактора **X** называется *уровнем фактора*, а область возможных значений фактора **X** при экспериментировании – *областью действия* его.

Чаще всего, с точки зрения наблюдения и экспериментирования, рассматриваемый в них объект представляют в виде кибернетической модели «черного ящика», графическое представление которой приведено на рис. 3.7.

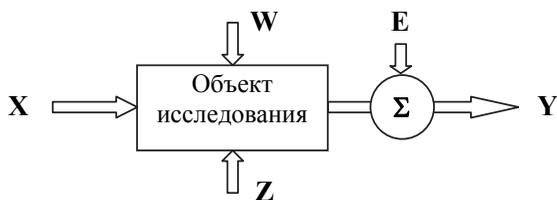


Рис. 3.7. Простейшая структурная схема объекта исследования с аддитивной помехой **E** на выходе

¹ Измеряемая физическая в физических экспериментах, вычисляемая математическая – в машинных вычислительных экспериментах.

На рис. 3.7 $\mathbf{X} = (x_1, x_2, \dots, x_n)^T$ – вектор входных контролируемых и управляемых переменных x , т. е. таких, которыми субъект-исследователь в эксперименте может варьировать по своему усмотрению; $\mathbf{Z} = (z_1, z_2, \dots, z_k)$ – вектор контролируемых, но неуправляемых переменных; $\mathbf{W} = (w_1, w_2, \dots, w_m)$ – вектор неуправляемых и неконтролируемых переменных; $\mathbf{Y} = (y_1, y_2, \dots, y_l)^T$ – вектор выходных переменных – значений показателей, характеристик, параметров, свойств, качеств объекта; $\mathbf{E} = (e_1, e_2, \dots, e_s)^T$ – вектор аддитивных помех; t – символ транспонирования матриц; Σ – символ оператора суммирования.

Выходную переменную \mathbf{Y} часто называют *зависимой, откликом*, зависимостью отклика от рассматриваемых факторов – *функцией отклика*, а ее геометрическое представление – *поверхностью отклика*.

Из изложенного ранее понятно, что объектом исследования могут быть реальные физические объекты; физические (натурные) модели реальных объектов; формальные модели (математические аналитические (знаковые), образные, алгоритмические, программные и др.) (см. табл. 2.1–2.3).

Приведенная на рис. 3.7 схема является обобщенной. В *пассивном эксперименте* (в наблюдениях) «существуют» (воспринимаются субъектом) только факторы \mathbf{Z} . В них субъект-исследователь находится в положении пассивного наблюдателя, фиксирующего состояние и изменения фактора \mathbf{Z} и отклика \mathbf{Y} , а все процессы, связанные с объектом, протекают без вмешательства субъекта-исследователя. В *активном эксперименте* предполагается, что субъект-экспериментатор может целенаправленно изменять факторы \mathbf{X} . Ясно, что и в эксперименте необходимо, чтобы субъект-исследователь мог учитывать влияние контролируемых факторов \mathbf{Z} , а при наблюдениях и экспериментировании нужно стараться организовать такое их проведение, которое позволило бы обнаружить и, по возможности, избежать или нивелировать влияние переменных \mathbf{W} .

Теперь перейдем к рассмотрению важнейшего этапа организации наблюдений и экспериментов – их планирования.

Термин *планирование* в рассматриваемом контексте будем понимать в следующих смыслах.

Планирование в широком смысле есть составление, *разработка плана решения* стоящей перед субъектом-исследователем задачи, представляющего собой систему мероприятий, предусматривающих формулировку цели и задач исследования; выбор методов, моделей и их характеристик, средств и условий; порядок, последовательность и сроки выполнения работ, операций, опыта, когда все они объединенно направлены на достижение поставленных целей исследования. Понятно, что если какая-то часть плана задана заранее, планирование относится к другим его составляющим.

Планирование в узком смысле есть рациональная организация только этапов задания и измерения значений переменных, т. е. физических величин (параметров) и состояний естественных и искусственных физических объектов, факторов в натуральных экспериментах, а также задания и вычисления переменных и факторов в машинных экспериментах. Здесь под *планом эксперимента* понимается совокупность сведений, данных, определяющих количество, условия и порядок проведения опытов.

В такой формулировке термин *планирование* применим как для пассивных (наблюдения), так и для активных экспериментов.

Следующее трактование этого термина связано с математической статистикой и ее составной частью – прикладной статистикой (см. следующую главу). Здесь под планированием (активного) эксперимента понимается, с одной стороны, раздел математического аппарата статистики, изучающий рациональную организацию измерений, подверженных погрешностям и описываемых вероятностно-статистическими моделями, а также построения по результатам таких измерений идентификационных моделей, описывающих связь Y с X , Z , E желательно с учетом присутствия W . С другой стороны, теоретический математический вывод (доказательство, расчет) наилучшего по принятому критерию плана эксперимента в смысле, указываемом ниже. Указанная выше связь обычно описывается регрессионной моделью (см. табл. 3.4, модель измерения)

$$Y = f(X, Z; \theta) + E, \quad (3.21)$$

где $\theta = (\theta_1, \theta_2, \dots, \theta_q)$ – вектор параметров модели. При этом, как правило, предполагается, что функция f является детерминированной, а параметры θ неизвестны, т. е. θ заменяется на Θ и подлежат опреде-

лению по эмпирическим Данным¹ с учетом или без учета погрешностей их измерения.

Забегая вперед (см. следующую главу), укажем, что функция регрессии описывает зависимость среднего (вдоль множества его возможных значений) значения Y от конкретных значений $X = x$ и $Z = z$. Регрессионная модель (3.21) призвана описать множество возможных значений Y вокруг его среднего путем введения E , учитывающего всякие случайности (погрешности измерения конкретных значений Y , влияния на них факторов W и т. д.).

План эксперимента есть совокупность значений управляемой переменной X и количество их повторений в каждом опыте эксперимента.

Ясно, что при этом речь идет только об активных экспериментах. Область значений факторов X , в которой могут находиться точки, отвечающие условиям проведения опытов по конкретному плану эксперимента, называется **областью планирования**. Упорядоченная совокупность назначаемых численных значений факторов X , соответствующая условиям проведения конкретного i -го опыта, есть **точка плана** $x_i = (x_{1i}, x_{2i}, \dots, x_{ni})^T$. Общая совокупность всех N таких точек плана, $i = 1, 2, \dots, N = \overline{1, N}$, образует **план эксперимента**, а совокупность из L различных точек плана – **спектр** плана.

Конкретный план эксперимента определяется целью, назначением эксперимента. С точки зрения системного подхода цель эксперимента должна определяться целью исследования объекта, той конечной задачей, ради которой проводится исследование объекта и, как следствие, ставится эксперимент. Иными словами, вначале должно выполняться планирование эксперимента в широком смысле, затем – в узком, после чего – в математической его постановке. В математической теории эксперимента чаще всего ограничиваются локальными целями. По ним эксперименты делятся на оценочные, идентификационные, сравнительные, отсеивающие и экстремальные.

¹ Здесь и в дальнейшем для краткости для Данных, полученных по итогам физического опыта (наблюдения, измерения, экспериментирования) на реальных объектах, будем использовать единый термин *эмпирические*, т. е. полученные опытным путем, в отличие от *модельных*, полученных с помощью расчетов по аналитическим моделям или при вычислительных экспериментах.

Целью *оценочных экспериментов* является определение или уточнение значений абсолютных констант, получение как можно более точных значений измеряемых величин (см. табл. 3.4).

Например, на практике со случайными погрешностями измеряется значение величины

$$y = f(\mathbf{x}; \boldsymbol{\theta}), \quad (3.22)$$

где $f(\cdot)$ – детерминированная (регулярная) функция одно- или многомерного аргумента $\mathbf{x} \in X$ и неизвестного одно- или многомерного параметра $\boldsymbol{\theta} \in \Theta$. Из-за наличия случайных погрешностей E измерения y при конкретных значениях \mathbf{x} , вместо (3.22) мы имеем $Y = y + E$, т. е. вместо y результат измерения есть $\hat{y} = y + \varepsilon$, где ε – конкретное значение погрешности, определяемое конкретным сочетанием факторов, действующих на объект, и условий измерения y в момент их выполнения. В более общей постановке считается, что \mathbf{x} также измеряется с погрешностью \mathbf{E}_x . Тогда (3.22) заменяется на

$$y = f(\mathbf{x} + \mathbf{E}_x; \boldsymbol{\theta}), \quad (3.23)$$

тем самым полагается, что

$$\hat{y} = f(x_1 + \varepsilon_{x_1}, x_2 + \varepsilon_{x_2}, \dots, x_n + \varepsilon_{x_n}; \theta_1, \dots, \theta_k) + \varepsilon_y. \quad (3.24)$$

Цель эксперимента – оценка параметров $\boldsymbol{\theta}$ или некоторых требуемых для решаемой задачи функций от них, в том числе $f(\mathbf{x}; \boldsymbol{\theta})$, либо проверка гипотез о параметрах $\boldsymbol{\theta}$. Под *планом эксперимента* понимается совокупность значений реальных физических величин, представляемых в (3.22) переменными \mathbf{x} , задаваемых в эксперименте.

Цель эксперимента – достижение максимальной точности или минимальной неопределенности результатов измерения y при минимальном количестве проведенных измерений и сохранении статистической достоверности результатов.

Цель *идентификационных экспериментов* – получение наиболее адекватных (по совокупности показателей) моделей, описывающих объект и его окружение (факторы), влияние факторов на объект.

Отсеивающие эксперименты ориентированы на выявление наиболее существенных факторов, действующих на объект в соответствующих условиях, и отсеивание малосущественных.

Задачи **сравнительных экспериментов** – выявление способов исследования объектов, которые дают лучшие конечные результаты в тех случаях, когда исходы (или результаты) экспериментов, полученных тем же методом в разных условиях или разными методами в тех же условиях, не согласуются друг с другом.

Наконец, цель **экстремальных экспериментов** – отыскание оптимальных (наилучших в заданном смысле, по конкретным критериям) режимов функционирования многофакторных объектов.

Таким образом, обобщая, можно сказать, что целью рассматриваемых «теоретических» экспериментов является либо оценивание параметров θ заданной функции f или в более общей постановке самой функции f и ее параметров θ , либо проверка гипотез о параметрах θ , способах исследования, равенства константы конкретному значению и т. п. Заметим, что подобные эксперименты в математике являются, во-первых, объектами исследования; во-вторых, инструментом для нахождения оптимальных по определенным критериям решений.

Следует в связи с этим заметить, что утверждения типа, «пользуясь объективными методами математической статистики и построенным на ее основе планом экспериментирования и методами оценивания, мы *вывели (получили, доказали)* следующую функциональную зависимость f », неверны. Корректнее было бы сказать, что гипотеза о наличии такой зависимости не отвергается. Ведь даже несколько подтверждений гипотезы является лишь необходимым условием не отвергнуть ее, но не достаточным, чтобы принять! В отличие от этого один корректно полученный отрицательный результат достаточен для ее опровержения.

Наконец, планирование вычислительного эксперимента связано с разработкой плана численного решения задачи, поставленной исследователем объекта по имеющейся в его распоряжении математической модели этого объекта. Это, например, задачи по поведению объекта в конкретных условиях, определению усредненных траекторий отклика, граничных и критических значений областей действия факторов и т. д.

Цель **вычислительного эксперимента** – расчет свойств, особенностей строения и поведения реально существующего или создаваемого объекта с помощью численного анализа его математической аналити-

ческой или алгоритмической модели. Эти расчеты в настоящее время обычно проводятся на ЭВМ, компьютере. Поэтому иногда синонимом термина *вычислительный эксперимент* выступает термин *машинный или компьютерный эксперимент*.

Отличие вычислительного эксперимента от физического видно из рис. 2.1, в упрощенном варианте оно отражено на рис. 3.8.

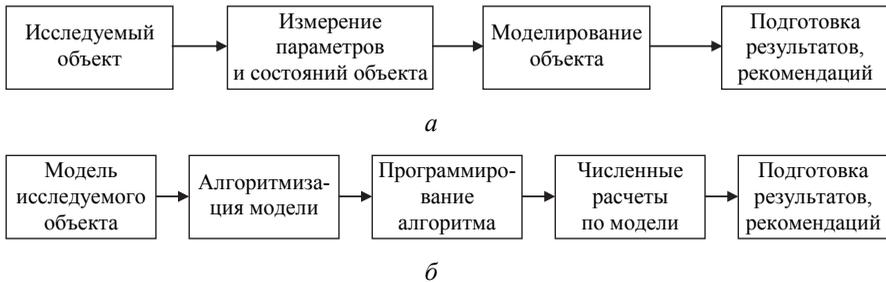


Рис. 3.8. Условная схема последовательности этапов экспериментирования:

а – физический эксперимент; *б* – вычислительный эксперимент

Ранее мы не отмечали, а теперь уместно заметить, что физические эксперименты с реальными объектами могут быть двух типов: натурные и лабораторные. В *натурных* экспериментах различным испытаниям подвергается сам объект, чтобы определить его свойства и возможности. В *лабораторных* же условиях создаются экспериментальный образец объекта, условия и аппаратура, которые имитируют («проигрывают») объект и условия его функционирования. Близкая к этому ситуация и в вычислительных экспериментах.

Сущность метода исследования, основанного на вычислительном эксперименте, состоит в том, что на основе аналитической или алгоритмической модели объекта «проигрывается» поведение объекта в разных условиях: начальных, при конкретных значениях параметров модели, соответствующих разным состояниям и режимам функционирования объекта. В результате этого находятся оптимальные, наихудшие, допустимые, граничные, критические и иные режимы работы объекта и значения его параметров, показателей состояния.

Важно отметить, что при использовании в вычислительном эксперименте аналитической модели в виде систем дифференциальных уравнений, вероятностной теории, теории оптимизации, массового

обслуживания и иных (см. следующую главу), объект «как бы условно проигрывается» на модели без повторения каждый раз всех действий процесса его функционирования, в то время как при использовании алгоритмической модели, описывающей механизм, правила, последовательность действий, развития процесса функционирования объекта, его работа «имитируется», т. е. каждый раз воспроизводится на модели. Например, если для определения пропускной способности автомобильного моста (максимальной скорости движения транспорта в каком-то сечении моста в конкретных условиях) используется аналитическая модель теории массового обслуживания, то, задавая разные параметры модели, касающиеся входного потока автомобилей и ограничений, вытекающих из правил дорожного движения, и рассчитывая по ней пропускную способность моста, мы как бы «проигрываем», какой она будет в разных ситуациях. При применении же алгоритма функционирования моста, мы имитируем движение по нему разных автомобилей с учетом их габаритов, скорости движения, расстояний между ними, ограничений по правилам движения и подсчитываем, сколько же максимально машин сможет пройти при этом в данном сечении моста.

Ясно, что вычислительные эксперименты имеют как сильные (достоинства), так и слабые (недостатки) стороны по сравнению с физическими, свои плюсы и минусы перед лабораторными. Среди достоинств вычислительных экспериментов отметим следующие:

- можно точно воспроизвести эксперимент;
- неожиданные результаты, полученные в них, можно обнаружить и проверить, повторяя эксперименты точно в тех же условиях;
- в них принципиально проще установить причину неожиданных результатов;
- они свободны от ряда ограничений, характерных для физических экспериментов: экономических, ресурсных, технических, кадровых и т. п.;
- можно быстро получить результат;
- удастся исследовать такие объекты и режимы их функционирования, которые невозможно провести с помощью физических экспериментов.

Однако достоверность результатов и выводов, полученных с помощью вычислительных экспериментов, нуждается в дополнительном

исследовании. Во-первых, модель лишь приближенно, ограниченно описывает объект. Во-вторых, могут иметь место значительные ошибки в реализации модели. В-третьих, расчеты порой выходят за пределы допустимой области адекватности модели объекту за счет погрешностей измерения, оценивания параметров, численного выполнения аналитических выражений, ограничения разрядов операндов, использования рекуррентных и расходящихся или плохо сходящихся итерационных процедур и т. д.

Еще одно важное замечание! Оно связано с тем, что задачи планирования экспериментов, как и более «узкие» задачи, например, идентификации объектов, являются **некорректными** в том смысле, что их результаты (решения) могут быть не единственными даже при отсутствии погрешностей измерения, а тем более при замене критериев их качества.

Критерии¹ качества плана эксперимента определяются целью эксперимента и могут формулироваться по-разному в зависимости от вида решаемой прикладной задачи, назначения плана и условий его применения.

В математической теории эксперимента они, как правило, сводятся к разнообразным критериям оптимальности плана. Во всех критериях отражается общая направленность планирования эксперимента, которую сокращенно можно отразить фразой:

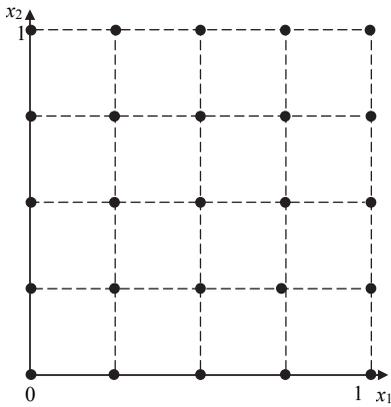
«Хороший план эксперимента тот, в котором реализовано правило “Меньше опытов + больше информации = выше качество результатов”».

Заметим, что критерий может быть не только скалярным, но и векторным, ориентированным на достижение с помощью плана нескольких целей, задач получения его результатов.

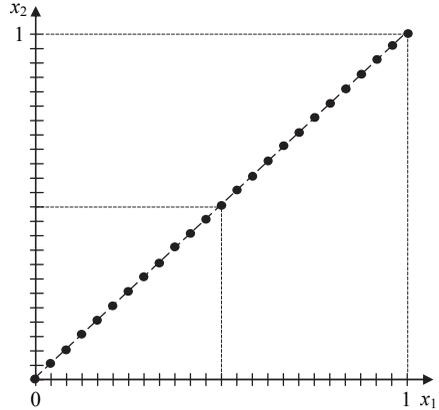
Для лучшего понимания идей планирования экспериментов и выделенного в рамке правила на рис. 3.9 приведены разные примеры.

Внимание! Попробуйте сами придумать показатели качества планов экспериментов на рис. 3.9 и сравнить их по этим показателям.

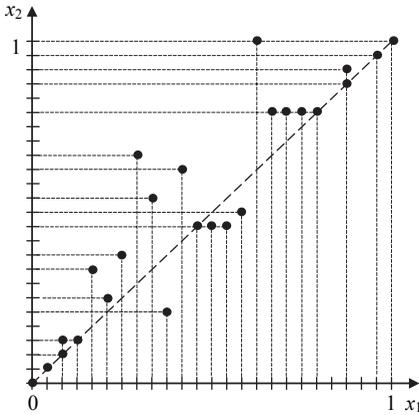
¹ **Критерий** (греч. kriterion – средство для суждения, решения) – мерило, правило, признак, на основании которого производится оценка, классификация чего-либо, принятие решения о чем-либо.



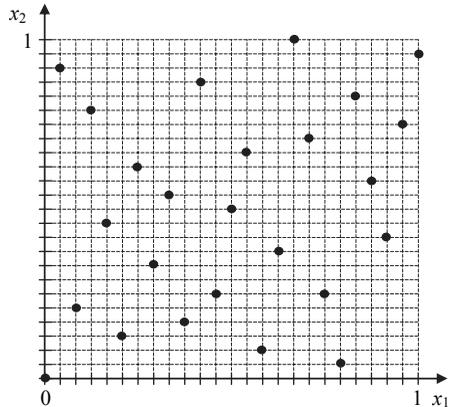
a



б



в



г

Рис. 3.9. Пример планов эксперимента при двумерном приведенном аргументе $x = (x_1, x_2)$; $x_1, x_2 \in [0, 1]$; количество измерений $N = 25$

Обычно главный критерий формулируется так, чтобы был достигнут основной результат эксперимента. Например, если эксперимент связан с измерением значения физической величины, то критерий может заключаться в достижении максимальной достоверности D результата измерения, т. е. минимального значения смещения, дисперсии,

доверительного интервала и т. п. Тогда дополнительным критерием для сравнения планов экспериментов может быть минимум величин функции потерь эксперимента $R = \Phi + \psi(D)$, где Φ – обобщенные (временные, материальные и т. д.) затраты, а $\psi(D)$ – некоторый функционал, зависящий от значения достоверности D и определяемый конечной целью исследования.

В качестве дополнительных критериев сравнения планов или достижения ими нескольких целей и задач могут быть также следующие.

Например, первый дополнительный критерий: эксперимент должен позволять получить не только наилучшую (по принятому критерию адекватности) на интервале времени и области условий функционирования объекта модель отклика или объекта, но и такую, которая при этом позволяет осуществлять прогноз поведения объекта за пределами этих интервалов и областей, в частности осуществлять прогноз вперед (анте-прогноз, forecasting) или назад (пост-прогноз, backcasting). Вторым дополнительным критерий: результаты эксперимента, в том числе итоговая модель объекта, должны быть представлены в компактной форме, удобной для их опубликования, хранения, сопоставления, применения; интерпретируемы в соответствующей предметной области их пользователя. Третий дополнительный критерий: получение при заданном количестве опытов эксперимента максимального количества данных (результатов, информации) не только для построения функции отклика, описания зависимостей (3.21), но и по всем переменным, входящим в (3.21). Например, чтобы найти их законы распределения, характеристики, исследовать связи и т. п. (ср. по этому критерию разные планы рис. 3.9).

Рассмотрим еще два пути планирования экспериментов, ориентированных на получение «хороших» планов.

Первый путь связан с **последовательными** планированием и проведением эксперимента. Суть его заключается в выборе момента остановки проведения опытов в зависимости от результатов предыдущих опытов эксперимента. Он сводится к использованию пошаговой стратегии как в планировании, так и в построении модели. При обычном (одношаговом, разовом) экспериментировании план (т. е. число N опытов эксперимента и правила их проведения) определяется один раз и на основании этого проводится эксперимент и получают соответствующие результаты (строятся модели, определяются значения, принимаются решения, гипотезы и т. п.). При последовательном планировании и проведении эксперимента после каждого шага реализации плана, т. е. после каждого очередного опыта, производится анализ получен-

ных результатов, на основании чего принимается решение о дальнейшей деятельности: прекратить эксперименты, продолжить их по тому же плану или отменить план. Промежуточными между одношаговым, разовым и последовательным экспериментами являются эксперименты, проводимые по усеченным последовательным планам. В них последовательная процедура проводится не неограниченное число раз, а только до заранее заданного числа N_0 опытов. Если до этого эксперимент не позволил получить желаемые результаты, то они получаются на основе N_0 опытов.

Второй путь связан с идеей *рандомизации* («случаивания») экспериментов. Он ориентирован на желание уменьшить влияние на результаты эксперимента неконтролируемых и неуправляемых факторов W (см. рис. 3.7). Это, например, такие факторы, как квалификация экспериментатора, неизвестные особенности используемых средств и методов измерения, погрешности измерительных приборов и их изменения вследствие старения, неправильной эксплуатации, влияния тех показателей внешней среды, которые мы не знаем как или не можем измерить, учесть. Идея данного варианта повышения качества результатов экспериментов и доверия к ним заключается в организации случайного порядка реализации опытов, т. е. выбора строк матрицы плана.

Матрица плана – принятая форма записи условий проведения эксперимента в виде прямоугольной таблицы размером $(N \times n)$, где N – число строк, количество планируемых опытов, а n – число управляемых факторов X_1, X_2, \dots, X_n . Строки матрицы соответствуют опытам, столбцы – факторам X_1, X_2, \dots, X_n , а в клетках отражены уровни факторов плана.

Матрица плана – это одна форма задания (представления плана). Другая, которая может быть гораздо компактнее, – это задание плана с помощью матрицы спектра плана и матрицы дублирования.

Матрица размером $(L \times n)$, составленная из всех строк плана, отличающихся уровнем хотя бы одного фактора X , называется **матрицей спектра плана**. Ясно, что все строки ее различны. Эту матрицу дополняет **матрица дублирования** размером $(L \times L)$ – квадратная диагональная матрица, в главной диагонали которой указано число N_i параллельных опытов в соответствующих точках спектра плана, т. е.
$$R = \{N_i, i = j; 0, i \neq j, i, j = \overline{1, L}\}.$$

Применение рандомизации позволяет свести систематически действующие неконтролируемые и неуправляемые факторы **W** к описанию их влияния случайными величинами и, следовательно, обрабатывать экспериментальные Данные по этим факторам статистически: находить средние значения, среднеквадратические отклонения, доверительные области и т. п.

§ 3.8. ТЕХНОЛОГИЧЕСКИЕ АСПЕКТЫ ЭКСПЕРИМЕНТИРОВАНИЯ

Экспериментирование на практике является отдельной областью деятельности человечества во многих сферах. Прежде всего это касается физики, химии, биологии, медицины, разработки и производства лекарств и т. д. Существуют различные автоматизированные системы научных исследований и комплексных испытаний, теория их разработки и практика применения. Для специалистов этой области обеспечение позиций получения качественных результатов эксперимента – главная задача. Однако постановщиком задач на эксперимент являются, как правило, прикладные специалисты-исследователи и практики, занятые в своей сфере. Именно они самостоятельно или в сотрудничестве со специалистами-экспериментаторами формулируют глобальные цели и задачи исследования и локальные для эксперимента, определяют показатели качества и то, насколько они достигнуты, интерпретируют эмпирические Данные и результаты и т. д. В связи с этим очень важно иметь представление о технологических процессах исследования хотя бы в общем виде на примере одной-двух прикладных областей, а также о технологических процессах собственно экспериментирования, правил их проведения, обработки и анализа результатов как элементов технологий экспериментирования.

На рис. 3.10 и 3.11 приведены примеры укрупненных схем технологических процессов исследования объектов с помощью их наблюдения (пассивного экспериментирования) и активного эксперимента, на рис. 3.12 – технологического процесса активного экспериментирования, а на рис. 3.13 – элементы методических рекомендаций по экспериментированию, их постановке, проведению, получению, обработке, анализу и интерпретации результатов. Советуем также вернуться к рис. 2.1–2.3 и самостоятельно увязать их с рис. 3.10–3.13.

Из приведенных схем, во-первых, становятся понятными место и роль экспериментов в исследованиях; во-вторых, по умолчанию ясно, что на каждом этапе процесса формируется, уточняется или принима-

ется от текущего этапа постановка задачи на каждый последующий этап. Процесс декомпозиции целей на рисунках не показан. Из рисунков виден также итерационный процесс исследования, предусматривающий возврат по мере необходимости к предыдущим этапам по мере поступления итогов, результатов, полученных на текущем и последующем этапах.

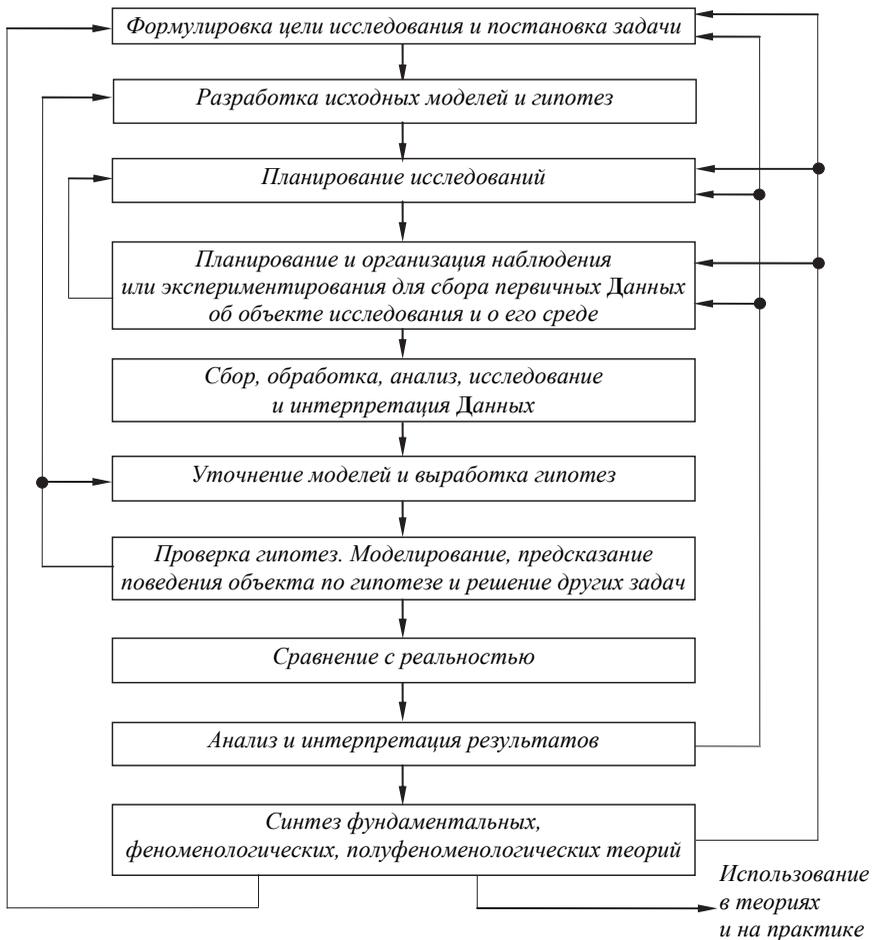


Рис. 3.10. Укрупненная схема технологического процесса в методах исследования, основанных на апостериорных Данных

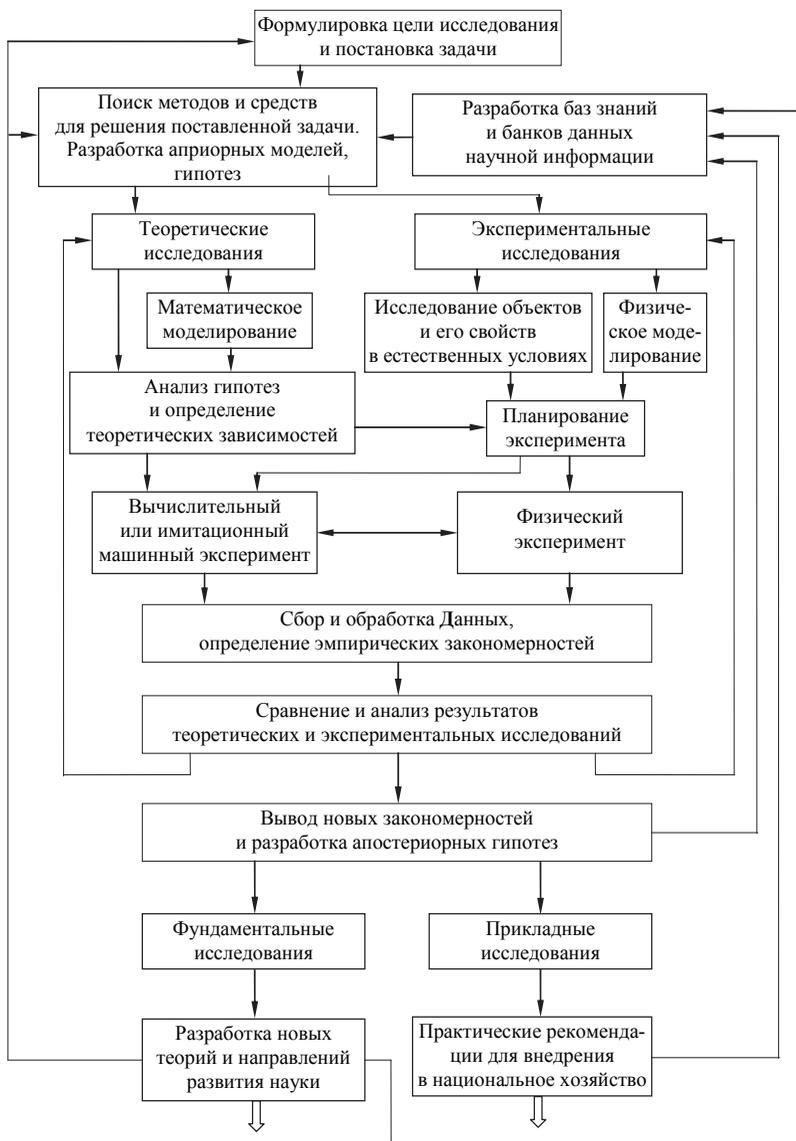


Рис. 3.11. Укрупненная схема технологического процесса в естественно-научных методах исследования



Рис. 3.12. Технологический процесс экспериментального исследования объекта



Рис. 3.13. Технологический процесс планирования натурального эксперимента в условиях априорной неопределенности

§ 3.9. ПРАВИЛА ЭКСПЕРИМЕНТИРОВАНИЯ

Как уже упоминалось, *важным аспектом исследования является соблюдение правил наблюдения и экспериментирования, их проведения, обработки, анализа и интерпретации результатов.* В качестве примера приведем правила анализа результатов наблюдения и экспериментирования, представленные ниже.

Правила анализа результатов наблюдения и экспериментирования

1. Первичное в анализе – постановка задачи: определение целей и задач его, методологии достижения цели.
2. Цели и задачи, используемые методы и средства анализа должны быть согласованы со средой, в которой осуществляется анализ, отвечать системным и специфическим требованиям (т. е. быть ингерентными им).
3. Требования – одна из моделей в постановке задачи анализа.
4. Анализ – итерационный процесс согласования различных этапов его.
5. Допускается и реально реализуется множественность моделей одного и того же объекта (явления, процесса).
6. Отношение эксперимент–модель взаимно, двояко:
 - а) эксперимент → источник информации для моделирования (модели, анализа), проверки и уточнения модели;
 - б) модель → источник организации эксперимента (анализа), она диктует вид эксперимента и анализа.
7. Модель есть способ существования знаний, планов, проектов и т. п.
8. Для детерминированных закономерностей характерна жесткая (механическая) причинность, определяющая поведение каждого элемента (единицы) совокупности. Для статистических закономерностей причинно-следственная связь обуславливается одновременным действием многих причин и отчетливо проявляется только в массовом процессе, свойственна всем совокупностям, а не отдельным экземплярам их. Необходимо иметь в виду, что сущность отражает внутреннюю, а явление – внешнюю сторону действительности. При этом сущность выступает как определяющее, а явление – как определяемое в закономерностях.
9. Анализ данных должен учитывать всех участников: объект, субъект, модель, среду, в которой объект и модель функционируют.
10. Среда это: термины, цели и задачи исследования, исходные и полученные Данные и результаты, измерительные шкалы, виды анализа, используемые модели, характеристики, технологии, алгоритмы, аппаратно-программные и другие средства, условия экспериментирования и функционирования объекта и т. д.
11. Термины (объектные и инструментальные) – базовый компонент моделирования и анализа, модели определяемых ими понятий.

12. Измерения бывают количественные и качественные; прямые, косвенные и совокупные; они неизбежно сопровождаются погрешностями, неопределенностями, могут не снять расплывчатой неполноты сведений об объекте.

13. Данные всегда представлены в определенной измерительной шкале (качественной (номинальной, порядковой), количественной (числовой)), которую следует выбирать максимально сильной, но не завышенной, соответствующей природе исследуемого объекта (явления, процесса).

14. При обработке и анализе данных можно выполнять только такие действия, которые допустимы для исследуемой шкалы. Нарушение этого правила может привести к неправомерной, ложной интерпретации результатов эксперимента.

15. При экспериментировании необходимо стремиться к однородности Данных и условий их получения. Необходимо помнить об относительности однородности: Данные, однородные по одному признаку, полученные для одной цели, могут быть неоднородными по другому признаку, для другой цели и наоборот.

16. При обработке Данных содержащееся в них количество информации не может быть увеличено. Обработка позволяет лишь представить информацию в более компактном, удобном для анализа виде. В лучшем случае – без потери полезной для анализа информации.

17. Для сигналов и данных со стандартным нормальным распределением выбросами можно считать их значения, превышающие порог $3,8 + 0,15(\log_2 N - 6)$, где N – объем выборки. Это гарантирует точность аппроксимации решений не хуже, чем 0,05 в диапазоне $30 \leq N \leq 10000$.

18. Необходимо помнить общее правило решения некорректных (по Адамару¹) обратных задач: не следует стремиться уравнивать левые и правые части систем уравнений точнее, чем это позволяют сделать погрешности получения экспериментальных данных.

19. Необходимо всегда учитывать ограниченность ресурсов (моделей, средств, ...).

20. При получении экспериментальных данных необходимо указывать погрешность, с которой они получаются. При этом значения

¹ Условия корректности задачи по Адамару: существование решения, единственность решения, непрерывная зависимость решения от начальных данных (см. стр. 44).

погрешностей рекомендуется указывать с точностью до одной-двух значащих цифр [12]. Рекомендуются следующие правила [12]: «Если полученное число (погрешности) начинается с цифры, равной или большей, чем 3, то в нем сохраняется лишь один знак; если же оно начинается с цифр, меньших 3, т. е. с цифр 1 и 2, то в нем сохраняются два знака. Результат округляется до того же десятичного разряда, которым оканчивается округленное значение абсолютной погрешности. Округление производится лишь в окончательном ответе, а все предварительные вычисления проводят с одним-двумя лишними значениями».

21. Рассматривая систему объект–признак–значение–уверенность, следует иметь в виду противоположные точки зрения на неполноту информации об объекте: неточность и неопределенность. Неточность ↔ значение. Неопределенность ↔ уверенность. Нечеткость ↔ субъективность. Незнание ↔ осведомленность. Нерепрезентативность ↔ достаточность. Неучтенность ↔ окружающая среда.

22. Данные требуют их исследования, а не только обработки и анализа.

23. Правдивые результаты можно получить, если придерживаться корректного выполнения всех технологических этапов операций (исследования, экспериментирования, сбора, обработки, анализа, интерпретации, применения).

24. Интерпретация результатов исследования обязательна и является одним из приемов повышения устойчивости результатов.

25. Назначение эксперимента – получение данных, позволяющих по ним построить наиболее правдоподобную (а не недостижимую истинную!) гипотезу, модель.

26. В основе любого эксперимента явно или неявно присутствует идея устойчивости результатов к сохранению или нарушению условий его проведения (идеи робастности, инвариантности, адаптивности, ...).

Добавим к этим правилам еще четыре важных комментария.

1. При проведении обработки Данных необходимо помнить следующее. Во-первых, Данные требуют их *исследования*, хотя чаще всего ведут только их обработку. Это связано с тем, что любая формализация, любая математическая обработка, даже самая изошренная и оптимальная, лишь облегчает движение к поиску истины на пути достижения цели исследования, решения проблемы и не может заменить логику и интуицию исследователя. Во-вторых, правдивые результаты можно получить, если придерживаться корректных технологий исследования, экспериментирования, обработки и анализа Данных, правиль-

но соблюдая все этапы. В-третьих, *интерпретация* результатов исследования обязательна и является одним из приемов повышения устойчивости выводов, особенно статистических. Согласно Хэммингу «цель расчетов – не числа, а понимание».

2. Назначение эксперимента – получение Данных, позволяющих по ним построить наиболее правдоподобную (а не недостижимую истинную!) гипотезу, модель строения, свойств, истории или функционирования исследуемого объекта. При этом необходимо иметь в виду, что модельная случайность или детерминированность – лишь возможные абстракции. Между строгой абсолютной случайностью и строгой причинностью лежит континуум возможностей, степеней связи. Полезность любой модели в приложениях не только и не столько в ее способности детально описать реальность в определенных условиях, сколько в ее способности дать приемлемые практические (в частности, инженерные) решения и прогнозы, в степени ее возможности помочь достичь цели исследования, получить полезный результат. Например, когда вероятностная модель позволяет не только подтвердить детерминированную закономерность, но и охарактеризовать необъяснимую (случайную, хаотическую) изменчивость, наблюдаемую при этом в физических явлениях.

3. В основе любого эксперимента явно или неявно присутствует идея устойчивости (цели исследования, состояния и поведения объектов, условий получения Данных, самих Данных, моделей и методов их обработки, результатов и выводов) к нарушению условий. Поэтому всегда надо анализировать решения и выводы на устойчивость в тех пределах, которые можно гарантировать при проведении эксперимента, учитывать возможную неустойчивость, измерительные шкалы и их преобразования, способ усреднения, засорения выборки, вводить корректировки или использовать робастные (т. е. устойчивые на классе моделей, условий, шкал) методы, быть аккуратным с оптимальными методами. В частности, использовать для этого приемы проверки допустимости каждого метода, модели и их пригодности в условиях получения Данных, проверять устойчивость оценок по возрастающей серии наблюдений и определять надежность результатов по многим сериям, вносить рандомизацию, формируя псевдовыборку, близкую к исходной, но с «разрушенной» искомой зависимостью, используя бутстреп-процедуры, приемы «складного ножа», другие приемы рандомизации, а также разные приемы планирования экспериментов.

4. Необходимо всегда иметь в виду возможность управления качеством результатов экспериментирования и решения поставленных задач [5].

5. Интерпретируя итоги анализа результатов экспериментальных исследований, необходимо всегда помнить, что *отсутствие доказательства чего-то* (например, связи, наличия мешающих или возмущающих факторов, включаемых в модельные соотношения, и т. п.), *не является доказательством отсутствия этого чего-то*.

ЗАКЛЮЧЕНИЕ

В главе даны основные понятия экспериментирования как метода исследования и одного из важнейших этапов моделирования объектов разной природы. Приведены основные понятия: величины; измерение, оценивание, вычисление, исчисление; идентификация; измерительные шкалы и др. Особое внимание уделено различию между этими понятиями (в частности между измерением и оцениванием), особенностям работы с данными, представленными в разных шкалах, а также основным метрологическим терминам.

Рассмотрены элементарные системные основы теории эксперимента: базовые понятия; разновидности экспериментов, получаемых в ходе экспериментов данных и построенных на их основе моделей; неполнота априорных и апостериорных Данных об исследуемом объекте; вопросы планирования экспериментов, технологические аспекты экспериментирования.

Следующий круг затронутых вопросов – математическое представление, временные и частотные характеристики сигналов, как детерминированных функций времени, разновидности сигналов.

ВОПРОСЫ ДЛЯ САМОПОДГОТОВКИ

1. Что такое наблюдение и экспериментирование? Охарактеризуйте их: сходство и отличие, положительные и отрицательные особенности в применении.

2. Что такое величина: физическая, математическая; аналоговая и цифровая?

3. Что понимается под экстенсивными и интенсивными физическими величинами? Когда и почему необходимо их различать?

4. Приведите определения и охарактеризуйте следующие термины: измерение, вычисление, исчисление; измерительная шкала и ее сила;

идентификация; оценивание и оценка; имитация, симуляция; генезис, диагностика, прогнозирование; метод, алгоритм.

5. Укажите разновидности измерительных шкал и допустимые действия в них.

6. Приведите вариант классификации и охарактеризуйте разновидности экспериментов.

7. Каковы разновидности объектов с точки зрения их модельного представления?

8. Укажите классы моделей, используемых для описания объектов различной природы.

9. Каковы разновидности экспериментальных данных?

10. Что такое неполнота априорных и апостериорных Данных об объекте? Виды неполноты.

11. Измерение и метрология. Основные рекомендуемые метрологические термины.

12. Погрешностный и неопределенностный подходы представления результатов измерений: сходство и отличие.

13. Что понимается под временным и частотным представлением сигналов? В чем их сходство, различие, преимущества?

14. Что такое функция, функционал, оператор?

15. Что такое дискретизация и квантование сигналов? Что понимается под термином *отсчет сигнала*?

16. Запишите оператор усреднения функции по аргументу (оператор усреднения по времени). Сравните его с оператором усреднения по траектории (реализации выборочной функции) случайных процессов и последовательностей (см. главу 4).

17. Приведите основные временные характеристики детерминированных сигналов.

18. Приведите основные частотные характеристики детерминированных сигналов.

19. Приведите основные соотношения теории детерминированных сигналов: равенство Парсеваля, соотношение неопределенности, связь между спектрами непрерывной и дискретизированной функциями, спектр произведения и свертки функций.

20. Что такое планирование наблюдений и экспериментов?

21. Что означает термин *фактор* в теории экспериментов, его уровень, область действия?

22. Что понимается под входными, контролируемыми, управляемыми и выходными переменными объекта в теории эксперимента?

23. Что такое отклик, зависимая переменная, функция и поверхность отклика в теории эксперимента?

24. В чем различие между планированием эксперимента в широком и узком смысле слова *планирование*? В чем сходство и отличие целей, планов эксперимента в том и другом случае?

25. Укажите и охарактеризуйте виды экспериментов по их назначению.

26. Приведите примеры последовательностей этапов физического, вычислительного, аналитического (расчетного) и алгоритмического (имитационного) экспериментов.

27. Что понимается под критерием качества эксперимента, его плана, результатов?

28. Нарисуйте схему технологического процесса наблюдений и экспериментирования.

29. Укажите важные, на ваш взгляд, технологические аспекты и правила экспериментирования и анализа результатов анализа и экспериментирования.

Глава четвертая
ФОРМАЛЬНЫЙ АППАРАТ ОПИСАНИЯ
И ИССЛЕДОВАНИЯ СТРУКТУР, СОСТОЯНИЙ
И ПОВЕДЕНИЯ ОБЪЕКТОВ:
КУОМОДНЫЕ АНАЛИТИЧЕСКИЕ МОДЕЛИ

§ 4.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

В предыдущих главах мы уже сталкивались с многообразием моделей, разработанных и используемых для описания и *исследования*¹ естественных и искусственных объектов различной природы (см., например, § 2.1–2.3 и рис. 3.4). При этом также обратили внимание на большое многообразие самих объектов с точки зрения их модельного представления (см. табл. 2.4 и рис. 3.3). Можно посчитать число возможных комбинаций объект–модель. Как уже упоминалось, вопрос выбора модели под конкретный объект, функционирующий в определенных условиях, должен всегда решаться с учетом цели моделирования и задач исследования объекта. Однако некоторые общие рекомендации по пригодности классов моделей необходимо знать с точки зрения не только их практического применения, но и выбора образовательных траекторий, особенно когда сходные математические («модельные») дисциплины предлагается изучать по выбору. Именно на этом акцентируется внимание в настоящей главе. В ней дается краткое перечисление и, по мере возможности и необходимости, пояснение классов, подклассов и отдельных видов моделей, ставятся задачи и определяются ограничения по их использованию.

Перейдем к рассмотрению различных формальных, прежде всего аналитически задаваемых и описываемых, моделей, сгруппировав их по соответствующим разделам математического аппарата исследований.

¹ Напомним, что слово *исследование* понимается в обобщенном виде (см. § 1.1).

§ 4.2. ДЕТЕРМИНИРОВАННЫЙ АППАРАТ. ЗАКОНОВЫЕ МОДЕЛИ

Эта разновидность моделей базируется на классическом детерминированном математическом аппарате. Согласно рис. 3.3 предпосылками его практического применения для исследования естественных и искусственных объектов является наличие причинно-следственных связей в функционировании и поведении объекта, его связи с окружающей средой. Они должны проявляться в виде «законов» – обязательных для всех объектов рассматриваемого вида, встроенных в ту же (условно близкую) среду и функционирующих в одинаковых условиях, повторений структур объекта, его особенностей, связанных с ним событий, процессов, состояний, явлений, значений параметров и характеристик, отражающих объект, и т. д.

Многообразии таких моделей рассматривается в «классических» разделах математики:

- арифметика;
- алгебра; алгебраические уравнения и их системы;
- геометрия;
- тригонометрия;
- аналитическая и дифференциальная геометрия;
- дискретная математика (математическая логика, исчисление высказываний; теории графов, кодирования, алгоритмов, автоматов; сетей и потоков и т. п.);
- теория чисел, включая комплексные числа, и числовых пространств;
- теория функций, функционалов, операторов и функциональных пространств (см. § 3.6):
 - **функция** – отображение одного множества чисел в другое числовое множество (примеры – \sin , \cos , \lg , ...);
 - **функционал** – отображение множества функций в множество чисел, функционального пространства – в числовое (пример – определенные интегралы, некоторые специальные функции, рассматриваемые вдоль множества значений параметров);
 - **оператор** – отображение множества функций в другое множество функций, одного функционального пространства – в другое (неопределенный интеграл, производная);
- дифференциальное и интегральное, операторное исчисление; интегродифференциальные уравнения и их системы;

- векторные исчисления, теория поля; тензоры;
- разложение и восстановление функций, теория числовых и функциональных рядов; частотный, спектральный анализ функций;
- топология;
- вычислительная математика, численные методы;
- другие разделы детерминированной математики.

Предпосылки практического применения детерминированного аппарата – «законно» повторяющиеся события, процессы, состояния, явления, значения параметров исследуемых объектов.

В качестве примера более детального разнообразия детерминированных моделей рассмотрим разновидности функциональных¹ моделей. Их многообразие отражено на рис. 4.1.

Представляется, что читатель знаком со многими видами функций, указанных на рис. 4.1, из школьного курса или первокурсных разделов высшей математики, посвященных теории функций. Поэтому ограничимся лишь небольшими комментариями.

1. Здесь и далее обсуждаются только точечные (см. табл. 2.3 и рис. 3.4) модели. Интервальные модели будут кратко рассмотрены на конкретном примере в § 4.9.

2. Не следует отождествлять детерминированные модели, условно названные на рис. 4.1 «катастрофическими», с моделями, используемыми в теории катастроф (см. § 4.6). Под прямоугольником «катастрофические» на рис. 4.1 представлены качественные графики плотности W распределения вероятностей, называемого гамма-распределением,

¹ Слово *функциональные* является производным от термина *функция* как математического понятия (см. § 3.1), а сочетание *функциональные модели* понимается как модели, представимые в виде математических функций. Очень часто такое сочетание заменяется другим – *функциональные модели*. Автор сознательно избегает этого словосочетания по двум причинам. Первая причина – чтобы отличить функциональные модели, представимые в виде функций переменных, от функциональных моделей как моделей, представимых функционалами. Вторая причина – сочетание *функциональные модели* в теории моделирования объектов очень часто используется в трактовке *любые*, а не только детерминированные, *модели*, предназначенные для описания функций (назначений, решаемых задач) элементов объекта, т. е. модели функций (а не модели в виде функций) в отличие от структурных, классификационных, поведенческих и других моделей объектов.

$$W(x; \alpha) = f(x; \alpha) = \begin{cases} 0, & x < a; \\ \frac{(x-a)^{\alpha-1}}{\lambda^\alpha \Gamma(\alpha)} \exp\left\{-\left(\frac{x-a}{\lambda}\right)^\alpha\right\}, & x \geq a, \end{cases} \quad (4.1)$$

где x – аргумент; a – параметр сдвига; λ – параметр масштаба; α – параметр формы распределения; $\Gamma(\alpha)$ – специальная гамма-функция [5]; $\exp\{\cdot\}$ – экспоненциальная функция. Из (4.1) видно, что при $\alpha = 0$ график $W(x)$ представляет собой график экспоненциально затухающей кривой e^{-x} , где $e = 2,718282$ – основание натурального алгоритма, при $\alpha \neq 0$ «катастрофически» меняется, переходя в график L-образной кривой при $\alpha < 0$ и «одногогорбой» унимодальной кривой при $\alpha > 0$. Именно в этом смысле используется слово «катастрофические». Вполне допустимо, что некоторые процессы «механизма» образования случайности, описываемой гамма-распределением, могут иметь катастрофический характер в смысле, используемом в теории катастроф (см. § 4.6).

3. Функции $f(x)$ с участками быстрого ускорения – это такие, которые в основной части области определения аргумента x изменяются сравнительно медленно, а в особых точках x_0 асимптотически устремляются к бесконечности. Такие модели используются для описания режимов функционирования объектов, получивших название «режимы с обострением»¹. Это режимы, имеющие длительную, близкую к стационарной стадию функционирования, развития, переходящую в стадию сверхбыстрого нарастания процессов в открытых нелинейных системах. Пример такой априорной функции:

$$m(\vartheta) = m_0 / \sqrt{1 - \vartheta^2/c^2}, \quad 0 \leq \vartheta \leq c, \quad (4.2)$$

описывающей массу $m(\vartheta)$ движущейся микрочастицы, имеющей в покое массу m_0 , со скоростью ϑ , когда скорость ее движения ϑ приближается к скорости света в вакууме c . Второй пример – гиперболические функции. Например, апостериорная функция вида

$$N(t) = N_0 / (1 - t/t_0), \quad (4.3)$$

¹ Обострение – по-англ. blow up.

которую ряд исследователей (С.П. Капица, С.П. Курдюмов и др.) используют для описания эмпирической кривой зависимости от времени t численности N людей на Земле (см. рис. П11.3 в [1]), где t – время в годах, $t_0 \approx (2025 \text{ г.})$, $T_0 \approx 10^8$.

Внимание! Проверьте справедливость применимости (4.3) по рис. П11.3 в [1], учитывая, что семимиллиардный житель Земли появился 31.10.2011 г.

§ 4.3. СТОХАСТИЧЕСКИЙ (ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЙ) АППАРАТ. МОДЕЛИ ОБЪЕКТИВНО ОБНАРУЖИВАЕМЫХ ЗАКОНОМЕРНОСТЕЙ

4.3.1. Условия применимости. Постановки задач

Стохастический аппарат – это разновидность математического аппарата, оперирующего со стохастическими аналитическими моделями.

Согласно табл. 2.3 и рис. 3.3 стохастические модели, во-первых, делятся на две разновидности – вероятностные и статистические; во-вторых, ориентированы на применение для описания и исследования объективно возможных закономерностей, отличающихся рядом особенностей.

Вероятностный аппарат (теория вероятностей)¹

В1. Вероятностные (априорные) модели (их разновидности, свойства, характеристики):

- случайные события A, B ;
- случайные величины X и векторы (многомерные величины) $\mathbf{X} = (X_1, X_2, \dots, X_n)$;
- скалярные $X(t)$ и векторные (многомерные) $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_n(t))$ случайные функции одного (процессы

¹ Теория вероятностей – раздел математики, связанный с введением и изучением вероятностных моделей, их свойств, характеристик, преобразований.

$\mathbf{X}(t)$, последовательности $\mathbf{X}(t_1), \mathbf{X}(t_2), \dots; \mathbf{X}(t), \mathbf{X}(t + \Delta t), \mathbf{X}(t + 2\Delta t) \dots$) и многих (поля $\mathbf{X}(\alpha_1, \alpha_2, \dots, \alpha_k)$) аргументов.

В2. Преобразования случайных элементов (величин, векторов, функций):

$$Y_1 = f_1(X_1, X_2, \dots, X_n), \quad k = 1, 2, \dots; \quad n = 1, 2, \dots;$$

⋮

$$Y_k = f_k(X_1, X_2, \dots, X_n), \quad \text{когда } k < n; \quad k = n; \quad k > n.$$

В3. Стохастические уравнения, в том числе интегродифференциальные; операторы; определители; геометрии; топологии и т. п.

В4. Неравенства и предельные теоремы теории вероятностей

В5. Вероятностные аналоги других разделов детерминированной математики.

Статистический (апостериорный) аппарат (математическая¹ и прикладная² статистика)

С1. Нахождение *числовых* значений статистических (эмпирических) или *количественное* оценивание вероятностных характеристик модели \mathcal{M} по имеющимся числовым данным – «эмпирическим» значениям или значениям случайных элементов (по выборкам из генеральной совокупности возможных значений x_1, x_2, \dots, x_N).

С2. Статистическая проверка (*качественная*) гипотез: соответствует ли имеющийся (с его свойствами) набор значений (выборка) x_1, x_2, \dots, x_N модели \mathcal{M} либо наоборот модель \mathcal{M} – набору.

С3. Статистические выводы: формулировка выводов о свойствах модели \mathcal{M} и ее характеристик, об их закономерностях, о связях и т. п. по имеющемуся набору значений x_1, x_2, \dots, x_N .

¹ Математическая статистика – раздел математики, связанный с построением и изучением статистических моделей, их разновидностей, свойств, характеристик по значениям случайных элементов (величин, векторов, функций), полученных в четко определенных (идеализированных) условиях.

² Прикладная статистика – раздел математической статистики, направленный на работу со значениями x_1, x_2, \dots, x_N в условиях, отличных от идеализированных, приближенных к реальным.

Теория статистической имитации

Раздел математики, связанный с построением аппарата получения значений x_1, x_2, \dots, x_N модели с заданными свойствами в следующих ситуациях:

- модель \mathcal{M} задана априори (вероятностная модель);
- апостериорное задание модели (статистическая модель);
- по ранее полученным или имеющимся значениям y_1, y_2, \dots, y_L , описываемым той же моделью, с теми же важными для исследования свойствами.

Рассмотрим эти разновидности, для чего вначале введем несколько необходимых понятий.

Согласно рис. 3.3 стохастические – это эксперименты, для которых выполняются следующие два необходимых (но не достаточных!) условия (требования стохастичности).

1. Они могут быть сколь угодно (теоретически, для мысленных вероятностных экспериментов, бесконечное множество раз) воспроизведены при одних и тех же условиях U (реально в пределах допустимых для решаемой задачи вариациях условий).

2. Хотя априори (до опыта) исходы эксперимента неизвестны, апостериори (после проведения мысленного или физического опыта) в результате многочисленных опытов обнаруживается следующая объективная закономерность (тенденция) в появлении исходов опыта (событий): частость (частота¹) $P^*(A)$ появления события A при неограниченном увеличении числа N проведенных опытов сходится к некоторой константе $p(A)$ в модельном описании данного эксперимента. Иными словами, для стохастических объектов (экспериментов) должен существовать (а для физических экспериментов иметь физический смысл) предел

¹ Слово *частота* в данном контексте используется в публикациях по теории вероятностей и математической статистике значительно чаще, чем слово *частость*. Однако поскольку в книге термин *частота* используется в спектральном анализе в другом понимании, во избежание путаницы будет в приложении к вероятностно-статистическим моделям использоваться для $P^*(A)$ именно термин *частость*.

$$\lim_{N \rightarrow \infty} P^*(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N} = p(A), \quad (4.4)$$

где $P^*(A) = N_A/N$ – частота события A ; N_A – число (количество) появлений события A в N опытах; $p(A)$ – константа.

В вероятностных и статистических моделях эта константа $p(A)$ принимается за конкретное значение $p(A)$ вероятности $P(A)$ случайного события A .

Именно такие исходы опытов подобных стохастических экспериментов (события A) называют *случайными*. Именно на описание таких объектов и экспериментов ориентирован стохастический математический аппарат.

Предпосылки практического применения стохастического аппарата – выполнение всех «закономерностных» условий стохастичности исследуемого объекта + вероятностная формулировка постановки задачи исследования.

Чтобы стохастический аппарат можно было корректно применять для описания объектов (экспериментов) и правильно интерпретировать полученные с его помощью результаты, мало, чтобы объект был стохастическим. Необходима еще соответствующая стохастическая постановка задачи. Понятно, что она делается под конкретное исследование и в конкретных условиях. Как мы уже указывали, стохастические аппарат и модели делятся на вероятностные и статистические. Отличие в формулировке постановок задач в них и подсказки по их содержанию наглядно видны из табл. 4.1. Заметим, что правый столбец табл. 4.1 касается имитационных статистических экспериментов.

Перейдем к более детальному рассмотрению этих отличий. Начнем с вероятностных моделей как базовых теоретических для всего стохастического аппарата.

Вероятностные – это аналитические математические модели, основанные на использовании в качестве меры $\mathbf{P}\{A\}$ возможности появления событий A в опытах *распределения вероятности* на полном множестве событий A .

Они изучаются в разделе математики – теории вероятностей, где такие модели вводятся (создаются, придумываются, строятся) и изучаются, включая исследование их свойств, характеристик, преобразований.

Постановки стохастических задач

Формулировки	Используемый аппарат		
	Вероятностный (априорный)	Статистический (апостериорный)	Статистический имитационный
Дано	Объект (эксперимент) «О», удовлетворяющий требованиям стохастичности		
Дано допол- нительно		Набор (последова- тельность) конкрет- ных значений, при- писываемых вероят- ностной модели \mathcal{M} (величин, векторов, функций) фиксир- ованного объема, полученных в задан- ных условиях (иде- альных – математи- ческая статистика или приближенных к реальности – при- кладная статистика)	Вероятностная или статистическая мо- дель \mathcal{M} либо набор их значений объ- емом N_1
Необходимо	Априори (до опыта) постро- ить (вероятност- ную) модель \mathcal{M} объекта и ее ха- рактеристики, исследовать их свойства, осо- бенности, пре- дельные случаи, поведение при функциональных преобразованиях и т. п.	Построить стати- стическую (в част- ности апостериор- ную вероятностную) модель объекта «О» по этому набору значений и/или сде- лать по набору вы- воды о свойствах и характеристиках модели	Получить (сымити- ровать) новый набор (последовательность с заданными свой- ствами) значений этой модели объема N_2
Результат	Описание моде- ли, ее характе- ристик и их свойств, аппарат работы с ними	Статистическая (эмпирическая или оценка вероятност- ной) модель (ее ха- рактеристики, пара- метры) и/или выво- ды о свойствах ее и объекта	Набор (последова- тельность с задан- ными свойствами) значений, соответ- ствующих модели \mathcal{M} или прошлому набору

Соответствие между физическими и математическими объектами

№ п/п	Сопоставимые понятия	Результаты физического эксперимента качественные	количественные	Вероятностное описание качественное	количественное
1	Множество И все возможных исходов, значе- ний «и»	Множество G всевоз- можных неразложимых взаимоисключающих исходов «g»	Множество Z всевоз- можных значений (реа- лизаций) параметров z объекта	Множество Ω эле- ментарных случай- ных событий ω	Метрическое числовое или функциональное пространство \mathcal{X}
2	Множество и фактических ис- ходов, значений	Множество $\mathcal{G} \in G$	Множество $Z \in Z$	Борелевское (под) множество $A \in \Omega$	Борелевское (под) множество $\mathcal{X} \in \mathcal{X}$
3	Совокупность И подмножеств множества ис- ходов И	Совокупность \mathcal{F} все- возможных $A \in G$, включая $A = G$	Совокупность Z все- возможных подмножеств $B \in Z$, включая $B = Z$	Борелевская σ - алгебра \mathfrak{G} подмно- жеств $A \in \Omega$	Борелевская σ -ал- гебра \mathcal{B} подмно- жеств $\mathcal{B} \in \mathcal{X}$
4	Совокупность (И, и)	Совокупность (состоя- ний) (G, \mathcal{F})	Совокупность (реализа- ции) (Z, Z)	Измеримое про- странство (основ- ное) (Ω, \mathfrak{G})	Измеримое простран- ство (выборочное) состояний $(\mathcal{X}, \mathcal{B})$
5	Мера μ воз- можности появ- ления исходов, значений	Частотная мера \mathbf{P}^* (частость событий $\mathbf{P}^*(A)$)	Частотная мера μ^* (распределение частот)	Вероятностная мера $\mathbf{P} > 0$ на (Ω, \mathfrak{G}) , $\mathbf{P}\{\Omega\} = 1$	Вероятностная мера $\mu \geq 0$ (закон распре- деления) такая, что $\mu(\mathcal{X}) = 1$
6	Совокупность (И, и, μ)	Совокупность $(G, \mathcal{F}, \mathbf{P}^*)$ (частостное пространство)	Совокупность (Z, Z, μ^*) (частостное про стран- ство)	Вероятностное про- странство $(\Omega, \mathfrak{G}, \mathbf{P})$ (основное)	Вероятностное про- странство $(\mathcal{X}, \mathcal{B}, \mu)$ (выборочное, состоя- ний)
7	Предмет иссле- дования, описа- ния	Состояние объекта	Значение физической величины – параметра объекта	Случайное событие A	Случайный элемент \aleph

В отличие от вероятностных статистические – это модели, получаемые с использованием аппарата математической и прикладной статистик. Заметим, что деление условно и лишь подчеркивает, как вводится, строится, используется, интерпретируется модель. По сути же в основе этих моделей находится вероятность: теоретическая в вероятностных моделях и эмпирическая (частость) в стохастических, а также сходство многих свойств характеристик аналогов.

В основе вероятностных моделей лежит идея формализации реальности (экспериментов) на качественном уровне с помощью теории случайных событий, а на количественном – с помощью аппарата случайных элементов (величин, векторов, функций и т.п.) (см. далее).

Соответствие между физическими и вероятностными объектами представлено в табл. 4.2 и 4.3.

Т а б л и ц а 4.3

Примеры случайных элементов

Пространство \mathcal{X}	Случайный элемент \aleph	
	Обозначение	Наименование
Одномерное вещественное пространство R^1 (числовая ось $(-\infty, +\infty)$)	X	Случайная величина
Пространство R^n $R^n = R^1 \times \dots \times R^1$	${}_n \mathbf{X}, \mathbf{X} = (X_1, X_2, \dots, X_n)$	n -мерный случайный вектор
Пространство F^1 функций непрерывного аргумента t	$X(t)$	Случайная функция непрерывного аргумента, случайный процесс, если t – время
Пространство \mathcal{F}^1 функций дискретного аргумента	$X(i), X(i\Delta t)$	Случайная последовательность (функция дискретного аргумента), временной ряд, если i – номер временного интервала, $\Delta t > 0$
Функциональное пространство F^n	${}_n \mathbf{X}(t),$ $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))$	n -мерная (векторная) случайная функция (процесс)
Функциональное пространство \mathcal{F}^n	${}_n \mathbf{X}(i),$ $\mathbf{X}(i) = (X_1(i), \dots, X_n(i))$	n -мерная (векторная) случайная последовательность (временной ряд)

Напомним, что *множество* – это совокупность, набор каких-либо объектов, называемых его *элементами*, обладающих общими для всех них характеристическими свойствами, особенностями. *Метрическое пространство* – это множество с заданной для всех его элементов, называемых *точками* пространства, мерой расстояний между ними. *Пространство событий, состояний* (точек пространства) – это их множество, характеризуемое некоторыми формализованными описаниями. Например, пространственными координатами x, y, z и временем t ; совокупностью подмножеств и мерой принадлежности к каждому подмножеству. *Фазовое пространство физической системы* – это множество всех ее возможных как-то формально описанных состояний.

4.3.2. Виды стохастических экспериментальных Данных

Прежде чем рассмотреть примеры задач, решаемых с использованием стохастического аппарата, рассмотрим, какие виды стохастического анализа экспериментальных данных используются чаще всего. Они представлены на рис. 4.2.



Рис. 4.2. Виды стохастического анализа Данных

4.3.3. Дескриптивный анализ

Дескриптивный (дескрипторный, описательный) анализ – простейший вид синтаксического анализа экспериментальных Данных, основанный на применении простейших *дескриптивных* (описательных) статистик. В частности эмпирических (статистических, выборочных аналогов) вероятности $\mathbf{P}(A) = \mathbf{M}\{\mathbf{I}(A)\}$ появления события A , где $\mathbf{I}\{A\}$ – индикатор появления события A в эксперименте, т. е. $\mathbf{I}\{A\} = 1$, если A появляется в эксперименте, и $\mathbf{I}\{A\} = 0$, если A не появляется в данном эксперименте, а имеет место событие \bar{A} , противоположное событию A ; математического ожидания $m_X = m_1 = \mathbf{M}\{X\}$, где $\mathbf{M}\{\cdot\}$ – оператор усреднения по вероятностной мере случайной величины X (иначе, оператор математического ожидания) (!) (см. прил. 1); дисперсии $\mathbf{D}\{X\} = D_X = \mathbf{M}\left\{[X - m_X]^2\right\}$ и среднеквадратического отклонения (СКО) $\sigma_X = +\sqrt{D_X}$; квантили порядка p $x_p = \arg[F_X(x) = p]$, где $F_X(x)$ – функция распределения случайной величины X , в частности *медианы* $x_{0,5}$, *квартилей* $x_{0,25}$, $x_{0,75}$, *децилей* $x_{0,1}$, $x_{0,2}$, ..., $x_{0,9}$, *процентилей* $x_{0,01}$, $x_{0,02}$, ..., $x_{0,99}$; *интерквартильной широты* $E_{p,q} = x_q - x_p$, $q > p$ (например, $p = 0,25$ и $q = 0,75$); *срединного отклонения*, определяемого соотношением $\mathbf{M}\{|X - m_X| \leq E_X\} = 1/2$; параметров сдвига $\mathbf{a} = (a_1, \dots, a_n)$, масштаба $\lambda = (\lambda_1, \dots, \lambda_n)$, формы $\alpha = (\alpha_1, \dots, \alpha_n)$, $\alpha_j = (\alpha_{1,j}, \dots, \alpha_{k,j})$, $j = \overline{1, n}$, и связи $\Phi = (\varphi_1, \varphi_2, \dots, \varphi_{n-1})$ случайных величин X_1, X_2, \dots, X_n , образующих n -мерный случайный вектор ${}_n\mathbf{X} = (X_1, \dots, X_n)$; моментных, квантильных, кумулянтных и иных коэффициентов асимметрии и эксцесса распределения (случайных величин, векторов, функций) (см., например, [20, 21]).

Цель дескриптивного анализа – численное, аналитическое описание статистики (распределения мгновенных значений) Данных.

Сделаем важные замечания по этому виду анализа.

1. Прежде всего обратим внимание, что помимо указанных выше вероятностных характеристик как характеристик генеральной совокупности рассматривают их статистические аналоги, найденные по выборке объема N . Например, среднее (среднеарифметическое) выборки

$\bar{x} = (x_1 + x_2 + \dots + x_N) / N$, где $x_i, i = \overline{1, N}$, i -е выборочное значение (значение, которое принимает X в i -м опыте), выборочная дисперсия (вариация выборки) $S^2 = \sum_{i=1}^N (x_i - \bar{x})^2 / (N - 1)$ и СКО выборки $S = +\sqrt{S^2}$; χ_p – p -я выборочная квантиль, равная $\chi_p = x_{(Np)}$, если (Np) – целое число, или $\chi_p = [x_{E\{Np\}} + x_{1+E\{Np\}}] / 2$, если (Np) – нецелое число, где $E\{Z\}$ – целая часть числа Z ; $x_{(1)}, x_{(2)}, \dots, x_{(N)}$ – элементы вариационного ряда – неубывающего ряда, полученного по выборке $\chi = (x_1, x_2, \dots, x_N)$, когда $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$; мода, антимода выборки и т. д. При этом статистики \bar{x}, S^2, S, χ_p и им подобные могут выступать как объективные детерминированные¹ характеристики выборки (тогда они называются выборочными характеристиками) либо как (случайные)² оценки \hat{Q} характеристик Q генеральной совокупности. В качестве характеристик Q при этом принимаются вероятностные характеристики модели – случайной величины X , выступающей как модель исследуемого в конкретных условиях объекта, т. е. как модель всей генеральной совокупности, получаемой для данного объекта в конкретных условиях его функционирования и экспериментирования с ним. Во втором случае мы получаем $\bar{x} = \hat{m}_X, S^2 = \hat{D}_X, S = \hat{\sigma}_X, \chi_p = \hat{x}_p$ и т. д.

2. Не все подобные вероятностные характеристики и, следовательно, их соответствующие выборочные аналоги – дескриптивные статистики – существуют³. Например, для некоторых распределений (Коши;

¹ Объективные детерминированные в том смысле, что их значения однозначно определяются апостериорными элементами выборки $\chi = (x_1, x_2, \dots, x_N)$.

² Случайные в том смысле, что априори их значения не определены, хотя апостериори определяются детерминированно, однозначно.

³ Заметим, может оказаться, что дескриптивная статистика как характеристика выборки существует, имеет смысл при разных N , хотя ее модельный аналог не существует. Это означает, что принятое модельное представление объекта некорректное, не адекватное объекту и/или условиям экспериментирования с ним. Тогда надо заменить модель.

Стьюдента и бета-II, Пирсона IX, Парето, Бёрра-II при некоторых соотношениях параметров формы, Пирсона V, Хольцмарка и др. [20]) не существуют моментные характеристики, даже начиная с математического ожидания m_1 ; для дискретных распределений квантиль может представлять собой не точку (иметь точечное значение), а интервал. В этих случаях необходимо либо переходить к другим характеристикам аналогичного назначения, сохраняя модель объекта (генеральной совокупности) либо менять модель на такую, адекватную выборке, для которой все нужные характеристики существуют.

3. Для решения одной и той же задачи могут использоваться разные характеристики. Например, для описания (определения) уровня, вокруг (около) которого флюктуируют (расположены, рассеяны) значения случайной величины, можно использовать математическое ожидание, квантиль x_p , в частности медиану $x_{0,5}$, параметр a положения распределения (или спектра (см. [20])) типа сдвига, когда $F(x; a) = F(x - a; 0)$; усечения слева, когда $F(x; a) = \{0 \text{ при } x \leq a; p > 0 \text{ при } x > a\}$; усечения справа, когда $F(x; a) = \{p < 1 \text{ при } x < a; 1 \text{ при } x \geq a\}$; мода $\text{moda } X = \left\{ x: W(x), P(x) = \max_x \right\} = \arg \max_x \{W(x), P(x)\}$, антимода $\text{amoda } X = \arg \min_x \{W(x), P(x)\}$ [20].

Здесь $F(x)$, $W(x)$, $P(x)$ – функция, плотность или ряд распределения вероятностей случайной величины X (см. подпараграф 4.3.4). Следовательно, для характеристики уровня выборки можно использовать соответствующие выборочные аналоги: \bar{x} , χ_p , $\chi_{0,5}$ и т. п.

Для описания рассеяния значений случайной величины можно использовать СКО σ , параметр масштаба распределения (случайной величины) λ , определяемый соотношением $F(x; \lambda) = F(x/\lambda; 1)$; $W(x; \lambda) = \frac{1}{\lambda} W(x/\lambda; 1)$; срединное отклонение E_X , срединное отклонение относительно медианы \mathcal{E}_X , определяемое соотношением $F_X(x_{0,5} + \mathcal{E}) - F_X(x_{0,5} - \mathcal{E}) \leq \frac{1}{2}$, среднее абсолютное отклонение от среднего $\mu_{||} = \delta_1 = \mathbf{M}\{|X - m_1|\}$, среднее абсолютное отклонение от

медианы $\delta_2 = \mathbf{M}\{|X - x_{0,5}|\}$, интерквартильную широту $|x_q - x_p|$, $q > p$, $q - p \approx 1$, в частности интерквартильную широту $|x_{0,75} - x_{0,25}|$, среднюю разность Джини $\delta_3 = \mathbf{M}\{|X_1 - X_2|\}$, где X_1 и X_2 – независимые случайные величины с одинаковым распределением $F(x)$, $W(x)$, $P(x)$, и др. [20], а также безразмерные характеристики рассеяния типа [20]: коэффициента вариации Пирсона $d_1 = \sigma/|m_1|$, $m_1 \neq 0$; коэффициента рассеяния Джини $d_2 = \delta_3/(2|m_1|)$, $m_1 \neq 0$; $d_4 = |(x_q - x_p)/x_{0,5}|$, $1 \geq q > p > 0$, $x_{0,5} \neq 0$, в частности $d_5 = |(x_{1-p} - x_p)/x_{0,5}|$, $0 < p < 0,5$, $d_6 = |(x_{0,75} - x_{0,25})/x_{0,5}|$, когда x_p – точка, а не интервал, коэффициент пиковости $d_6 = |x_q - x_p|/\sigma_X$, $1 \geq q > p > 0$.

Для характеристики формы распределения можно использовать параметры α формы распределений, моментные и квантильные коэффициенты асимметрии, коэффициенты асимметрии Пирсона, моментные и квантильные коэффициенты эксцесса, показатели полимодальности и т. п. (см. [20]). Аналогичное, как увидим далее, справедливо для характеристик связи случайных векторов и функций.

Понятно, что для описания выборок можно использовать выборочные аналоги рассмотренных вероятностных характеристик, в частности размах выборки $|x_{(N)} - x_{(1)}|$, интерквартильную широту выборки $|x_q - x_p|$.

В связи с многообразием характеристик сходного назначения при решении теоретических и практических задач возникает проблема выбора из них одной или двух. Какую выбрать? В общей постановке ответ прост: ту, которая, во-первых, существует для рассматриваемых объекта и условий экспериментирования; во-вторых, позволяет решить вашу задачу наилучшим образом, т. е. с наилучшим качеством либо при наименьших требуемых затратах, либо наиболее просто, например с наименьшими погрешностями оценки, определяемой по той же выборке. Окончательный ответ можно получить только при решении конкретной задачи, в конкретных условиях экспериментиро-

вания. При отсутствии априорных сведений, возможно, при этом придется действовать перебором или использовать **вектор-характеристику** (см. [5]) – систему из минимального набора родственных по назначению, как можно более простых и близких по сложности характеристик, отражающих в совокупности все многообразие свойств и особенностей случайного элемента как модели объекта с учетом условий экспериментирования. Примеры теоретических и выборочных характеристик представлены в табл. 4.4. Но прежде чем переходить к таблице, введем важные понятия.

Первым по важности является понятие **вариационный ряд**. Под ним чаще всего понимают упорядоченную, ранжированную в порядке возрастания (или убывания) значений выборку, обозначаемую через $(\chi) = (x_{(1)}, x_{(2)}, \dots, x_{(N)})$, где $x_{(1)}$ – наименьшее из значений членов выборки x_1, \dots, x_N ; $x_{(2)}$ – следующее по величине значение, ...; $x_{(N)}$ – максимальное значение членов исходной выборки χ . Для векторных $\mathbf{x}_i = (x_{1,i}, \dots, x_{n,i})$ неравенство $\mathbf{x}_i \leq \mathbf{x}_k$ понимается как пересечение неравенств $x_{j,i} \leq x_{j,k}$, $j = \overline{1, n}$, $i, k = \overline{1, N}$. Вариационный ряд по такой выборке можно получить, например, если составить вариационные ряды каждой из величины X_j , т. е. $x_{j,(1)}, \dots, x_{j,(N)}$, $j = \overline{1, n}$, а затем в качестве $\mathbf{x}_{(1)}$ взять $(x_{1,(1)}, \dots, x_{n,(1)})$, в качестве $\mathbf{x}_{(2)}$ – $(x_{1,(2)}, \dots, x_{n,(2)})$ и т. д. Реже в понятие **вариационный ряд** вкладывают упорядоченный сгруппированный ряд вариант $x_{(1)}, \dots, x_{(L)}$ и соответствующие им частоты распределения. Особенно удобно это понятие применять для так называемых **интервальных вариационных рядов**, использующих группирование данных. Как уже указывалось, под **группированием** понимают замену значений x_1, \dots, x_N , попадающих в заданный l -й интервал (интервал группирования) $[x_{[l]} - \Delta x_l / 2, x_{[l]} + \Delta x_l / 2)$, $l = \overline{1, L}$, где Δx_l – ширина l -го интервала, на середину этого интервала $x_{[l]}$. В этом случае вместо вариационного ряда $x_{(1)}, \dots, x_{(N)}$ получают ряд из упорядоченного набора границ интервалов или тройки: $x_{[l]}$, Δx_l и частостей P_l^* попадания в них значений x_1, \dots, x_N .

Характеристики-аналоги случайной величины (СВ) X и ее выборки χ

№ п/п	Вероятностные характеристики	Выборочные характеристики
1. Характеристики уровня (сдвига, положения диапазона значений)		
1.1	<p>Математическое ожидание (математическое (теоретическое) среднее, среднее вероятностное) (!)¹</p> $m_X = \mathbf{M}\{X\} = \begin{cases} \sum_{i=-\infty}^{\infty} x_i \mathbf{P}(X = x_i); \\ X - \text{дискретная СВ}; \\ \int_{-\infty}^{\infty} xW(x)dx; \\ X - \text{абсолютно} \\ \text{непрерывная СВ} \end{cases}$	<p>Выборочное среднее (среднее арифметическое)</p> $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i; \quad [\bar{x}] = \frac{1}{N} \sum_{l=1}^L N_l x_{[l]}$
1.2	<p>Квантиль порядка p ($0 \leq p \leq 1$) (!)</p> $x_p : F_X(x_p) = p, \quad \mathbf{P}\{X \leq x_p\} = p$	<p>Выборочная квантиль</p> $\chi_p = \hat{x}_p = \begin{cases} x_{(pN)}, \\ \text{если } pN - \text{целое число}; \\ [x_{(1+E\{pN\})} + x_{(E\{pN\})}] / 2, \\ \text{если} \end{cases}$ <p>pN – нецелое число; $E\{z\}$ – целая часть числа z</p>
1.3	Нижняя квартиль $x_{1/4}$	Нижняя квартиль выборки $\hat{x}_{1/4} = \chi_{1/4}$
1.4	Верхняя квартиль $x_{3/4}$	Верхняя выборочная квартиль $\hat{x}_{3/4} = \chi_{3/4}$
1.5	Медиана	Выборочная медиана $\chi_{0,5} = \hat{x}_{0,5} = x_{((N+1)/2)}$, если N – нечетное число; $\chi_{0,5} = \hat{x}_{0,5} = [x_{(N/2)} + x_{(1+N/2)}] / 2$, если N – четное число
1.6	<p>Мода $\text{Mo}\{X\} = \begin{cases} \arg \max_x W_X(x), \\ x_i : \mathbf{P}\{X = x_i\} = \max_p \end{cases}$</p>	$\hat{\text{Mo}}\{X\} = \chi_{\uparrow}$ – значение x_i , $x_{(i)}$, соответствующее максимальной частоте (частости) $\max_l \hat{P}_l = \max_l N_l / N, \quad l = \overline{1, L}$

Продолжение табл. 4.4

№ п/п	Вероятностные характеристики	Выборочные характеристики
2. Характеристики рассеяния (степени разброса, флуктуаций) значений		
2.1	Дисперсия (!) $D_X = \mathbf{D}\{X\} = \mathbf{M}\left\{(X - m_X)^2\right\} =$ $= \mathbf{M}\{X^2\} - (\mathbf{M}\{X\})^2$	Выборочная дисперсия $S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$
2.2	Среднеквадратическое (стандартное) отклонение $\sigma = \sigma_X = +\sqrt{D_X}$	Выборочное СКО (стандартное отклонение выборки) $S = +\sqrt{S^2}$
2.3	Интерквартильная ширина (размах квантилей) $E_{p,q} = x_q - x_p, q > p$	Выборочная интерквартильная ширина $\varepsilon_{p,q} = \chi_q - \chi_p = \hat{x}_q - \hat{x}_p, q > p$
2.4	Интерквартильная ширина (размах квантилей) $E_{1/4,3/4} = x_{3/4} - x_{1/4}$	Размах квартилей выборки $\varepsilon_{1/4,3/4} = \hat{x}_{3/4} - \hat{x}_{1/4} = \chi_{3/4} - \chi_{1/4}$
2.5	Интердецильная ширина $E_{1/10,9/10} = x_{0,9} - x_{0,1}$	Размах децилей выборки $\varepsilon_{1/10,9/10} = \hat{x}_{0,9} - \hat{x}_{0,1} = \chi_{0,9} - \chi_{0,1}$
2.6	Квантильное Q_{1-2p} (квартильное $Q_{0,5}$, децильное $Q_{0,8}$) отклонение $Q_{1-2p} = [x_{1-p} - x_p] / 2, p < 0,5;$ $(Q_{0,5} = [x_{3/4} - x_{1/4}] / 2, Q_{0,8} =$ $= [x_{0,9} - x_{0,1}] / 2)$	Выборочное квантильное q_{1-2p} (квартильное $q_{0,5}$, децильное $q_{0,8}$) отклонение $q_{1-2p} = [\hat{x}_{1-p} - \hat{x}_p] / 2, p < 0,5$
2.7	Срединное отклонение (отклонение от среднего) $E_1 : \mathbf{P}\{ X - m_X \leq E_1\} =$ $= F_X(m_X + E_1) - F_X(m_X - E_1) \leq 1/2$	Выборочное срединное отклонение $\varepsilon_1 : \hat{F}(\bar{x} + \varepsilon_1) - \hat{F}(\bar{x} - \varepsilon_1) \leq 1/2,$ $\hat{F}(x) = \sum_{i=1}^N \mathbf{I}\{x_i \leq x\} / N$ – эмпирическая функция распределения (накопленная частота); $\mathbf{I}\{A\} = 1, \mathbf{I}\{\bar{A}\} = 0$ – индикатор события A ; N_X – число членов вариационного ряда, находящихся левее x , т. е. значения которых $x_i \leq x$
2.8	Срединное отклонение от медианы $E_2 : F_X(x_{0,5} + E_2) - F_X(x_{0,5} - E_2) \leq 1/2$	Выборочное срединное отклонение от медианы $\varepsilon_2 : \hat{F}_2(\hat{x}_{0,5} + \varepsilon_2) - \hat{F}_2(\hat{x}_{0,5} - \varepsilon_2) \leq 1/2$

Продолжение табл. 4.4

№ п/п	Вероятностные характеристики	Выборочные характеристики
2.9	Среднее (абсолютное) отклонение (от среднего) $\delta_1 = \mathbf{M}\{ X - m_X \}$	Выборочное среднее отклонение $\hat{\delta}_1 = \overline{ x } = \sum_{i=1}^N x_i - \bar{x} / N$
2.10	Среднее (абсолютное) отклонение от медианы $\delta_2 = \mathbf{M}\{ X - x_{0,5} \}$	Выборочное среднее отклонение от медианы $\hat{\delta}_2 = \sum_{i=1}^N x_i - \hat{x}_{0,5} / N$
2.11	Средняя разность Джини $\delta_3 = \mathbf{M}\{ X_1 - X_2 \}$, X_1, X_2 – независимые величины с распределением величины X ; $x_{[l]}$ – середина l -го интервала группирования	$\delta_3 = \sum_{i=1}^N \sum_{j=1}^N x_i - x_j / N^2 =$ $= 2 \sum_{i=1}^{N-1} i(N-1) x_{i+1} - x_i / N^2$ Для интервального ряда, группированного с шагом Δ_X , $\delta_3 = 2 \sum_{i=1}^N \hat{F}(x_{[l]}) [1 - \hat{F}(x_{[l]})]$
2.12	Размах b - a финитной величины (финитного распределения) $X \in [a, b]$; $\mathbf{P}\{X < a\} = 0$, $\mathbf{P}\{X > b\} = 0$	Размах выборки $x_{(N)} - x_{(1)} = \hat{x}_{\max} - x_{\min}$
3. Характеристики формы распределения (вида, формы, рассеяния значений)		
3.1	(Моментный) коэффициент асимметрии (!) $\gamma_1 = \mu_3 / \sigma^3 = \mu_3 / \mu_2^{3/2}$, $b_1 = \gamma_1^2$, $\mu_k = \mathbf{M}\{(X - m_X)^k\}$ – центральный момент k -го порядка (!)	Коэффициент асимметрии (моментный) выборки (асимметрия выборки) $\gamma_1^* = \mu_3^* / \sigma_3^*$, γ_1^* – любая оценка γ_1 ; $\tilde{\gamma}_1 = A_S = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{\left(\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \right)^3}$
3.2	Квантильный коэффициент асимметрии (!) $\alpha_1 = \frac{x_p + x_{1-p} - 2x_{0,5}}{2(x_p - x_{1-p})}$, $p \neq 1/2$, например $p = 3/4$	Выборочная квантильная асимметрия $\hat{\alpha}_1 = \frac{\hat{x}_p + \hat{x}_{1-p} - 2\hat{x}_{0,5}}{2(\hat{x}_p - \hat{x}_{1-p})} =$ $= \frac{\chi_p + \chi_{1-p} - 2\chi_{0,5}}{2(\chi_p - \chi_{1-p})}$

Продолжение табл. 4.4

№ п/п	Вероятностные характеристики	Выборочные характеристики
3.3	Коэффициент асимметрии Пирсона $\alpha_2 = S_k = \frac{m_X - \text{Mo}\{X\}}{\sigma_X}$	Пирсоновская асимметрия $\hat{\alpha}_2 = \frac{\bar{x} - \hat{\text{Mo}}\{X\}}{S} = \frac{\bar{x} - x_{\uparrow}}{S}$
3.4	Среднемедианная асимметрия $\alpha_3 = \frac{m_X - x_{0,5}}{\sigma_X}, \alpha_3 \leq 1$	Среднемедианная асимметрия выборки $\alpha_3 = \frac{\bar{x} - \hat{x}_{0,5}}{S} = \frac{\bar{x} - \chi_{0,5}}{S}$
3.5	(Моментные) коэффициенты эксцесса $\gamma_2 = (\mu_4/\mu_2^2) - 3; b_2 = \mu_4/\mu_2^2$	Выборочный эксцесс $b_2^* = \mu_4^*/(\mu_2^*)^2; \gamma_2^* = b_2^* - 3;$ $\tilde{b}_2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^2}; \tilde{\gamma}_2 = E\tilde{b}_2 - 3$
3.6	Квантильный коэффициент эксцесса $\beta_1 = \frac{x_p - x_{1-p}}{x_q - x_{1-q}}, q > p, q \approx 1,$ напрмер $q = 0,9, p = 0,75$	Квантильный эксцесс выборки $\hat{\beta}_1 = \frac{\hat{x}_p - \hat{x}_{1-p}}{\hat{x}_q - \hat{x}_{1-q}} = \frac{\chi_p - \chi_{1-p}}{\chi_q - \chi_{1-q}}$
4. Другие «средние»		
4.1	Степенное ожидание порядка k $\psi_k = \left[\mathbf{M}\{X^k\} \right]^{1/k},$ $\psi_1 = m_1 = m_X, k = 1, 2, \dots$	Среднее порядка k $(\bar{x}^k)^{1/k} = \hat{\psi}_k = \left(\sum_{i=1}^N x_i^k / N \right)^{1/k}, k = 1, 2, \dots$
4.2	Гармоническое ожидание $\psi_{-1} = 1 / (\mathbf{M}\{1/X\})$	Гармоническое среднее $\hat{\psi}_1 = \frac{N}{\sum_{i=1}^N (1/x_i)} = \frac{N}{\sum_{i=1}^N x_i^{-1}}$
4.3	Геометрическое ожидание $\psi_0 = \lim_{k \rightarrow \infty} \left[\mathbf{M}\{X^k\} \right]^{1/k}, X > 0$	Геометрическое среднее $\beta = \left(\prod_{i=1}^N x_i \right)^{1/N} = \exp \left\{ \sum_{i=1}^N \ln x_i / N \right\}, x_i > 0$
4.4	Квадратичное ожидание $\psi_2 = \sqrt{m_2} = \left[\mathbf{M}\{X^2\} \right]^{1/2}$	Квадратичное среднее $\psi_2 = \sqrt{\bar{x^2}} = \sqrt{\sum_{i=1}^N x_i^2 / N}$

№ п/п	Вероятностные характеристики	Выборочные характеристики
4.5	Центральный момент k -го порядка $\mu_k, k \geq 2$ $\mu_k = \mathbf{M}\{(X - m_X)^k\}$	а) m_X – известно: $\hat{\mu}_k = \sum_{i=1}^N (x_i - m_X)^k / N$; б) m_X – неизвестно: $\tilde{\mu}_k = \sum_{i=1}^N (x_i - \bar{x})^k / N$
5. Вероятностные характеристики (характера рассеяния, концентрации значений)		
5.1	Вероятность события A $\mathbf{P}\{A\} = \mathbf{M}\{\mathbf{I}(A)\}$, $\mathbf{I}(A)$ – индикатор события A , т. е. $\mathbf{I}(A) = 1, \mathbf{I}(\bar{A}) = 0$, \bar{A} – событие, противоположное A	Частота (частость) события A $\hat{\mathbf{P}}\{A\} = \hat{p} = \hat{\mathbf{P}}(A) = \sum_{i=1}^N \mathbf{I}_i(A) / N = N_A / N$, N_A – число опытов, в которых появилось событие A
5.2	Функция распределения вероятностей $F(x) = \mathbf{P}\{X \leq x\} = \mathbf{M}\{\mathbf{I}(X \leq x)\}$	Эмпирическая функция распределения $\hat{F}(x) = N_{(x_i \leq x)} / N = \sum_{i=1}^N \mathbf{I}(x_i \leq x) / N$, $N_{(x_i \leq x)}$ – количество элементов выборки, значения которых не превышают x , т. е. номер ближайшего к x слева члена вариационного ряда. Эмпирическая функция распределения группированного ряда, гистограммная эмпирическая функция $\tilde{F}(x) = \sum_{i=1}^N \mathbf{I}(x_i \leq x_{[l]} + \Delta x_l / 2) / N$ при $x \in [x_{[l]} - \Delta x_l / 2, x_{[l]} + \Delta x_l / 2]$
5.3	Плотность распределения вероятностей $W(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\mathbf{P}\left\{x_{[l]} - \frac{\Delta x_l}{2} < X \leq x_{[l]} + \frac{\Delta x_l}{2}\right\}}{\Delta x_l}$	Гистограмма $\hat{W}(x) = N_l / (N \Delta x_l) = \hat{\mathbf{P}}_l / \Delta x_l$ для $x \in [x_{[l]} - \Delta x_l / 2, x_{[l]} + \Delta x_l / 2]$ Полиграмма k -го порядка $\hat{W}^*(x) = \frac{k}{N(x_{(rk+k)} - x_{(rk)})}$, $x \in (x_{(rk+k)} - x_{(rk)})$, $r = \overline{1, m}$, $m = E\{N/k\}$, $E\{z\}$ – целая часть числа z

¹ Здесь и далее (!) означает, что условия существования выполняются.

Рассмотрим еще несколько понятий, связанных со случайными величинами X и выборками $\chi = (x_1, \dots, x_N)$.

Минимальным значением выборки x_{\min} называется наименьшее значение выборки χ , **максимальным** x_{\max} – наибольшее, т. е. $x_{\min} = x_{(1)}$, $x_{\max} = x_{(N)}$. Разность $x_{\max} - x_{\min} = x_{(N)} - x_{(1)}$ называется **размахом выборки**.

Заметим, что оценка $\hat{Q}_1(\vartheta)$ считается **эффeктивной** оценки $\hat{Q}_2(\vartheta)$ той же характеристики $Q(\vartheta)$, если $\Delta_{\hat{Q}_1}(\vartheta) \leq \Delta_{\hat{Q}_2}(\vartheta)$, что для несмещенных оценок равносильно неравенству $\mathbf{D}\{\hat{Q}_1(\vartheta)\} \leq \mathbf{D}\{\hat{Q}_2(\vartheta)\}$ или $\sigma_{\hat{Q}_1}(\vartheta) \leq \sigma_{\hat{Q}_2}(\vartheta)$.

Вспомним, что средний квадрат $\Delta_Q^2(\vartheta)$ погрешности оценки $\hat{Q}(\vartheta)$ есть

$$\Delta_Q^2(\vartheta) = \varepsilon_Q^2(\vartheta) + \sigma_Q^2(\vartheta),$$

где $\varepsilon_Q(\vartheta)$ – смещение оценки, т. е.

$$\varepsilon_Q(\vartheta) = \mathbf{M}\{\hat{Q}(\vartheta) - Q(\vartheta)\},$$

$$\sigma_Q(\vartheta) = +\sqrt{\mathbf{D}\{\hat{Q}(\vartheta)\}}.$$

4.3.4. Анализ и идентификация распределения значений

Вторым по последовательности и по повышению информативности об исследуемом объекте и эксперименте с ним является анализ одномерных, реже двумерных, иногда многомерных распределений вероятностей случайных элементов (величин, векторов, функций – процессов, последовательностей) и/или их аналогов – статистических (эмпирических) законов распределения выборок.

Цель анализа распределений – получение графических образов и/или аналитического описания распределения вероятностей значений случайных элементов как моделей параметров (физических величин)

реальных объектов, исходя из предъявляемых к такому описанию требований:

- а) отражение «механизма» (физики) порождения этих значений;
- б) адекватность описания в рамках имеющихся надных;
- в) пригодность для решения прикладных задач, например, прогнозирования вероятностей значений за пределами интервала получения данных, определения процентной доли значений величины в какой-то области через вероятность попадания в эту область (см. рис. 4.3);
- г) простота описания;
- д) простота последующей имитации выборочных значений;
- е) другие требования.

В качестве законов распределения как характеристик случайных элементов или генеральных совокупностей экспериментальных данных используются следующие. **Функция** (интегральный закон) **распределения вероятностей** (ФРВ)

$$F(x) = \mathbf{P}\{X \leq x\}, \quad F(x_1, x_2, \dots, x_n) = \mathbf{P}\left\{\bigcap_{i=1}^n (X_i \leq x_i)\right\},$$

$$F(x) = \mathbf{P}\{X(t) \leq x\}, \quad F(x_1, x_2, \dots, x_n; t_1, \dots, t_n) = \mathbf{P}\left\{\bigcap_{i=1}^n (X_i(t_i) \leq x_i)\right\}.$$

Плотность распределения (дифференциальный закон) **вероятностей** (ПРВ), которая для абсолютно непрерывных распределений определяется как $W(x) = dF(x)/dx$, $W(x_1, \dots, x_n) = \partial^n F(x_1, \dots, x_n) / \partial x_1 \dots \partial x_n = W_{X_1}(x_1)W_{X_2}(x_2 | x_1)W_{X_3}(x_3 | x_1, x_2) \dots W_{X_n}(x_n | x_1, \dots, x_{n-1})$, где $W_{X_i}(x_i | A)$ – (условная) плотность распределения вероятностей при условии (обозначается символом $|$) A : условие $A = x_i$ – есть $X_i = x_i$; условие $A = (x_1, x_2)$ есть $(X_1 = x_1) \cap (X_2 = x_2), \dots$; условие $A = (x_1, \dots, x_{n-1})$ есть

$\bigcap_{i=1}^{n-1} (X_i = x_i)$. **Ряд распределения** (РР) $P(x)$ есть набор пар (x_j, p_j) значений x_j дискретной случайной величины X и вероятностей $p_j = \mathbf{P}\{X = x_j\}$, с которыми величина X принимает это значение x_j . Как известно, ФВР существует для всех случайных элементов, ПРВ –

для абсолютно непрерывных и дискретных¹ случайных элементов, а ряд распределения – только для дискретных случайных величин, векторов и функций (см. рис. 4.3).

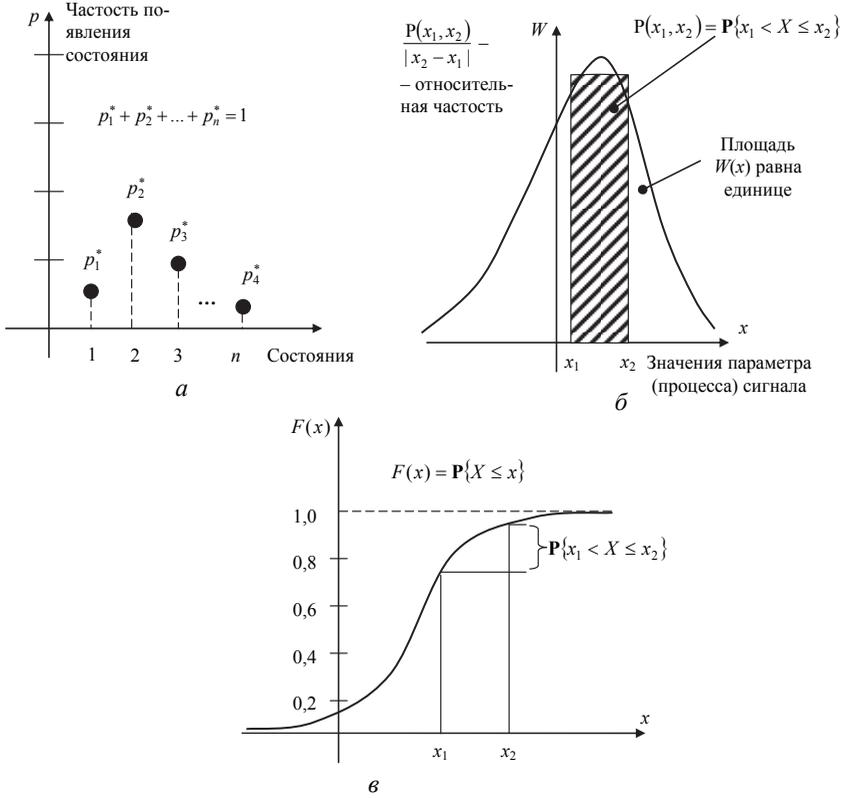


Рис. 4.3. Графическая иллюстрация законов распределения:

частота $P(A) = P^*(A)$ появления события A : отношение числа опытов, когда A появилось, к общему числу опытов (в % или долях от 0 до 1); $a - A$ – событие, состояние; $\bar{b} - A$ – попадание значений случайной величины X в область (x_1, x_2) ; $\bar{b} - A$ – попадание значений X в область $(-\infty, x)$ вверху или (x_1, x_2) в центре; p^* – конкретное значение частоты или вероятности p для РР; $P(A)$ – вероятность события A ; $P\{\cdot\}$ – оператор вероятности как меры; W, F – плотность и функция распределения случайной величины X

¹ В этом случае используются обобщенные дельта-функции $\delta(x), \delta(x_1, \dots, x_n)$.

Выборочными аналогами вероятности события A являются частота $\mathcal{P}(A)$ появления A в экспериментах, т. е. отношение $\mathcal{P}(A) = p^*(A) = N_A/N$ числа N_A появлений события A в серии из N опытов к количеству N проведенных опытов эксперимента. Выборочными аналогами ФРВ и ПРВ являются их гистограммные или полиграммные статистики. Гистограммные эмпирические ФРВ $\mathcal{F}_1(x)$ и ПРВ (гистограммы) $\mathcal{W}_1(x)$ получаются разбиением диапазона значений $[x_{(1)}, x_{(N)}]$ на n состыкованных участков $(x_{[l]}, x_{[l+1]})$, где $x_{[l]}$ – левая граница участка; $x_{[l+1]} = x_{[l]} + \Delta x_l$, Δx_l – ширина l -го полуинтервала группирования, $x_{[1]} = x_{(1)}$, $x_{[n+1]} = x_{(n)}$. Значения эмпирических ПРВ равны $\mathcal{W}_1(x) = N_{[l]}/(N\Delta x_l)$ для $x \in (x_{[l]}, x_{[l+1]})$, а для эмпирических ФРВ – $\mathcal{F}_1(x) = (N_{[1]} + \dots + N_{[l]})/N$ для $x \in (x_{[l]}, x_{[l+1]})$. Вторая, полиграммная, эмпирическая ФВР определяется как $\mathcal{F}_2(x) = l/n$ для $x \in (x_{[l]}, x_{[l+1]})$.

Для автоматизации упорядочения и априорного или апостериорного выбора аналитической модели, описывающей (аппроксимирующей, идентифицирующей) теоретическое или эмпирическое распределение, можно рекомендовать метод моделетеки [5, 20].

Знание законов распределения позволяет определять процентную долю значений случайных элементов или выборок, попадающих в соответствующую область допустимых значений, а также наличие выбросов, однородность выборок (по наличию и количеству эмпирических мод и антимод выборочного распределения) и решать другие подобные задачи.

4.3.5. Анализ взаимосвязей случайных величин и векторов

Следующая важная задача обработки и анализа экспериментальных данных – это определение взаимосвязи исследуемых объектов (предметов, показателей, параметров, явлений, процессов) путем исследования взаимосвязи случайных величин, векторов и функций (процессов, последовательностей), с помощью которых описываются (представляются) эти объекты. Обратим еще раз внимание на то, что статистиче-

ские методы позволяют определить наличие связи между объектами и их взаимосвязи, оценить направление (тенденцию), вид (функциональная, статистическая), характер (линейная, нелинейная; однозначная, неоднозначная) и степень тесноты (степень разброса значений одних элементов при фиксированных значениях других) или, иначе, степень неопределенности значений одной величины при конкретных значениях другой. Однако эти методы не позволяют определить причины, выявить, что первично в этой связи, первопричина, а что вторично, следствие, кто ведущий, а кто ведомый.

Рассмотрим вначале вопросы количественной оценки связи и взаимосвязи на примере случайных качественных факторов, количественных величин и векторов.

Дисперсный (дисперсионный¹, факторный) анализ

Дисперсный анализ – это статистический метод анализа экспериментальных надных, зависящих от действия качественных факторов. Он используется для выявления совместного влияния разных факторов, не поддающихся количественному измерению, на исследуемый показатель (физическую или численную величину).

Цель дисперсного анализа – установление факта наличия или отсутствия влияния на исследуемую случайную величину (вектор) качественных факторов, сопутствующих реальному проявлению этой величины (вектора).

Суть метода – выявление влияния наличия или отсутствия учитываемых и неучитываемых факторов на вариацию исследуемого показателя. Для этого общая дисперсия D_X как характеристика вариации рассеяния значений показателя X расчленяется на части, соответствующие раздельному и смешанному воздействию различных качественных факторов, и остальную дисперсию, обусловленную влиянием всех неучтенных факторов. Рассмотрим для простоты однофакторный дисперсный анализ.

В основе *однофакторного анализа* лежит представление исследуемого показателя X в виде случайных величин

¹ Термин *дисперсионный* используется чаще, но термин *дисперсный* лучше отражает суть метода. К тому же под дисперсионным анализом понимают анализ через корреляционные отношения (см. далее).

$$X_{ij} = a_j + \Xi_{ij}, \quad i = \overline{1, I_j}, \quad j = \overline{1, J}, \quad \sum_{j=1}^J I_j = N, \quad (4.5)$$

или их выборочных значений

$$x_{ij} = a_j + \xi_{ij}, \quad i = \overline{1, N_j}, \quad j = \overline{1, J}, \quad \sum_{j=1}^J N_j = N, \quad (4.6)$$

где X_{ij} – случайные величины, представляющие результирующий признак x ; a_j – среднее (математическое ожидание) признака x при j -м значении качественного фактора (например, $a_j, j = \overline{1, J}$ – среднее значение мутности воды, измеренной j -м измерителем одного и того же типа); ξ_{ij} – реализации Ξ_{ij} – случайные (как правило, считаются нормально распределенными) отклонения результирующего показателя x от средних $\mathbf{M}\{\Xi_{ij}\} = 0, \mathbf{D}\{\Xi_{ij}\} = \sigma^2$; N_j – число экспериментальных значений (наблюдений) при j -м значении качественного фактора (полученных с использованием j -го измерителя); N – общее число экспериментальных данных.

Представим a_j в виде

$$a_j = a + \varepsilon_j, \quad \sum_{j=1}^J \varepsilon_j = 0, \quad (4.7)$$

где $a = (a_1 + \dots + a_J)/J$ – общее (генеральное) среднее, $\varepsilon_j = a_j - a, j = \overline{1, J}$ – главные эффекты фактора.

Найдем следующие эмпирические (выборочные) характеристики:

$$\bar{x} = \sum_{j=1}^J \sum_{i=1}^{N_j} x_{ij}; \quad \bar{x}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_{ij}; \quad (4.8)$$

$$S^2 = \sum_{j=1}^J \sum_{i=1}^{N_j} (x_{ij} - \bar{x})^2 = S_R^2 + S_A^2; \quad (4.9)$$

$$S_R^2 = \sum_{j=1}^J \sum_{i=1}^{N_j} (x_{ij} - \bar{x}_j)^2; \quad (4.10)$$

$$S_A^2 = \sum_{j=1}^J N_j (\bar{x}_j - \bar{x})^2, \quad (4.11)$$

где \bar{x} , S^2 – среднее и средний квадрат отклонений значений признака по всей выборке объема N ; \bar{x}_j и S_R^2 – среднее и сумма квадратов отклонений внутри j -й, $j = \overline{1, J}$, группы; S_A^2 – сумма квадратов отклонений между группами.

Примем в качестве рабочей гипотезу H_0 , что качественный фактор не влияет на x , т. е. что все средние a_j одинаковы. Тогда общее выборочное среднее \bar{x} не должно статистически значимо отличаться от \bar{x}_j . В случае, если распределения остатков подчиняются нормальному (Гаусса, см. [20]) распределению, гипотеза H_0 может быть проверена с использованием статистики Фишера:

$$F = \frac{S_A^2 / (J - 1)}{S_R^2 / (N - 1)}, \quad (4.12)$$

имеющей распределение Фишера с $J - 1$ и $N - 1$ степенями свободы (см. [20]) и принята, если $F \leq F_\alpha(J - 1, N - 1)$, где α – ошибка 1-го рода – вероятность отвергнуть гипотезу H_0 , когда она верна, т. е.

$$\mathbf{P} \left\{ F < F_\alpha(J - 1, N - 1) \middle| {}^1 H_0 \right\} = \alpha, \quad (4.13)$$

где $F_\alpha(\cdot)$ – взятое из таблицы распределения Фишера граничное значение критерия Фишера, а вертикальная черта означает условие, что H_0 верна.

¹ Здесь и далее вертикальная черта означает: «при условии, что стоящее за ним выражение, утверждение, верно, истинно, имеет место быть» (см. прил. 1, П1.4).

Если гипотеза H_0 отвергается, то принимается решение о том, что изучаемый качественный фактор влияет на результирующий признак x . При этом \bar{x}_j принимается за оценку \hat{a}_j , а $(\hat{\sigma})^2$ для x находится как

$$(\hat{\sigma})^2 = S_R^2 / (N - J). \quad (4.14)$$

Регрессионный и скедастический анализ

Регрессионный анализ – совокупность статистических методов построения функциональной (функциональной, детерминированной, числовой) зависимости одного результирующего показателя y (случайной величины Y) от значений других существенных (x_1, \dots, x_k) и несущественных (ξ_1, \dots, ξ_n) факторов.

Аналитически эта зависимость представляется вероятностной моделью

$$Y = f(X_1, \dots, X_k; \Xi_1, \dots, \Xi_n) + E, \quad (4.15)$$

где Y – случайная величина, описывающая результирующий показатель y ; $X_1, \dots, X_k; \Xi_1, \dots, \Xi_n$ – случайные величины, описывающие значения случайных факторов $x_1, \dots, x_k; \xi_1, \dots, \xi_n$; $f(\cdot)$ – неслучайная (детерминированная) функция; E – случайная величина, описывающая погрешности измерения y . Обычно в качестве функции $f(\cdot)$ выбирается **функция регрессии** $m_Y(x_1, \dots, x_k; \xi_1, \dots, \xi_n)$ – условное математическое ожидание Y при конкретных значениях $X_i = x_i, i = \overline{1, k}, \Xi_j = \xi_j, j = \overline{1, n}$ (отсюда и название анализа), являющаяся наилучшей аппроксимацией по среднеквадратическому критерию (СК-аппроксимации), т. е. по минимуму СКО, функциональной зависимости значений Y от конкретных значений $X_1, \dots, X_k; \Xi_1, \dots, \Xi_n$, усредненных вдоль всех их возможных значений с учетом вероятностей их появления (!):

$$m_Y(x_1, \dots, x_k; \xi_1, \dots, \xi_n) = \mathbf{M} \left\{ Y \left| \bigcap_{i=1}^k (X_i = x_i), \bigcap_{j=1}^n (\Xi_j = \xi_j) \right. \right\}. \quad (4.16)$$

В этом случае в качестве эмпирического аналога зависимости (4.15) принимается

$$\hat{y} = m_Y(x_1, \dots, x_k; \xi_1, \dots, \xi_n; \hat{\theta}_1, \dots, \hat{\theta}_s), \quad (4.17)$$

где $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_s)$ – оценки (выборочные значения) параметров функции $m_Y(\cdot)$, найденные по экспериментальным Данным. Поскольку ξ_1, \dots, ξ_n – значения несущественных факторов, их либо опускают, либо на этапе оценивания параметров θ приравнивают к существенным и отбрасывают лишь после статистической проверки значимости их влияния на \hat{y} (см. ниже). В дальнейшем будем их просто опускать.

Таким образом, *цель регрессионного анализа* – получение и исследование аналитического описания функциональной зависимости среднего значения Y (или Y_1, \dots, Y_m) от конкретных значений влияющих на него признаков (факторов, регрессоров), удовлетворяющего совокупности определенных требований: отражение «механизма» зависимости (случайности), простота и т. п. (см. анализ распределения вероятностей и требования к модели).

Обратим внимание, что в (4.17) мы сознательно в качестве y выбираем не его возможное случайное значение, которое может быть в опытах, а значение, определяемое согласно (4.16), т. е. среднее из возможных случайных. Это очень важно помнить, чтобы не впасть в схоластическую дискуссию, почему по значениям x_1, \dots, x_k можно по функции (линии) регрессии предсказать (определить) значение y (точнее его среднее значение!), но нельзя по y определить x_1, \dots, x_k (имеет место необратимость функции регрессии). Или, по-другому, почему по функции регрессии $m_Y(x)$ при заданном x получаем одно y , а для того же y по $m_X(y)$ получаем другое x (имеет место неравенство $m_Y(x) \neq m_X(y)$). Отсюда ясна бесплодность поиска какого-либо аналога, модификации функций регрессии, чтобы эту задачу решить. Еще раз заметим, что функция регрессии, например $m_Y(x)$, позволяет определять $y = m_Y(x)$ как среднее из возможных значений Y при данном x , а не какое-то конкретное y_i , которое может быть в i -м эксперименте в паре $(x, y_1), (x, y_2), \dots, (x, y_N)$.

Обычно регрессионный анализ предполагает решение четырех задач:

1) априорный или апостериорный выбор модели – функции, описывающей (или аппроксимирующей) $m_Y(x_1, \dots, x_k; \theta_1, \dots, \theta_s)$;

2) оценивание параметров $\theta_1, \dots, \theta_s$;

3) статистическая проверка гипотез о значимости влияния на Y каждого i -го фактора (величины) X_i , $i = \overline{1, k}$; адекватности модели данным, включая проверку гипотезы о свойствах остатков ε ;

4) решение прикладной задачи (идентификация зависимости, нахождение значений y за пределами полученных значений x_1, \dots, x_k , т. е. при требуемых x'_1, \dots, x'_k , в частности решение прогнозных задач, и т. п.).

При прогнозировании находят аналитическое описание $m_Y(x_1, \dots, x_k; \theta_1, \dots, \theta_s)$, затем оценивают $\theta_1, \dots, \theta_s$ по выборке, проверяют, какие переменные x_1, \dots, x_k оставить в выражении функции регрессии, после чего находят искомые (заметим еще раз – на уровне средних значений!) \hat{y} для заданных текущих значений x'_1, \dots, x'_k по соотношению

$$\hat{y} = m_Y(x'_1, x'_2, \dots, x'_k; \hat{\theta}_1, \dots, \hat{\theta}_s). \quad (4.18)$$

Разность между полученными в дальнейшем экспериментальными значениями y и предсказанными по модели (4.18) значениями \hat{y} называется *остатками* ε , т. е.

$$y - \hat{y} = \varepsilon. \quad (4.19)$$

Их можно считать реализациями E из (4.15), по их описанию дополнительно судить о качестве прогноза, а также дополнительно проверять адекватность гипотезы, положенной в основу идентификации зависимости Y от X_1, \dots, X_k .

Свойства функций регрессии на примере двух величин X и Y (!)

$$m_Y(x) = \mathbf{M}\{Y | (X = x)\} = \int_{-\infty}^{\infty} y W_Y(y | x) dy; \quad (4.20)$$

$$m_X(y) = \mathbf{M}\{X|(Y=y)\} = \int_{-\infty}^{\infty} xW_X(x|y)dx \quad (4.21)$$

представлены в приложении 2 (табл. П2.1); $W_Y(y|x)$, $W_X(x|y)$ – условные плотности распределения вероятностей. Графические изображения функций регрессии называются **кривыми регрессии**.

Скедастический анализ – это совокупность статистических методов определения рассеянности, разброса, флюктуаций, вариации одной зависимой случайной величины Y при конкретных значениях других величин (факторов) x_1, \dots, x_k .

Цель скедастического анализа – построение и исследование удовлетворяющей совокупности требований аналитической функциональной зависимости условного среднеквадратического отклонения результирующего признака Y или $\mathbf{Y} = (Y_1, \dots, Y_m)$ от конкретных значений x_1, \dots, x_n влияющих признаков (факторов, регрессоров) X_1, \dots, X_n .

Обычно он выполняется через условные дисперсии

$$\sigma_Y^2(x) = \mathbf{M}_{\text{усл}} \left\{ [Y - m_Y(x)]^2 \right\} = \int_{-\infty}^{\infty} [y - m_Y(x)]^2 W_Y(y|x) dy; \quad (4.22)$$

$$\begin{aligned} \sigma_X^2(y) &= \mathbf{M}_{\text{усл}} \left\{ [X - m_X(y)]^2 \right\} = \\ &= \int_{-\infty}^{\infty} [x - m_X(y)]^2 W_X(x|y) dx. \end{aligned} \quad (4.23)$$

Функции, описывающие зависимости $\sigma_Y^2(x)$ и $\sigma_X^2(y)$, называются **скедастическими**¹, а их графическое изображение – **скедастическими кривыми**. Свойства скедастических функций представлены в прил. 2 (табл. П2.2).

Как следует из табл. П2.1 и П2.2, из всех функций $f(x_1, \dots, x_k)$, приближающих (аппроксимирующих) зависимость Y от x_1, x_2, \dots, x_k на уровне детерминированного (числового, неслучайного, функциональ-

¹ От греч. skedassimo – рассеиваю.

ного) описания, наилучшее СК-приближение дает функция регрессии $m_Y(x_1, \dots, x_k)$ вдоль всех значений Y и X_1, \dots, X_k . Минимальное значение среднего квадрата погрешности Δ_{\min}^2 при этом равно

$$\Delta_{\min}^2 = \mathbf{M} \left\{ \sigma_Y^2(X) \right\} = \mathbf{D}\{Y\} - \mathbf{M} \left\{ [m_Y(X) - m_Y]^2 \right\}. \quad (4.24)$$

Следовательно, во-первых, выбирая любое приближение (аппроксимацию) зависимости Y от X_1, \dots, X_k функцией $\psi(x_1, \dots, x_k)$ и сравнивая получаемые при этом значения Δ_ψ^2 с Δ_{\min}^2 , можно судить о близости приближения (аппроксимации). Например, выбирая $\psi(x)$ в виде $\psi(x) = a_0 + a_1x + a_2x^2 + \dots + a_lx^l + \dots + a_nx^n$ для аппроксимации зависимости Y от X , можно найти Δ_l^2 для разных $l = 0, 1, 2, \dots, n$ и такое минимальное $l' < n$, при котором $\Delta_{l'}^2 \approx \Delta_{\min}^2$, когда дальнейшее увеличение $l > l'$ не только не целесообразно, но и с учетом статистических погрешностей может увеличить Δ^2 при апостериорной (выборочной) аппроксимации.

Во-вторых, как видно из (4.24), величина

$$\zeta_{Y|X}^2 = \mathbf{M} \left\{ \sigma_Y^2(x) \right\} / \sigma_Y^2 \quad (4.25)$$

позволяет судить о степени вариации (разбросанности, рассеянности, неопределенности) Y относительно условного среднего $m_Y(x)$. Равенство $\zeta_{Y|X}^2 = 0$ означает, что $\sigma_Y^2(x) = 0$, т. е. зависимость Y от X является однозначной функциональной, а не статистической. Наоборот, если величины независимы, то $\sigma_Y^2(x) = \sigma_Y^2$, $m_Y(x) = m_Y$, $\zeta_{Y|X}^2 = 1$, т. е. вариации (разброс, рассеяние) Y от m_Y или неопределенность Y при фиксированном $X = x$ наибольшие из возможного. Заметим, что это имеет место не только, когда Y и X независимы, но и всегда, когда $m_Y(X) = m_Y$. При этом важно знать, какой вид имеет $m_X(y)$ (рис. 4.4).

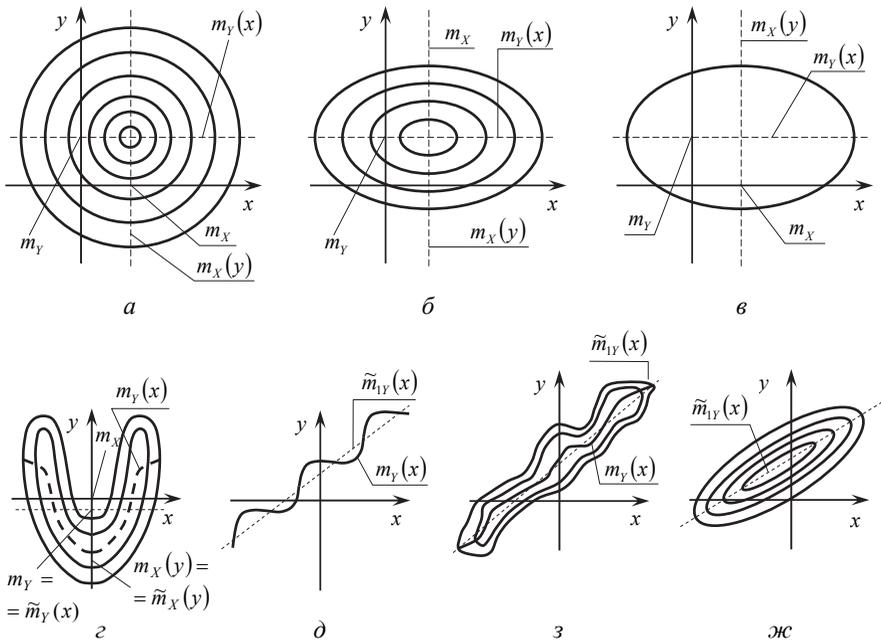


Рис. 4.4. Примеры диаграмм рассеяния, функций регрессии и прямых средней квадратической регрессии

Представим разложение дисперсий (см. свойство 7 из табл. П2.2 прил. 2) в другом виде:

$$\begin{aligned}
 \sigma_Y^2 &= \mathbf{D}\{m_Y(X)\} + \mathbf{M}\{\sigma_Y^2(X)\} = \sigma_{1Y}^2 + \sigma_{2Y}^2 = \\
 &= \mathbf{M}\left\{[m_Y(X) - m_Y]^2\right\} + \mathbf{M}\left\{[Y - m_Y(X)]^2\right\}; \quad (4.26)
 \end{aligned}$$

$$\begin{aligned}
 \sigma_X^2 &= \mathbf{D}\{m_X(Y)\} + \mathbf{M}\{\sigma_X^2(Y)\} = \sigma_{1X}^2 + \sigma_{2X}^2 = \\
 &= \mathbf{M}\left\{[m_X(Y) - m_X]^2\right\} + \mathbf{M}\left\{[X - m_X(Y)]^2\right\}. \quad (4.27)
 \end{aligned}$$

Отсюда видно, что первое слагаемое σ_1^2 отражает отличие функции регрессии от константы – математического ожидания рассматриваемой величины, т. е. вариации, рассеяние условных средних от безусловных одномерных средних, а второе слагаемое σ_2^2 – рассеяние (вариации) величины относительно своей функции регрессии (своего условного среднего, условного математического ожидания).

Корреляционный анализ

Корреляционный анализ (КА) – совокупность методов оценивания корреляционных характеристик связи и проверки гипотез об их значимости по выборочным данным.

Цель корреляционного анализа – получение количественных (численных) значений показателей о наличии, степени нелинейности и тесноте связи рассматриваемых величин как между собой, так и с третьими факторами.

Обычно при этом решаются три задачи:

1) построение **корреляционного поля** (поля рассеяния) – графического отображения в n -мерном пространстве эмпирических значений $(x_{1,i}, x_{2,i}, \dots, x_{n,i})$, $i = \overline{1, N}$, случайных величин X_1, \dots, X_n ;

2) нахождение эмпирических значений (оценок) различных (см. далее) корреляционных характеристик парной и множественной связей величин X_1, X_2, \dots, X_n вектора $\mathbf{X} = (X_1, \dots, X_n)$;

3) статистическая проверка гипотез о значимости этих связей.

Очень часто под корреляционным анализом понимают объединение регрессионно-скедастического и собственно корреляционного анализов. Тогда построение аналитических зависимостей типа $m_Y(x_1, \dots, x_n)$, $\sigma_Y(x_1, \dots, x_n)$ считается первой задачей корреляционного анализа, а получение и анализ численных значений показателей связи типа коэффициентов корреляции, корреляционных отношений и т. п. – второй задачей корреляционного анализа. Мы разделили эти две задачи по видам анализов: регрессионный, скедастический, корреляционный.

Рассмотрим КА вначале на примере величин X и Y с математическими ожиданиями $m_X = m_{10}$ и $m_Y = m_{01}$ и среднеквадратическими отклонениями σ_X , σ_Y .

Простейшей характеристикой (показателем) взаимосвязи X и Y является **коэффициент корреляции** (КК)¹ (!)

$$\rho_{XY} = \mathbf{M}\{(X - m_X)(Y - m_Y)\} / (\sigma_X \sigma_Y), \quad (4.28)$$

выборочный аналог (ЭХ-оценка $\hat{\rho}_{XY}$) которого определяется соотношением

$$\hat{\rho}_{XY} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) / \hat{\sigma}_X \hat{\sigma}_Y, \quad (4.29)$$

где (x_i, y_i) – пара значений X и Y , полученных в i -м опыте (эксперименте); $\hat{\sigma}_X, \hat{\sigma}_Y$ – оценки СКО – выборочные характеристики S_X, S_Y (см. табл. 4.4), в частности:

$$\bar{x} = \hat{m}_X = \frac{1}{N} \sum_{i=1}^N x_i, \quad \bar{y} = \hat{m}_Y = \frac{1}{N} \sum_{i=1}^N y_i. \quad (4.30)$$

Однако КК ρ_{XY} как характеристика связи X и Y должен иметь весьма ограниченное применение. Дело в том, что его корректное использование возможно, во-первых, если обе функции регрессии прямые; во-вторых, когда значения $|\rho_{XY}|$ близки к единице. Часто рекомендуют $\rho \geq 0,6$. В противном случае можно получить ложное представление о виде, характере, степени тесноты связи величин X и Y (см. свойства в табл. П2.3, П2.4 прил. 2 и рис. 4.4 и 4.5)².

Еще один пример того, что числовые характеристики, включая прямую средней квадратической регрессии, не всегда хорошо отражают истинную ситуацию, представлен на рис. 4.6.

Он построен по данным, придуманным английским статистиком Ф.Дж. Энскомбе и получившим название «квартет Энскомбе»³.

¹ От *лат.* *correlatio* – соотношение, связь.

² При функциональной зависимости $Y = f(x)$ для каждого x рассеяние Y и «точечная» линейно-корреляционная связь зависят от модуля производной $|f'(x)|$. Поэтому при нелинейной $f(x)$ рассеяние Y для каждого x будет разным, что в итоге и приводит к выводу: функциональная связь аналогична статистической, если о ней судят по коэффициенту корреляции (см. рис. 4.4, *д, е, ж*).

³ Квартет Энскомбе. Элект. ресурс <http://habrahabr.ru/blogs/statistics/91128>.

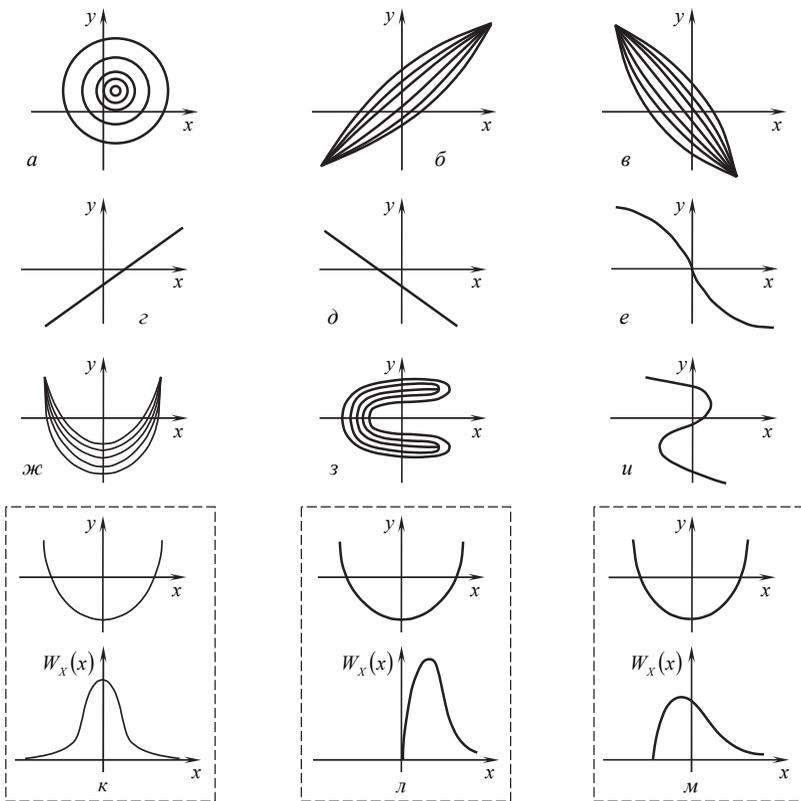


Рисунок	<i>a</i>	<i>б</i>	<i>в</i>	<i>г</i>	<i>д</i>	<i>е</i>	<i>ж</i>	<i>з</i>	<i>и</i>	<i>к</i>	<i>л</i>	<i>м</i>
ρ_{XY}	0	0,7	$-\overline{0,6}$	1	-1	?	0	0	?	0	0,7	0,3
$\eta_{Y X}$	0	0,7	0,6	1	1	1	0,5	0	?	1	1	1
$\eta_{X Y}$	0	0,7	0,6	1	1	1	0	0,7	1	0	1	0,6
χ_{XY}	0	0,7	$-\overline{0,6}$	1	-1	-1	0	0	?	0	1	0,8

Рис. 4.5. Примеры статистических (*a*, *б*, *в*, *ж*, *з*) и функциональных (*г*, *д*, *е*, *и*, *к*, *л*, *м*) зависимостей и соответствующих им значений характеристик связи случайных величин X и Y :

ρ_{XY} и χ_{XY} – коэффициенты корреляции и конкорреляции; $\eta_{Y|X}$, $\eta_{X|Y}$ – корреляционные отношения; ? – значение зависит от закона распределения $W_X(x)$

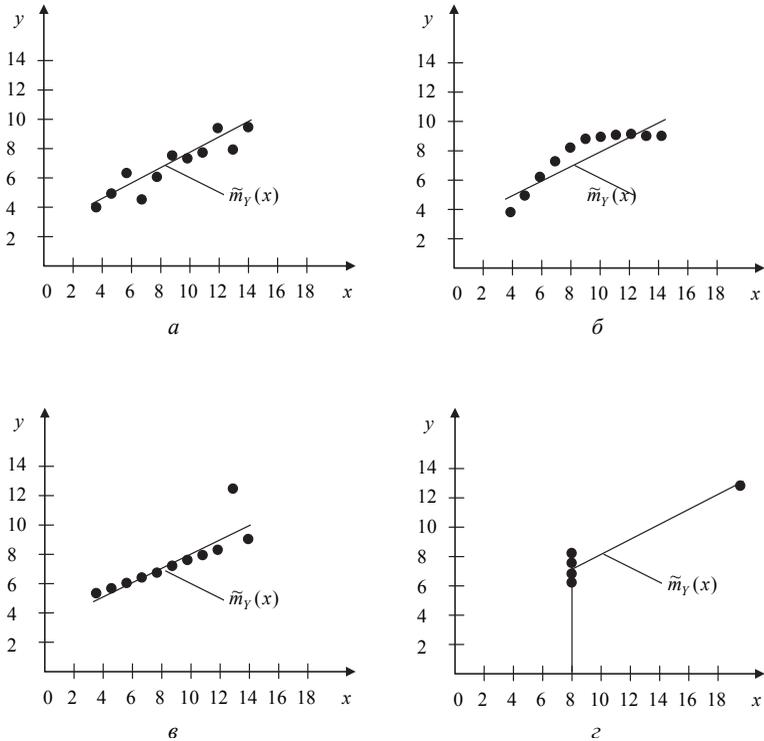


Рис. 4.6. Диаграммы рассеяния случайных величин X, Y из «квартета Энскомбе»:

для всех рисунков среднее $m_X = 9,0$; $m_Y = 7,5$; $D_X = 10,0$; $D_Y = 3,75$; коэффициент корреляции $\rho_{XY} = 0,82$; прямая среднеквадратической регрессии $\tilde{m}_Y(x) = 3 + 0,5x$

В прил. 2 (табл. П2.3) $\text{sign}(x)$ – сигнум-функция, т. е.

$$\text{sign}(x) = \begin{cases} -1, & x < 0; \\ 0, & x = 0; \\ +1, & x > 0. \end{cases} \quad (4.31)$$

В связи с этим было введено множество других характеристик связи (см., например, [20]), а также вектор-характеристика связи [5].

Рассмотрим из них только некоторые.

Корреляционные отношения (КО) (индекс корреляции):

$$\eta_{Y/X} = \frac{1}{\sigma_Y} \sqrt{\mathbf{M}\{[m_Y(X) - m_Y]^2\}} = \sqrt{\frac{\mathbf{D}\{m_Y(X)\}}{\mathbf{D}\{Y\}}}; \quad (4.32)$$

$$\eta_{X/Y} = \frac{1}{\sigma_X} \sqrt{\mathbf{M}\{[m_X(Y) - m_X]^2\}} = \sqrt{\frac{\mathbf{D}\{m_X(Y)\}}{\mathbf{D}\{X\}}}. \quad (4.33)$$

Из соотношений (4.32)–(4.33) можно получить

$$\zeta_{Y|X}^2 = 1 - \eta_{Y|X}^2; \quad (4.34)$$

$$\zeta_{X|Y}^2 = 1 - \eta_{X|Y}^2; \quad (4.35)$$

$$\eta_{Y|X}^2 = \rho_{XY}^2 + \frac{1}{\sigma_Y^2} \mathbf{M}\{[m_Y(X) - \tilde{m}_Y(X)]^2\} = d_{Y|X}; \quad (4.36)$$

$$\eta_{X|Y}^2 = \rho_{XY}^2 + \frac{1}{\sigma_X^2} \mathbf{M}\{[m_X(Y) - \tilde{m}_X(Y)]^2\} = d_{X|Y}, \quad (4.37)$$

где ζ – коэффициент квадратично усредненной вариации; $\tilde{m}_Y(x)$, $\tilde{m}_X(y)$ – среднеквадратические (СК) прямые регрессии; d – коэффициент детерминации.

При этом видно, что значения $\zeta_{Y|X}$ и $\zeta_{X|Y}$ отражают степень рассеяния Y относительно функций регрессии, а $\eta_{Y|X}$ и $\eta_{X|Y}$ – степень нелинейности функции регрессии.

Нетрудно убедиться, что если в качестве аппроксимирующих функций $m_Y(x)$, $m_X(y)$ выбрать линейную, т. е. положить $\tilde{m}_Y(x) = ax + b$, $\tilde{m}_X(y) = cy + d$, то

$$\tilde{m}_Y(x) = m_Y + \frac{\sigma_Y}{\sigma_X} \rho_{XY} (x - m_X) = m_Y + r_{Y|X} (x - m_X); \quad (4.38)$$

$$\tilde{m}_X(y) = m_X + \frac{\sigma_X}{\sigma_Y} \rho_{XY} (y - m_Y) = m_X + r_{X|Y} (y - m_Y), \quad (4.39)$$

т. е. если истинная функция регрессии $m_Y(x)$, $m_X(y)$ прямая, то она совпадает с прямой среднеквадратической регрессии $\tilde{m}_Y(x)$ и/или $\tilde{m}_X(y)$ соответственно.

Кроме того, нетрудно показать, что

$$\mathbf{D}\{m_Y(X)\} = \sigma_Y^2 \rho_{XY}^2 + \mathbf{M}\left\{[m_Y(X) - \tilde{m}_Y(X)]^2\right\}; \quad (4.40)$$

$$\mathbf{D}\{m_X(Y)\} = \sigma_X^2 \rho_{XY}^2 + \mathbf{M}\left\{[m_X(Y) - \tilde{m}_X(Y)]^2\right\}. \quad (4.41)$$

Если функции регрессии линейные, то

$$\sigma_{2Y}^2 = \mathbf{M}\left\{\sigma_Y^2(X)\right\} = \sigma_Y^2 (1 - \rho_{XY}^2) = \Delta_{1Y}^2; \quad (4.42)$$

$$\sigma_{2X}^2 = \mathbf{M}\left\{\sigma_X^2(Y)\right\} = \sigma_X^2 (1 - \rho_{XY}^2) = \Delta_{1X}^2, \quad (4.43)$$

что лишний раз подтверждает тот факт, что σ_2^2 характеризует вариации величин относительно своей функции регрессии.

В заключение заметим следующее.

1. Распределения (например, нормальное, Гаусса), у которых $\sigma_Y(x) = \sigma_Y$ и $\sigma_X(y) = \sigma_X$ для всех x, y , называются **гомоскедастическими**¹, а величины X и Y с такими распределениями – одинаково рассеянными, равнорассеянными, равноизменчивыми. В противном случае распределения называются **гетероскедастическими**², а величины X и Y – неодинаково, различно рассеянными.

2. Как видно из разложений дисперсий, ни одна из условных дисперсий не может быть больше соответствующей безусловной (одномерной, частной, маргинальной).

3. Если функции регрессии $m_X(y)$ и $m_Y(x)$ могут быть использованы для количественного прогнозирования Y при известных x или X по y , то $\sigma_Y^2(x)$, $\sigma_X^2(y)$ или лучше $\sigma_Y(x)$, $\sigma_X(y)$ являются характери-

¹ От. греч. homos – равный, одинаковый.

² От. греч. heterogenes – разнородный, неоднородный.

стиками, влияющими на качество прогноза для каждого конкретного x или y . Качество же аппроксимации (и прогноза) вдоль всех возможных значений аргументов функции регрессии характеризуются средним квадратом $\Delta_{\min}^2 = \sigma^2(1 - \eta^2)$ (ср. с (4.24), (4.33)–(4.37), (4.42), (4.43) и см. п. 6 в табл. П2.1 и п. 8, 9 в табл. П2.2 прил. 2).

Как видно из (4.36), (4.37), КО $\eta_{Y|X}$ и $\eta_{X|Y}$ характеризуют степень средней нелинейности функций регрессии, т. е. усредненное вдоль всех значений аргумента квадратичное отклонение функции регрессии от среднеквадратичной прямой регрессии. Это позволяет ввести *коэффициент нелинейности функции* регрессии в виде

$$v_{Y|X} = \sqrt{1 - \rho_{XY}^2 / \eta_{Y|X}^2}; \quad (4.44)$$

$$v_{X|Y} = \sqrt{1 - \rho_{XY}^2 / \eta_{X|Y}^2}. \quad (4.45)$$

В частности, если $m_Y(x) = \tilde{m}_Y(x)$, т. е. функция регрессии является прямой, то

$$\eta_{Y|X} = |\rho_{XY}|; \quad (4.46)$$

$$v_{Y|X} = 0. \quad (4.47)$$

Аналогичное справедливо и для $m_X(y)$. Другое дело, когда связь между X и Y нелинейная статистическая или функциональная (см. рис. 4.4 и 4.5). В этом случае из равенства $\rho_{XY} = 0^1$, т. е. из *некоррелированности* X и Y , следует, что X и Y либо независимы, либо зависимы, но эта зависимость нелинейная, неоднозначная в области задания величин X и Y .

Сопоставляя ρ_{XY} , $|\rho_{XY}|$, $\eta_{Y|X}$ и $\eta_{X|Y}$, можно более подробно описать тенденцию, вид, характер и степень тесноты связи величин (см. прил. 2, табл. П2.4).

Из определений и свойств корреляционных отношений $\eta_{Y|X}$ и $\eta_{X|Y}$ видно, что они теоретически предпочтительнее коэффициентов корреляции.

¹ Если $\rho_{XY} = 0$, то величины X и Y называются *некоррелированными*.

ляции как характеристик нелинейной связи. Однако как индикатор связи КО имеют два принципиальных недостатка. Во-первых, они не чувствительны к направлению связи, т. е. не позволяют определить, изменяются ли X и Y в одном или противоположных направлениях. Во-вторых, сложность их определения: чтобы найти КО, надо знать или прямо либо косвенно оценить по выборке функции регрессии, что является сложной задачей, и, кроме того, знание функций регрессии позволяет ответить на многие вопросы о взаимосвязи величин без знания КО или КК.

Коэффициент конкорреляции (ККК). В связи с этим рассмотрим еще одну характеристику связи – **коэффициент конкорреляции**¹ χ_{XY} [5, 20]. Он вводится соотношением

$$\chi_{XY} = \frac{K_{XY}}{\sqrt{K_{XX}K_{YY}}}, \quad (4.48)$$

где

$$K_{XY} = \mathbf{M}\left\{[F_X(X) - \mathbf{M}\{F_X(X)\}][F_Y(Y) - \mathbf{M}\{F_Y(Y)\}]\right\}; \quad (4.49)$$

здесь $F_X(x)$ – функция распределения величины X .

Отметим его некоторые важные свойства. Во-первых, он существует для всех распределений, т. е. для всех случайных векторов. Это связано с тем, что после преобразования случайной величины X в $\Xi = F_X(X)$ и Y в $\Psi = F_Y(Y)$ значения Ξ и Ψ принадлежат диапазону $[0, 1]$. Следовательно для Ξ и Ψ всегда разрешима проблема моментов. Более того, если $F_X(x)$ и $F_Y(y)$ – абсолютно непрерывные функции, то Ξ и Ψ имеют равномерные на $[0, 1]$ распределения. При этом $\mathbf{M}\{\Xi\} = \mathbf{M}\{F_X(X)\} = 1/2$, $\mathbf{M}\{\Psi\} = \mathbf{M}\{F_Y(Y)\} = 1/2$, $K_{XX} = K_{YY} = 1/12$. Следовательно, для всех абсолютно непрерывных распределений имеем

$$\chi_{XY} = 12\mathbf{M}\{F_X(X)F_Y(Y)\} - 3. \quad (4.50)$$

¹ От начального слога *лат. constans* – постоянный и *позднелат. correlatio* – соотношение, взаимосвязь, т. е. сохраняющий информацию о связи.

Поэтому для оценивания χ_{XY} можно использовать выборочный (эмпирический) коэффициент корреляции Спирмена [20] или оценку

$$\hat{\chi}_{XY} = \frac{\hat{K}_{XY}}{\sqrt{\hat{K}_{XX}\hat{K}_{YY}}}; \quad (4.51)$$

$$K_{XY} = \frac{1}{N-1} \sum_{i=1}^N \left[\hat{F}_X(x_i) - \frac{1}{N} \sum_{j=1}^N \hat{F}_X(x_j) \right] \times \\ \times \left[\hat{F}_Y(y_i) - \frac{1}{N} \sum_{j=1}^N \hat{F}_Y(y_j) \right], \quad (4.52)$$

а для абсолютно непрерывных распределений

$$\chi_{XY} = \frac{12}{N} \sum_{i=1}^N \hat{F}_X(x_i) \hat{F}_Y(y_i) - 3, \quad (4.53)$$

где $\hat{F}_X(x)$ – любая состоятельная оценка функции распределения $F_X(x)$.

Во-вторых, любое монотонное взаимно однозначное преобразование X и/или Y не изменяет модуля $|\chi_{XY}|$ (см. свойство 8 в табл. П2.3 прил. 2). Именно это свойство породило название характеристики *конкор* (см. сноску на стр. 192) как коэффициента, сохраняющего информацию о связи, и позволяет называть ККК *верускорреляцией*¹, иначе *истинной корреляцией*.

Среди парных характеристик связи особый интерес представляют различные информационные коэффициенты, например

$$R_{X,Y} = \frac{2\mathbf{I}(X,Y)}{H(X) + H(Y)}, \quad (4.54)$$

¹ От *лат. verus* – истинный, основанный на истине или действительности; действительный; настоящий; правдивый; заслуживающий доверия.

где $\mathbf{I}(X, Y)$ – мера количества информации по К. Шеннону; $H(X)$, $H(Y)$ – энтропии (по К. Шеннону, см. далее) случайных величин X и Y , а также статистика χ^2 как мера связи:

$$\chi^2 = N \left(\sum_{i=1}^I \sum_{j=1}^J \frac{N_{ij}^2}{N_{i\bullet} N_{\bullet j}} - 1 \right). \quad (4.55)$$

Здесь N_{ij} – число элементов выборки, попадающих в таблицу сопряженности (ячейки таблицы рассеянности) переменных x и y размером $I \times J$, $i = 1, 2, \dots, I$, I – число строк таблицы; $j = 1, 2, \dots, J$, J – число столбцов таблицы; $N = \sum_{i=1}^I \sum_{j=1}^J N_{ij} = \sum_{i=1}^I N_{i\bullet} = \sum_{j=1}^J N_{\bullet j}$ – объем выборки;

$$N_{i\bullet} = \sum_{j=1}^J N_{ij} \text{ – сумма количества элементов по строке; } N_{\bullet j} = \sum_{i=1}^I N_{ij} \text{ –}$$

сумма по столбцу. Заметим, что $R_{X,Y}$ и χ^2 равны нулю тогда и только тогда, когда X и Y независимы, а $R_{X,Y} = 1$ – тогда и только тогда, когда между X и Y существует взаимно однозначная функциональная зависимость. Отметим также, что на практике используются различные другие характеристики парной связи признаков, измеренных в разных шкалах, в том числе вводимых с целью различить причину и следствие, силу, тесноту связи и т. п., а также приводятся итоги их имитационного сравнения, подчеркивается важность увязки характеристики связи с измерительными шкалами. Показано, что для всех шкал применимы коэффициенты связи χ^2 Пирсона, построенный на χ^2 коэффициент Чупрова и информационные коэффициенты, в то время как коэффициент корреляции Спирмена (выборочный аналог конкорв) пригоден для характеристики связи данных, измеренных в шкалах порядка, интервалов, отношений. При этом подчеркивается, что использование неинвариантного в данной шкале коэффициента связи приводит к привнесению неверной информации.

Корреляционные матрицы

Теперь рассмотрим характеристики связи n -мерного вектора ${}_n\mathbf{X} = (X_1, \dots, X_n)$. Это прежде всего **матрица парных корреляций** (МПК) или **корреляционная матрица** (КМ) $\rho = \{ \rho_{ij}, i, j = \overline{1, n} \}$, где ρ_{ij} – коэффициент корреляции **пары** величин X_i и X_j . Ясно, что $\rho_{ii} = 1$, $|\rho_{ij}| \leq 1$, а матрица ρ является симметричной неотрицательно определенной (теплицевой), т. е. все ее главные определители, составленные из элементов, расположенных в левом верхнем углу матрицы, не отрицательны. Это свойство КМ можно учитывать для уточнения оценок парных КК, а также обязательно надо учитывать при имитации случайных векторов с требуемой КМ при задании КМ (см. [20], а также подпараграф 4.3.8).

Аналогично можно ввести матрицы парных корреляционных отношений, парных конкорреляций и других характеристик парной связи [20].

Другие характеристики множественной связи

Как следует из предыдущего изложения, величины X_i и X_j , входящие в вектор $\mathbf{X} = (X_1, \dots, X_n)$, описывающий n показателей исследуемого объекта, могут быть независимыми, зависимыми стохастически (случайно), функционально или смешанным образом – функционально-стохастически, как между собой (парно), так и с другими элементами X_k , $k = \overline{1, n}$, $k \neq i, j$. Эти зависимости могут проявляться напрямую, косвенно и даже ложно. Напрямую, когда мы исследуем связь между X_i , на значения которого влияют другие X_k , $k = \overline{1, n}$, $k \neq i$, и этими другими либо связь между X_i и X_j , если одна из них связана с другой независимо от остальных X_k , $k = \overline{1, n}$, $k \neq i, j$, либо если две величины X_i и X_j или три и более величины связаны между собой независимо от остальных и т. п. Косвенная связь – это когда рассматриваемые величины причинно между собой не связаны, а кажущаяся, виртуально проявляющаяся при ее аналитическом, имитационном или графическом исследовании связь является следствием стохастической

связи каждой из них (например, X_i и X_j) с одной или несколькими остальными (X_k , $k = \overline{1, n}$, $k \neq i, j$). Ложная связь, частный случай косвенной, когда, например, X_i и X_j причинно, стохастически или напрямую функционально не связаны, а каждая из них функционально связана с одной (X_k) или несколькими остальными величинами¹.

Решение подобных задач следует начинать с выявления функциональных связей.

Как мы уже знаем, для этого необходимо находить соответствующие функции регрессии, предварительно устанавливая сам факт функциональной связи, используя набор соответствующих КК, КО, ККК, т. е. $\rho_{X_i X_j} = \rho_{ij}$, $\eta_{X_i | X_j} = \eta_{i|j}$, $\eta_{X_j | X_i} = \eta_{j|i}$, $\chi_{X_i X_j} = \chi_{ij}$ и т. д.

После определения функциональной связи можно перейти к исследованию стохастической связи, «убрав» функциональную связь. Рассмотрим некоторые из подобных приемов. Прежде чем переходить к ним, сделаем очень важное замечание.

Использование только парных характеристик связи любых пар случайных величин X_i и X_j n -мерного вектора $\mathbf{X} = (X_1, \dots, X_n)$ может привести к ложным, неправильным выводам, когда на значения X_i , X_j влияют другие величины вектора. Так, если коэффициент парной корреляции $\rho_{X_i X_j}$ уменьшается или даже равняется нулю при фиксированных значениях других случайных величин (т. е. при переходе к частному КК по сравнению с парным КК), то следует ожидать, что зависимость между X_i и X_j полностью, определяющим образом,

¹ Заметим, что если $X_i = f_i(X_k)$ и $X_j = f_j(X_k)$, то при взаимно однозначных f_i и f_j между X_i и X_j будет иметь место связь $X_i = \varphi(X_j)$ или $X_j = \psi(X_i)$. В этом случае имеет место лжефункциональная связь между X_i и X_j , «наведенная» величиной X_k . Этот факт может быть обнаружен, если установить факт связи X_i , X_j , X_k , отдельно проанализировать парные и множественные связи этих величин, а затем сопоставить между собой значения индикатора связи либо введением соответствующих вектор-характеристик (см. далее).

значительно или значимо (смотря, как уменьшается $\rho_{X_i X_j}$) определяется благодаря третьим факторам, т. е. влиянием других величин вектора \mathbf{X} , которые вносят «ложную» корреляцию (наводят ложную связь) между X_i и X_j . Если же при фиксации значений третьих факторов $\rho_{X_i X_j}$ (т. е. частная корреляция – см. далее) возрастает, то это означает, что третьи факторы «маскировали» истинную взаимосвязь X_i и X_j , декоррелировали эту связь.

Именно поэтому при анализе n -мерных векторов необходимо помимо парных находить другие множественные характеристики связи, к рассмотрению которых мы и переходим.

Другие числовые характеристики связи элементов n -мерного вектора $\mathbf{X} = (X_1, \dots, X_n)$

Рассмотрим вначале ситуацию, когда необходимо оценить связь между любой X_i , $i = \overline{1, n}$, и всеми другими X_k , $k = \overline{1, n}$, $k \neq i$. Тогда вначале априори, апостериори или смешанными приемами определяется функция регрессии $m_{X_i}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. После этого можно найти *парные* КК, КО, ККК между X_i и

$$Y = m_{X_i}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n), \quad (4.56)$$

т. е. охарактеризовать наличие стохастической или функциональной связи, а также тенденцию, вид, характер и степень тесноты связи X_i и всех других величин X_k , $k = \overline{1, n}$, $k \neq i$, описанными выше приемами.

Однако, к сожалению, определять с приемлемой точностью функцию регрессии затруднительно даже в двумерном, а тем более, в многомерном случае, если объем данных невелик. Выход из этого в использовании зависимости X_i от X_k , $k = \overline{1, n}$, $k \neq i$, на уровне аппроксимации функции регрессии линейной среднеквадратической регрессией, нахождение которой существенно проще и сводится к нахождению только попарных корреляций. В этом случае вместо

$Y = m_{X_i}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ вводится в рассмотрение среднеквадратическая линейная регрессия

$$Z = \sum_{\substack{k=1, \\ k \neq i}}^n \tilde{\alpha}_k X_k = \tilde{m}_{X_i}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) = X'_i, \quad (4.57)$$

где $\tilde{\alpha}_k$ есть значения α_k , обеспечивающие минимум

$$\begin{aligned} \min_{\mathbf{a}} \Delta_{X_i}^2 &= \min_{\mathbf{a}} \mathbf{M} \left\{ \left[X_i - \alpha_0 - \sum_{\substack{k=1 \\ k \neq i}}^n \alpha_k X_k \right]^2 \right\} = \\ &= \mathbf{M} \left\{ [X_i - X'_i]^2 \right\}, \end{aligned} \quad (4.58)$$

и рассматривают **сводный (совокупный, множественный) коэффициент корреляции** (СКК) между X_i и совокупностью $X'_i = (X_k, k = \overline{1, n}, k \neq i)$ (!):

$$r_{i \cdot (\cdot)'} = r_{i \cdot (1, \dots, n)'} = \sqrt{1 - \frac{\|\mathbf{\rho}\|}{\|\mathbf{\rho}_{ii}\|}} = \sqrt{1 - \frac{\mathbf{D}\{X'_i\}}{\mathbf{D}\{X_i\}}}, \quad (4.59)$$

где $\|\mathbf{\rho}\|$ – определитель корреляционной матрицы $\mathbf{\rho} = (\rho_{ij}, i, j = \overline{1, n})$ всего вектора $\mathbf{X} = (X_1, \dots, X_n)$; $\|\mathbf{\rho}_{ij}\|$ – определитель, полученный из $\|\mathbf{\rho}\|$ вычеркиванием i -й строки и j -го столбца и умноженный на $(-1)^{i+j}$, т. е. алгебраическое дополнение, $(1, \dots, n)' = (1, 2, \dots, i-1, i+1, \dots, n)$.

Как следует из (4.59), СКК есть частная разновидность корреляционного отношения вида (4.32), (4.33) в случае линейной регрессии. Он определяет модуль парной корреляции между X_i и линейной среднеквадратической регрессией Z (4.57) как функцией величины $X_k, k = \overline{1, n}, k \neq i$, и совпадает с максимальным значением модуля

парного коэффициента корреляции $|\rho_{X_i Y}| = |\rho_{X_i Z}|$ между X_i и Y (4.56) (максимум достигается, когда $Y = Z$).

Отсюда следует, что СКК характеризует (подобно модулю коэффициента корреляции) степень тесноты линейной связи X_i и X'_i : чем лучше случайная величина X_i приближается линейными комбинациями других случайных величин вектора X_k , $k = \overline{1, n}$, $k \neq i$, тем ближе СКК к единице; чем хуже линейное приближение, тем этот коэффициент ближе к нулю. В частности, если $m_{X_i}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = m_{X_i}$, то СКК равен нулю. Ясно, что всегда имеют место неравенства $0 \leq r_{i(1, \dots, n)} \leq 1$.

Обратим внимание на следующие свойства СКК $r_{i(\cdot)}$:

1) значение $r_{i(\cdot)}$ всегда неотрицательно и $0 \leq r_{i(\cdot)} \leq 1$;

2) если $r_{i(\cdot)} = 0$, то X_i не может быть линейно связано с другими $(X_1, \dots, X_n)'$. Однако при этом возможна нелинейная связь X_i с $(X_1, \dots, X_n)'$;

3) если $r_{i(\cdot)} = 1$, то это означает, что X_i связана с остальными величинами $(X_1, \dots, X_n)'$ линейной функциональной зависимостью;

4) если $r_{i(\cdot)}$ принимает значение между 0 и 1, то связь между X_i и $(X_1, \dots, X_n)'$ может быть как статистической линейной, так и нелинейной, в том числе функциональной. Чем ближе $r_{i(\cdot)}$ к единице, тем теснее связана X_i с $(X_1, \dots, X_n)'$ и эта связь ближе к линейной функциональной.

В качестве оценки СКК $r_{i(1, \dots, n)}$ можно использовать его выборочное значение

$$\hat{r}_{i(1, \dots, n)} = \sqrt{1 - \frac{\|\hat{\rho}\|}{\|\hat{\rho}_{ii}\|}} \quad (4.60)$$

либо

$$\hat{r}_{i(1,\dots,n)'} = \sqrt{1 - \frac{\hat{\mathbf{D}}\{X'_i\}}{\hat{\mathbf{D}}\{X_i\}}}, \quad (4.61)$$

где $\hat{\boldsymbol{\rho}} = \{\hat{\rho}_{ij}, i, j = \overline{1, n}\}$, $\hat{\rho}_{ij}$ – оценки (выборочные значения) парных коэффициентов корреляции,

$$\hat{\mathbf{D}}\{X'_i\} = \frac{1}{N-1} \sum_{l=1}^N (\hat{x}'_{i,l} - \bar{x}'_{i,l})^2; \quad (4.62)$$

$$\hat{\mathbf{D}}\{X_i\} = \frac{1}{N-1} \sum_{l=1}^N (x_{i,l} - \bar{x}_i)^2; \quad (4.63)$$

$$\bar{x}_i = \frac{1}{N} \sum_{l=1}^N x_{i,l}; \quad (4.64)$$

$$\hat{x}_{i,l} = \hat{\alpha}_0 + \frac{1}{N} \sum_{\substack{k=1, \\ k \neq i}}^N \hat{\alpha}_k x_{k,l}, \quad (4.65)$$

где $x_{i,l}$, $x_{k,l}$ – значение X_i , X_k , полученное в l -м ($l = \overline{1, N}$) опыте эксперимента; $\hat{\alpha}_0$, $\hat{\alpha}_k$, $k = \overline{1, n}$, $k \neq i$ – оптимальные значения α , найденные методом наименьших квадратов, т. е. минимизацией

$$\sum_{l=1}^N \left[x_{i,l} - \alpha_0 - \sum_{\substack{k=1 \\ k \neq i}}^N \alpha_k x_{k,l} \right]^2 = \min_{\boldsymbol{\alpha}}. \quad (4.66)$$

По аналогии с СКК можно использовать совокупный коэффициент конкорреляции I рода, если в соотношениях типа (1.54), (1.55) вместо X_k ввести $\varphi_k(X_k)$, где $\varphi_k(\cdot)$ – подбираемые взаимно однозначные функции, обеспечивающие минимум аналога (4.58), т. е.

$$\min_{\alpha, \varphi} \Delta_{\varphi}^2 = \min_{\mathbf{M}} \left\{ \left[\begin{array}{c} X_i - \alpha_0 - \sum_{\substack{k=1, \\ k \neq i}}^n \alpha_k \varphi_k(X_k) \end{array} \right]^2 \right\} = \mathbf{M} \left\{ \left[X_i - X'_i \right]^2 \right\}, \quad (4.67)$$

а в (1.57) заменить корреляционную матрицу \mathbf{p} на конкорреляционную χ .

Сводный коэффициент конкорреляции II рода (СККК_{II}) можно ввести, если заменить в (4.60) корреляционные определители $\|\mathbf{p}\|$, $\|\mathbf{p}_{ii}\|$ на конкорреляционные $\|\chi\|$, $\|\chi_{ii}\|$. Полученный таким образом СККК_{II} будет соответствовать коэффициенту корреляции между $\Xi_i = F_{X_i}(X_i)$ и функцией регрессии

$$\tilde{m}_{\Xi_i}(\Xi_1, \dots, \Xi_{i-1}, \Xi_{i+1}, \dots, \Xi_n), \quad (4.68)$$

где $\Xi_k = F_{X_k}(X_k)$, $F_{X_k}(x)$ – функция распределения величины X_k .

Теперь рассмотрим зависимость пар величин (X_i, X_j) от других величин X_k , $k = \overline{1, n}$, $k \neq i, j$. Чтобы убрать влияние на X_i и X_j функциональных связей их с X_k , $k = \overline{1, n}$, $k \neq i, j$, образуем величины (полагаем условно, что $j > i$)

$$X_i'' = x_i - m_{X_i}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, X_n); \quad (4.69)$$

$$X_j'' = x_j - m_{X_j}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, X_n) \quad (4.70)$$

и

$$Y_i = X_i - X_i'', \quad Y_j = X_j - X_j''. \quad (4.71)$$

Затем рассмотрим КК $\rho_{Y_i Y_j}$, КО $\eta_{Y_i|Y_j}$, $\eta_{Y_j|Y_i}$, ККК $\chi_{Y_i Y_j}$ и использовать те приемы, которые были описаны выше. В связи с затруднением определения функции регрессии в условиях малых выборочных данных опять-таки ограничимся аппроксимацией функциональных зависимостей X_i и X_j от X_k , $k = \overline{1, n}$, $k \neq i, j$, прямыми средней квадратической регрессии. Иными словами, вместо (4.70) рассмотрим

$$Z_i = x_i - \tilde{m}_{X_i} (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, X_n); \quad (4.72)$$

$$Z_j = x_j - \tilde{m}_{X_j} (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, X_n), \quad (4.73)$$

где $\tilde{m}_{X_i}(\cdot)$ – соответствующие линейные среднеквадратические регрессии (см. (4.57)).

Теперь можем найти КК, КО и ККК для Z_i , Z_j . Коэффициент корреляции между Z_i , Z_j называется **частным коэффициентом корреляции** (ЧКК) и находится как

$$\rho_{ij \cdot (1, \dots, n)} = \rho_{Z_i Z_j} = -\frac{\|\mathbf{p}\|}{\|\mathbf{p}_{ii}\| \|\mathbf{p}_{jj}\|}, \quad (4.74)$$

где \mathbf{p} – корреляционная матрица вектора ${}_n \mathbf{X} = (X_1, \dots, X_n)$; $\|\mathbf{p}\|$ – определитель \mathbf{p} ; $\|\mathbf{p}_{ij}\|$ – алгебраическое дополнение элемента $\rho_{X_i X_j} = \rho_{ij}$ в корреляционной матрице \mathbf{p} . Понятно, что ЧКК имеет все свойства коэффициента корреляции. Ясно также, что ЧКК $\rho_{ij \cdot (1, \dots, n)}$ характеризует «чистую» линейную статистическую связь двух случайных величин X_i и X_j вектора $\mathbf{X} = (X_1, \dots, X_n)$, когда исключается линейное влияние величин (факторов) X_k , $k = \overline{1, n}$, $k \neq i, j$ на X_i , X_j .

Из (4.74) следует, что если величины, образующие n -мерный вектор, попарно независимы или хотя бы некоррелированы, то ЧКК $\rho_{ij \cdot (1, \dots, n)}$ будет равен нулю. С другой стороны, если хотя бы одна из

величин X_i и X_j есть линейная комбинация от X_k , $k = \overline{1, n}$, $k \neq i, j$, то, как и следовало ожидать, ЧКК $|\rho_{ij \cdot (1, \dots, n)}|$ будет также равняться нулю.

Если же X_i и X_j не зависят от X_k , $k = \overline{1, n}$, $k \neq i, j$, то $\rho_{ij \cdot (1, \dots, n)} = \rho_{ij} = \rho_{X_i X_j}$.

Аналогично разложению дисперсии можно разложить коэффициент корреляции между X_i и X_j с учетом линейного среднеквадратического влияния на них величин X_k , $k = \overline{1, n}$, $k \neq i, j$. Можно убедиться, что парный КК ρ_{ij} между X_i и X_j разложим в виде

$$\rho_{ij} = \frac{\mathbf{M} \left\{ \left[\mathbf{M} \left\{ X_i \mid (X_1, \dots, X_n)'' \right\} - \mathbf{M} \{ X_i \} \right] \times \left[\mathbf{M} \left\{ X_j \mid (X_1, \dots, X_n)'' \right\} - \mathbf{M} \{ X_j \} \right] \right\}}{\sigma_{X_i} \sigma_{X_j}} + \frac{\mathbf{M} \left\{ \left[X_i - \mathbf{M} \left\{ X_i \mid (X_1, \dots, X_n)'' \right\} \right] \times \left[X_j - \mathbf{M} \left\{ X_j \mid (X_1, \dots, X_n)'' \right\} \right] \right\}}{\sigma_{X_i} \sigma_{X_j}}. \quad (4.75)$$

Иными словами, $\rho_{ij} = \rho'_{ij} + \rho''_{ij}$.

Здесь ρ'_{ij} характеризует ту часть парной корреляции, которая вызвана влиянием случайных величин X_k , $k = \overline{1, n}$, $k \neq i, j$. Второе же слагаемое определяет «чистую» статистическую линейную (корреляционную) связь X_i и X_j , когда функциональное (через регрессию) воздействие X_k , $k = \overline{1, n}$, $k \neq i, j$, на них исключено.

Заметим, что ρ'_{ij} при замене $\mathbf{M}\left\{X_i \mid (X_1, \dots, X_n)''\right\}$ и $\mathbf{M}\left\{X_j \mid (X_1, \dots, X_n)''\right\}$, т. е. $m_{X_i}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$ и $m_{X_j}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$ на их линейную среднеквадратическую аппроксимацию $\tilde{m}_{X_i}(\cdot)$ и $\tilde{m}_{X_j}(\cdot)$, представляет собой ЧКК $\rho_{ij \cdot (1, \dots, n)''}$, в то время как второе слагаемое ρ''_{ij} есть усредненный КК величин X_i'' и X_j'' (см. (4.69)–(4.71)).

Сравнивая ρ_{ij} с $\rho_{ij \cdot (1, \dots, n)''}$, можно судить о влиянии X_k , $k = \overline{1, n}$, $k \neq i, j$, на X_i , X_j . Так, если $|\rho_{ij}| = |\rho_{ij \cdot (1, \dots, n)''}|$, то это означает, что X_k , $k = \overline{1, n}$, $k \neq i, j$, не влияет на уровне линейной зависимости на X_i , X_j . Если $|\rho_{ij}| < |\rho_{ij \cdot (1, \dots, n)''}|$, то это означает, что третьи факторы (величины X_k , $k = \overline{1, n}$, $k \neq i, j$) влияют на зависимость между X_i и X_j в среднем в противоположных направлениях, т. е. «маскируют» истинную («чистую») связь между X_i и X_j . Наоборот, если $|\rho_{ij}| > |\rho_{ij \cdot (1, \dots, n)''}|$, то это также означает, что X_k , $k = \overline{1, n}$, $k \neq i, j$, влияют на X_i , X_j , внося «ложную» корреляцию.

Рассмотрим такой пример. Пусть нас интересует связь между случайными величинами X , Y и Z . Для этого мы нашли их парные КК ρ_{XY} , ρ_{XZ} , ρ_{YZ} . Если $\rho_{YZ} = 0$, то ρ_{XZ} и ρ_{XY} можно было бы принять в качестве характеристик «чистой» связи между X и Z , X и Y соответственно. Если же $\rho_{YZ} \neq 0$, то «чистая» связь X с Y и X с Z «замаскирована» (затуманена) наличием связи между Y и Z . Найдем поэтому частные КК (например, $\rho_{XY \cdot Z}$). Этот ЧКК характеризует «чистую» (на уровне линейного влияния третьего фактора) тесноту линейной связи между X и Y при исключении влияния на них величины Z . Если $|\rho_{XY}| = |\rho_{XY \cdot Z}|$, то Z не влияет (линейно) на X и Y . Если $|\rho_{XY}| < |\rho_{XY \cdot Z}|$,

то это означает, что Z влияет на X и Y так, что затушевывает, маскирует связь между X и Y . Если же $|\rho_{XY}| > |\rho_{XY \cdot Z}|$, то Z влияет на X и Y , внося «ложную» (дополнительную) корреляцию.

Аналогично ЧКК можно ввести частные корреляционные отношения, частные коэффициенты конкорреляции I и II типов. Можно убедиться в том, что

$$\rho_{ij \cdot (1 \dots n)} = \frac{\rho_{ij \cdot (2, \dots, n)} - \rho_{i1 \cdot (2, \dots, n)} \rho_{j1 \cdot (2, \dots, n)}}{\sqrt{(1 - \rho_{i1 \cdot (2, \dots, n)}^2)(1 - \rho_{j1 \cdot (2, \dots, n)}^2)}}. \quad (4.76)$$

Эта формула позволяет последовательно определять $\rho_{ij \cdot n}$, $\rho_{ij \cdot (n-1, n)}$, ..., $\rho_{12 \cdot (3, \dots, n)}$ по ρ_{ks} , $k, s = \overline{1, n}$.

В частности, при $n = 3$, т. е. для трех величин X_1 , X_2 и X_3 , из (4.76) получаем, например,

$$\rho_{12 \cdot 3} = (\rho_{12} - \rho_{13}\rho_{23}) \left[(1 - \rho_{13}^2)(1 - \rho_{23}^2) \right]^{-\frac{1}{2}}; \quad (4.77)$$

$$r_{1 \cdot (23)} = \sqrt{(\rho_{12}^2 + \rho_{13}^2 - 2\rho_{12}\rho_{13}\rho_{23}) / (1 - \rho_{23}^2)}. \quad (4.78)$$

Заметим, во-первых, что $r_{i \cdot (1, \dots, n)} \in [0, 1]$, в то время как $\rho_{ij \cdot (1, \dots, n)} \in [-1, 1]$. Во-вторых, значения множественного коэффициента корреляции $r_{i \cdot (1, \dots, n)}$ зависят от *всех* частных корреляций и, если даже некоторые из них принимают нулевое значение, $r_{i \cdot (1, \dots, n)}$ может равняться единице. Как следует из (4.76), всегда $r_{i \cdot (1, \dots, n)} \geq |\rho_{ji \cdot (1, \dots, n)}|$. Следовательно, если $r_{i \cdot (1, \dots, n)} = 0$, то все $\rho_{ij \cdot (1, \dots, n)}$ при любых $j = \overline{1, n}$, $j \neq i$, также будут равны 0, т. е. X_i будет некоррелирована со всеми остальными X_j , $j \neq i$. В то же время, если $r_{i \cdot (1, \dots, n)} = 1$, то, по крайней мере, один коэффициент $\rho_{ij \cdot (1, \dots, n)}$ должен быть равен по модулю 1,

т. е. $\sigma_{i(1,\dots,n)}^2 = 0$ и все точки распределения лежат на плоскости (линии) регрессии, или, иными словами, X_i является строго линейной функцией от $(X_1, \dots, X_n)'$. Эти факты позволяют предостеречь от отбрасывания тех величин $(X_1, \dots, X_n)'$ при прогнозе значений X_i по регрессии, которые слабо коррелированы с X_i . С другой стороны, если каждая пара величин (X_j, X_s) , $j = \overline{1, n}$, $s = \overline{1, n}$, $j \neq i$, $s \neq i$, некоррелирована, то

$$r_{i(1,\dots,n)}' = \sum_{\substack{j=1, \\ j \neq i}}^n \rho_{ij}^2. \quad \text{В-третьих, важным в приложениях и для проверки}$$

значений оценок могут быть соотношения

$$\rho_{ij(1,\dots,n-1)}'' = \frac{\rho_{ij(2,\dots,n)}'' - \rho_{in(1,\dots,n-1)}' \rho_{jn(1,\dots,n-1)}'}{\sqrt{(1 - \rho_{in(1,\dots,n-1)}'^2)(1 - \rho_{jn(1,\dots,n-1)}'^2)}}, \quad (4.78)$$

где $(1, \dots, n-1)'$ не содержит i , а $(1, \dots, n-1)''$ — не содержит j ,

$$\rho_{12} = (\rho_{123} - \rho_{132}\rho_{231}) / \sqrt{(1 - \rho_{132}^2)(1 - \rho_{231}^2)}. \quad (4.79)$$

Для многомерных нормальных распределений $\rho_{ij(1,\dots,n)}''$ есть условный коэффициент корреляции X_i и X_j , т. е. коэффициент, найденный по их условным распределениям, полученным в предположении, что другие величины X_k , $k = \overline{1, n}$, $k \neq i$, $k \neq j$, принимают некоторые фиксированные значения. В общем случае частный коэффициент корреляции $\rho_{ij(1,\dots,n)}''$ характеризует корреляцию между остатками $E_i = X_i - \tilde{m}_{X_i}(X_{ij}'')$ и $E_j = X_j - \tilde{m}_{X_j}(X_{ij}'')$, усредненными вдоль всех значений $X_{ij}'' = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$ по вероятностной мере X_i , X_j и X_{ij}'' , т. е.

$$\rho_{ij(1,\dots,n)}'' = \mathbf{M}\{\hat{E}_i \hat{E}_j\}. \quad (4.80)$$

Между множественным и частным коэффициентами корреляции существует следующая зависимость:

$$r_{i(1,\dots,n)'}^2 = 1 - (1 - \rho_{i1}^2)(1 - \rho_{i2(1)'}^2)(1 - \rho_{i3(1,2)'}^2)\dots(1 - \rho_{in(1,\dots,n-1)'}^2), \quad (4.81)$$

позволяющая вычислить множественный коэффициент по частным.

Следующим обобщением является **частный множественный коэффициент корреляции** между X_i и $X'_{(k+1,n)} = (X_{k+1}, \dots, X_n)'$, исключая зависимость от $X'_{(1,k)} = (X_1, \dots, X_k)'$, где знак «'» означает отсутствие X_i в том наборе, в котором i входит во множество $(k+1, k+2, \dots, n)$ или во множество $(1, 2, \dots, k)$, $k < n$:

$$r_{i(k+1,\dots,n)'(1,\dots,k)'} = \sqrt{\frac{r_{i(1,\dots,n)'}^2 - r_{i(1,\dots,k)'}^2}{1 - r_{i(1,\dots,k)'}^2}}. \quad (4.82)$$

Его квадрат характеризует относительное уменьшение среднего квадрата ошибки предсказания X_i по $\tilde{m}_{X_i}(X_1, \dots, X_k)'$, где знак «'» означает исключение из перечня величины X_i .

Заметим, что коэффициент (4.59) есть частный случай (4.82), когда $(k+1, \dots, n)'$ сводится к одному j , а $(1, \dots, k)'$ заменяется на $(1, \dots, n)''$.

Аналогично может быть получен **коэффициент частной множественной корреляции** заменой r на χ .

Множественные коэффициенты корреляции позволяют характеризовать в среднем качество прогноза X_i (точнее, средний квадрат E_i) с помощью линейной средней квадратической регрессии. При использовании же нелинейной регрессии следует перейти к корреляционным отношениям или конкорам.

Множественное (сводное) корреляционное отношение $\eta_{i(1,\dots,n)'}$ совпадает по определению с корреляционным отношением η_{X_i/x_i} .

Частное корреляционное отношение для $X_{i(1,\dots,n)''}$ и $X_{j(1,\dots,n)''}$ определяется выражениями

$$\eta_{i(1,\dots,n)''} = \mathbf{M} \left\{ \begin{matrix} m_{\circ}^2 \\ X_{i(1,\dots,n)''} \end{matrix} (X_{j(1,\dots,n)''}, (X_1, \dots, X_n)'' \right\} \sigma_{i(1,\dots,n)''}^{-2}. \quad (4.83)$$

Частное множественное корреляционное отношение определяется как

$$\eta_{i(k+1, \dots, n)' \cdot (1, \dots, k)'} = \sqrt{\frac{\eta_{i(1, \dots, n)'}^2 - \eta_{i(1, \dots, k)'}^2}{1 - \eta_{i(1, \dots, k)'}^2}}. \quad (4.84)$$

Его квадрат характеризует относительное уменьшение среднего квадрата ошибки предсказания X_i по $m_{X_i}(X_1, \dots, X_n)'$ по сравнению с предсказанием по $m_{X_i}(X_1, \dots, X_k)'$. Заметим, что всегда при $k < n$ $\eta_{i(1, \dots, k)'} \leq \eta_{i(1, \dots, n)'}$, $|r_{i(1, \dots, k)'}| < |r_{i(1, \dots, n)'}$, а любой **множественный коэффициент корреляции** мажорирует любой парный или частный коэффициент корреляции, характеризующий статическую связь той же результирующей величины X_i .

Наконец, заметим, что в ряде случаев рассматривают коэффициент корреляции $r_{i < j \cdot (1, \dots, n)''}$ между X_i и разностью $X_j - m_{X_j}(X_{i,j}'')$, который называется **частью корреляции**, является несимметричным по отношению к индексам i, j , так как $r_{i < j \cdot (1, \dots, n)''} \neq r_{j < i \cdot (1, \dots, n)''}$ (в отличие от ρ_{ij} и $\rho_{ij \cdot (1, \dots, n)''}$) и удовлетворяет выражениям:

$$r_{i < j \cdot (1, \dots, n)''}^2 \leq \rho_{ij \cdot (1, \dots, n)''}^2; \quad (4.85)$$

$$r_{i < j \cdot (1, \dots, n)''}^2 = \rho_{ij}^2 + r_{i < (n-1) \cdot n}^2 + r_{i < j \cdot (n-2)(n-1) \cdot n}^2 + \dots + r_{i < 1 \cdot (2, \dots, n)''}^2. \quad (4.86)$$

Как уже упоминалось, функции регрессии можно использовать для наилучшего (в смысле среднего квадрата ошибки) прогноза значений одной из величин по известным значениям других величин.

Рассмотрим, например, величины X и Y . Наилучший в смысле среднего квадрата ошибки (погрешности) прогноз значений Y для фиксированного $X = x$ получается, если положить, что прогнозное значение $\hat{Y} = m_Y(x)$. При этом средний квадрат ошибки прогноза Δ^2 , полу-

ченный усреднением вдоль всех значений X и Y , т. е. $\Delta^2 = \mathbf{M} \left\{ [Y - m_Y(X)]^2 \right\}$, равняется

$$\Delta^2 = \Delta_{\min}^2 = \sigma_Y^2 (1 - \eta_{Y/X}^2). \quad (4.87)$$

Если же значение Y прогнозировать по прямой средней квадратической регрессии $\tilde{m}_Y(x)$, то при этом погрешность прогноза Δ^2 будет равна

$$\Delta^2 = \Delta_1^2 = \sigma_Y^2 (1 - \rho_{XY}^2). \quad (4.88)$$

Аналогично, если значение X_i прогнозировать по $m_{X_i}(x_1, \dots, x_n)'$, то погрешность Δ^2 будет равна

$$\Delta^2 = \Delta_{\min}^2 = \sigma_{X_i}^2 (1 - \eta_{i \cdot (1, \dots, n)'}^2) = \sigma_{X_i}^2 (1 - \eta_{X_i | \mathbf{X}_i'}^2). \quad (4.89)$$

Если же X_i прогнозировать по линейной средней квадратической регрессии $\tilde{m}_{X_i}(x_1, \dots, x_n)'$, то получаем

$$\Delta^2 = \Delta_1^2 = \sigma_{X_i}^2 (1 - r_{i \cdot (1, \dots, n)'}^2) \text{ и т. д.} \quad (4.90)$$

Приведенные соотношения позволяют дать еще одну интерпретацию рассмотренным характеристикам связи вектора \mathbf{X} и позволяют, сопоставляя Δ^2 , определять, насколько удачна та или иная функция, используемая для прогнозирования X_i или для аппроксимации функции регрессии $m_{X_i}(\cdot)$, от наилучшей, т. е. от истинной функции регрессии.

В заключение сделаем несколько замечаний.

1. Используя регрессионный, скедастический и другие разновидности корреляционного анализа, можно решить следующие задачи: определить наличие, вид, характер, тесноту связи случайных величин вектора, выявить ее «чистоту», спрогнозировать значения, лежащие за пределами наблюдаемых значений независимых переменных, и определить достоверность прогнозируемых значений.

2. Используя функции регрессии, можно прогнозировать среднее значение признака Y , пользуясь выражением

$$\hat{y} \approx \hat{m}_Y(x_1, \dots, x_n) + \hat{\varepsilon} \quad (4.91)$$

или

$$\hat{y} \approx f(x_1, \dots, x_n) + \varepsilon, \quad (4.92)$$

где $\hat{\cdot}$ – символ оценки; f – функция, найденная по минимуму Δ_Y^2 ; ε – невязка. При этом, как правило, по коэффициенту вариации судят о тесноте связи. Как уже упоминалось, относительно ε выдвигаются некоторые гипотезы. Например, что ε – реализация нормальной случайной величины. Если такая гипотеза подтверждается, то это служит косвенным подтверждением верности подбора функции $f(\cdot)$ для описания зависимости y от (x_1, \dots, x_n) . При прогнозировании значений y по (4.91) и (4.92), зная свойства ε , можно определить погрешность или доверительный интервал возможных прогнозных значений y . Второй путь – это подобрать функцию, аппроксимирующую скеластическую функцию (аналогично аппроксимации функции регрессии), найти ее прогнозное значение на тот же интервал упреждения и по нему судить о возможном рассеянии y на этом отрезке упреждения.

3. При использовании регрессионных аппроксимаций для определения значений Y по x_1, x_2, \dots, x_n необходимо иметь в виду следующее.

Из близости $\eta_{Y|X_1, \dots, X_n}^2$ к $r_{Y \cdot X_1, \dots, X_n}^2$, т. е. из близости $v_{Y|X_1, \dots, X_n}$ к нулю, еще не значит, что зависимость величины Y может быть приемлемо точно описана линейной СК-регрессией. Это лишь означает, что либо факторы X_1, X_2, \dots, X_n несущественны для Y (т. е. по значениям X_1, \dots, X_n нельзя оценить с достаточной уверенностью значения Y), либо необходимо перейти к нелинейному описанию зависимости Y от X_1, \dots, X_n (см. рис. 4.4 и 4.5).

4. При практическом построении регрессионных зависимостей Y от X_1, X_2, \dots, X_n необходимо учитывать важные нюансы.

Во-первых, чтобы каждый вводимый в модель фактор X_1, X_2, \dots, X_n обладал большой вариабельностью влияния на Y для исключения неправильной интерпретации получаемых результатов.

Во-вторых, необходимо либо избегать включения в модель сильно связанных между собой факторов, либо, если они были включены ранее, убрать один из двух или трех сильно связанных факторов.

В-третьих, необходимо учитывать возможность влияния на поведение Y линейных и нелинейных трендов, описываемых регулярными функциями времени. Иными словами, рассматривая Y как случайную величину и исследуя ее развитие во времени, нужно отделить трендовый компонент от Y и рассматривать регрессионные зависимости отклонений Y от тренда. Надо учитывать, что игнорирование наличия трендов в Y , а также в X_1, X_2, \dots, X_n может привести к завышению степени влияния величин X_1, X_2, \dots, X_n на Y , т. е. к *ложной (наведенной) корреляции*. Наличие связи между X_1, \dots, X_n может привести к ошибке: коэффициент детерминации или множественный коэффициент корреляции может быть близок к единице, но модель может обладать малой прогностической способностью. При сильной коррелированности X_1, \dots, X_n необходимо либо уменьшать число предикторов, регрессоров X_1, X_2, \dots, X_n , оставив наименее коррелированные, либо использовать автоматические приемы выбора X_1, \dots, X_n , либо использовать преобразования X_1, \dots, X_n , в том числе нелинейные, в новые некоррелированные переменные.

В-четвертых, необходимо учитывать возможную засоренность экспериментальных значений как Y , так и X_1, X_2, \dots, X_n . В этом случае модель типа (4.92) нужно дополнить аддитивной или мультипликативной составляющей, характеризующей засорение, и использовать различные робастные процедуры (см. далее).

В-пятых, необходимо учитывать, что точность аппроксимации зависимости Y от X_1, X_2, \dots, X_n , следовательно, прогноза Y с использованием этой зависимости, определяется точностью измерений исходных значений x_1, x_2, \dots, x_n , а также тех значений y , по которым строится аппроксимация этой зависимости.

В-шестых, решая прогнозные задачи, необходимо стремиться к тому, чтобы функция, описывающая зависимость Y от X_1, \dots, X_n , имела, с одной стороны, физическую интерпретацию, с другой – была как можно проще, т. е. при представлении в нелинейном случае в виде какого-то ряда имела как можно меньшее количество членов ряда. Добавление новых членов ряда, теоретически желательное с точки зрения минимизации Δ_Y^2 , на практике может привести к существенному увеличению эмпирического значения Δ_Y^2 по сравнению с его значением при меньшем числе членов ряда из-за характера нелинейности удлиненного ряда за пределами диапазона значений x_1, \dots, x_n , по которым находились оценки параметров ряда.

В-седьмых, как уже упоминалось, множественный коэффициент корреляции изменяется в пределах от 0 до 1. Он инвариантен к изменению начала отсчетов и масштабов переменных и не может быть меньше любого парного или частного коэффициента корреляции. Добавление новых признаков в регрессионную зависимость не может уменьшить множественный коэффициент корреляции. Этот факт можно использовать для проверки корректности расчетов коэффициентов корреляции. В связи с этим по значениям парных, частных и множественных корреляций Y с X_1, \dots, X_n можно принимать решение о включении или невключении в регрессионную зависимость Y от X_1, \dots, X_n тех или иных признаков X_1, \dots, X_n . Нежелательно при этом ориентироваться только на какой-то один тип КК. Лучше использовать вектор-характеристики из разных коэффициентов связи. Так, пусть КК ρ_{YX_1} – мал, $\rho_{X_1X_2}$ – велик, ρ_{YX_2} – мал, но $\rho_{YX_1 \cdot X_2}$ имеет большое значение. Тогда МКК $r_{Y \cdot (X_1 X_2)} \geq \rho_{YX_1 \cdot X_2}$, т. е. МКК может быть большим при малых парных корреляциях и отбрасывать X_1, X_2 в выражении для Y из-за малости их парных корреляций с Y неразумно.

Наконец, еще раз отметим, что все характеристики связи отражают лишь факт и степень связи, но не первопричину, даже если они вводились для этого. Причина может быть найдена лишь на основе содержательного, логического, прикладного анализа задачи, а не формально, когда она вряд ли статистически может быть выяснена.

4.3.6. Представление и анализ результатов стохастических экспериментов, описываемых случайными функциями (процессами, последовательностями)

4.3.6.1. Модели случайных функций

Как следует из изложенного ранее, случайные функции есть разновидность случайных элементов. Их различие в том, что каждому элементарному исходу опыта здесь ставится не число (значение случайной величины) или вектор (совокупность значений случайного вектора), а функция. Ранее мы условились, что будем рассматривать только случайные функции непрерывного временного аргумента t (случайные процессы) или дискретного (t_1, t_2, \dots, t_N либо $t_i = i\Delta t$, $i = \overline{0, \dots, N-1}$, случайные последовательности или временные ряды). Для конкретики далее в основном будем рассматривать процессы.

Отсюда следует, что для случайных процессов возможны два типа характеристик: *ансамблевые*, получаемые усреднением по ансамблю, т. е. по вероятностной мере, по всем траекториям (реализациям) случайного процесса в фиксированные моменты времени t , например, t_1, t_2, \dots, t_n для n -мерных характеристик, или *траекторные*, полученные усреднением по одной отдельной траектории (реализации, выборочной функции) (см рис. 3.4 и (3.1), а также прил. 1 [5, 20]).

Рассмотрим только ансамблевые характеристики. При этом отсчеты случайного процесса в фиксированные моменты времени t_1, t_2, \dots представляются как случайные величины и исследование случайных процессов $X(t)$ и $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_n(t))$ сводится к исследованию случайных векторов ${}_n\mathbf{X} = (X(t_1), X(t_2), \dots, X(t_n))$ либо ${}_n\mathbf{X} = (X_1(t_1), X_2(t_2), \dots, X_n(t_n))$, $X_i = X(t_i)$ или $X_i = X_i(t_i)$ ¹.

Основные свойства законов распределения, характеристических и кумулянтных функций, моментов и кумулянтов при этом сохраняются. Правда, добавляются некоторые новые, связанные с появлением аргу-

¹ Такой подход – следствие описания (задания) случайных функций с помощью семейства конечномерных распределений. С другими способами задания СФ (с помощью основного и выборочного вероятностных пространств, канонических и неканонических моделей, дифференциальных и конечно-разностных уравнений и т. п.) можно ознакомиться, например, в [5, 20].

мента t . В частности, свойство **стационарности** – независимости характеристик от начала отсчета времени t_1 , а зависимость всех характеристик от разности времен $t_2 - t_1 = \tau_1$, $t_3 - t_1 = \tau_2, \dots, t_n - t_1 = \tau_{n-1}$.

Чаще всего **стационарным** называют процесс, для которого функция распределения ${}_2F_X(x_1, x_2; t_1, t_2)$ отсчетов $X_1 = X(t_1)$ и $X_2 = X(t_2)$ удовлетворяет условию

$${}_2F_X(x_1, x_2; t_1, t_2) = {}_2F_X(x_1, x_2; t_2 - t_1). \quad (4.93)$$

Аналогично, если для двух процессов $X_i(t)$ и $X_j(t)$, $i \neq j$, имеет место соотношение

$${}_2F_{\mathbf{X}}(x_1, x_2; t, t + \tau) = {}_2F_{\mathbf{X}}(x_1, x_2; \tau), \quad (4.94)$$

где x_1 – значение $X_i(t)$, а x_2 – значение $X_j(t + \tau)$, т. е. ${}_2F_{\mathbf{X}}(\cdot)$ не зависит от начала отсчета времени t , то процессы $X_i(t)$ и $X_j(t)$ называются **стационарно связанными**.

Внимание! Самостоятельно графически проиллюстрируйте условия стационарности (4.93), (4.94).

Если же выполняются более жесткие условия, аналогичные (4.93), (4.94), но для k -мерного ($k \geq 2$) распределения, то такие процессы называются **строго стационарными** или **строго стационарно связанными**. В последнем случае условие стационарной связанности для $X(t)$ и $Y(t)$ выглядит в виде

$$\begin{aligned} & {}_kF_{\mathbf{X}}(x_1, x_2, \dots, x_l, y_{l+1}, \dots, y_k; t_1, t_1 + \tau_1, \dots, t_1 + \tau_{k-1}) = \\ & = {}_kF_{\mathbf{X}}(x_1, \dots, x_l, y_{l+1}, \dots, y_k; \tau_1, \dots, \tau_{k-1}), \quad 1 < l \leq k. \end{aligned} \quad (4.95)$$

Из нестационарных случайных процессов чаще всего рассматривают такие, которые получаются из стационарных с помощью следующих добавок.

Тренд – аддитивная $a(t)$ или мультипликативная $\lambda(t)$ функция, описывающая устойчивое систематическое изменение среднего значения или масштаба $X(t)$ от времени в течение длительного периода. Это понятие относительное, определяемое целью исследования и представляющее собой сглаженную тенденцию изменения $X(t)$ на

большом отрезке времени. При аддитивном тренде часто говорят о регрессии $X(t)$ по t .

Сезонность – периодически повторяющаяся аддитивная или мультипликативная компонента $s(t)$ функции $X(t)$, период колебания которой известен и обычно считается постоянным, определяемым внешними причинами.

Цикличность (колебания, циклы) – несистематическая колебательная аддитивная или мультипликативная компонента функции $X(t)$, не имеющая регулярной (сезонной) периодичности и допускающая, что период колебаний может меняться циклами, нерегулярным образом.

Например, рассмотрим апостериорное описание нестационарных процессов $X(t)$ с аддитивным $a(t)$ и мультипликативным $\lambda(t)$ трендами, аддитивными сезонностью $s(t)$ и цикличностью $c(t)$, т. е. про-

цессов вида $X(t) = a(t) + s(t) + c(t) + \overset{\circ}{\lambda}(t) X(t)$, по реализациям $Y(t) = X(t) + E(t)$, где $E(t)$ – шумовая компонента. Для них вначале по реализациям $y(t)$ процесса $Y(t)$ находят описание $\hat{a}(t)$ тренда $a(t)$, затем по $y_1(t) = y(t) - \hat{a}(t)$ – описание $\hat{s}(t)$, по $y_2(t) = y(t) - \hat{a}(t) - \hat{s}(t)$ находят представление $\hat{c}(t)$, после чего $\hat{\lambda}(t)$ и $\overset{\circ}{X}(t)$, в частности, путем логарифмирования $y_3(t) = y_2(t) - \hat{c}(t)$, а затем уже представление

остатков $E(t)$ через $\hat{e}(t) = y(t) - \hat{a}(t) - \hat{s}(t) - \hat{c}(t) - \hat{\lambda}(t) \hat{x}(t)$. При этом отличие модели $E(t)$ от ожидаемой (заданной), в частности наличие в $\hat{e}(t)$ какой-либо систематичности или коррелированности отсчетов означает, что исходная для идентификации модель $Y(t)$ не адекватна имеющимся данным $y(t)$. Поэтому ее надо скорректировать, в том числе корректировкой моделей компонент $X(t)$, включая модель $\overset{\circ}{X}(t)$, добавлением авторегрессионных компонент в $X(t)$ или в остатки $E(t)$.

4.3.6.2. Характеристики связи: корреляционный и регрессионный анализ случайных функций

Так же, как для случайных векторов, корреляционный (КА) и регрессионный (РА) анализ случайных функций состоит из трех этапов. Не останавливаясь на них подробно, рассмотрим различие КА и РА случайных функций от их аналогов для случайных векторов, которое сводится к различию характеристик, используемых в анализе.

По аналогии с функциями регрессии, скедастическими функциями, корреляционными, конкорреляционными моментами и коэффициентами, корреляционными отношениями можно ввести их аналоги для случайных функций. Они широко и хорошо описаны. Поэтому мы здесь ограничимся лишь небольшими комментариями, полагая, что СФ являются стационарными.

1. Следует заметить, что для случайных функций собственные (авто) функции регрессии не могут быть произвольными. На них накладываются определенные ограничения, описанные в [5].

2. В зависимости от функции регрессии случайные функции (СФ) можно разбить на два класса: линейные (точнее, линейно-коррелированные, или линейно-связанные (с линейной регрессией, ЛР-функции) [5]) и нелинейные (с нелинейной регрессией). Хотя класс ЛР-функций относительно широк, класс нелинейных СФ, т. е. функций с нелинейной регрессией, гораздо шире. А, как следует из корреляционной теории случайных векторов, корреляционное описание (и, как следствие, спектральное описание) СФ пригодно только для ЛР-функций. Можно убедиться, что для многих нелинейных СФ диапазон значений нормированных корреляционных функций (НКФ) – аналогов коэффициентов корреляции – очень мал (см. примеры в [20]), что делает их и получаемые по ним спектральные характеристики малопригодными для нелинейных СФ. В этом случае надо переходить к конкорреляционным функциям (ККФ), полуконкорреляционным (ППКФ) и к дисперсионным функциям, аналогам корреляционных отношений (см. [5, 20]).

3. Основные свойства корреляционных функций приведены в прил. 2, табл. П2.5.

Указанные в табл. П2.5 свойства КФ позволяют описать некоторые приложения КФ. Помимо определения степени и тесноты линейной статистической и функциональной связи отсчетов $X(t)$ и $X(t + \tau)$ либо $X(t)$, $Y(t + \tau)$ применения КФ позволяют решать следующие задачи.

Если

$$Y(t) = s(t) + X(t), \quad (4.96)$$

где $X(t)$ – случайный процесс, а $s(t)$ – периодическая функция, то, согласно свойству 14 в табл. П2.5,

$$R_{YY}(\tau) = R_{ss}(\tau) + R_{XX}(\tau). \quad (4.97)$$

Если $R_{XX}(\tau) \approx 0$ при $|\tau| \geq \tau_k$, то из (4.97) и свойства 5 периодичности КФ следует, что при $|\tau| > \tau_k$ $R_{YY}(\tau) \approx R_{ss}(\tau)$. Это дает основание выделять периодическую компоненту в (4.96) в смеси его с шумом $X(t)$ и более точно находить период $s(t)$.

Свойство 8 позволяет использовать взаимные КФ для определения задержки в системе, с помощью корреляционно-экстремальных систем определять скорость движения объектов (например, с помощью датчиков, отстоящих на расстоянии L друг от друга, по аргументу τ_{\max} максимума КФ $R_{XY}(\tau)$ сигналов $x(t)$ и $y(t)$ на выходе датчиков определять скорость протяжки пряжи, нити, стали, бревна, ... как L/τ_{\max} , или локализовать источник шума, поворачивая голову либо микрофоны так, чтобы разница во временном сдвиге поступления сигналов в левое и правое ухо (микрофон) была нулевой, т. е. чтобы $R_{XY}(\tau)$ достигала максимума), определять неидентичность каналов в многоканальных системах (по свойствам 8 и 10) и т. п.

Свойство 12 позволяет использовать $R_{XY}(\tau)$ для измерения импульсных переходных (весовых) функций $h(t)$ линейных динамических систем. Согласно свойству 12, если входной сигнал такой системы δ -коррелирован (белый шум), то $R_{XY}(\tau) = h(\tau)$. Заметим, что эта задача нахождения $h(\tau)$ по $R_{XY}(\tau)$ из свойства 12 некорректная по Адамару. Взаимная КФ может оказаться предпочтительнее перед КФ, если надо выделять периодические компоненты из аддитивной смеси периодический сигнал–шум, если найти $R_{XY}(\tau)$, а в качестве $y(t)$ выбирать периодический сигнал с переменным периодом или усредненный исходный сигнал.

Аналогичное справедливо для конкорреляционных, полуконкорреляционных и дисперсионных функций, которые можно использовать и для нелинейных СФ и систем.

Их специфические свойства представлены в прил. 2, табл. П2.6.

4. По аналогии с ЧКК можно ввести разные варианты **частных корреляционных функций** (ЧКФ). Действительно, рассматривая различные отсчеты $X_1 = X(t_1), \dots, X_n = X(t_n)$ скалярной СФ $X(t)$ или отсчеты $X_1 = X_1(t_1), X_2 = X_2(t_2), \dots, X_n = X_n(t_n)$ векторной СФ $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_n(t))$, по аналогии с (4.59), (4.74) можно ввести разные (переставляя местами $X_1(t), X_2(t), \dots, X_n(t)$) сводные и частные корреляционные функции, для которых, разумеется, будут справедливы соотношения вида (4.76)–(4.82). Например, можно ввести частную автокорреляционную функцию ЧАКФ (!)

$$\begin{aligned} & \rho_{1n \cdot (2,3,\dots,n-1)}(t_1, \dots, t_n) = \\ & = \rho_{ZZ}(t_1, t_n; t_2, \dots, t_{n-1}) = -\frac{\|\boldsymbol{\rho}(t)\|}{\|\boldsymbol{\rho}_{11}(t)\| \|\boldsymbol{\rho}_{nn}(t)\|}, \end{aligned} \quad (4.98)$$

где $\|\boldsymbol{\rho}(t)\|$ – определитель корреляционной матрицы $\boldsymbol{\rho}(t) = \{\rho(t_i, t_j), i, j = \overline{1, n}\}$; $\|\boldsymbol{\rho}_{ij}(t)\|$ – алгебраическое дополнение элемента $\rho_{XX}(t_i, t_j)$ в корреляционной матрице $\boldsymbol{\rho}(t)$. Понятно, что $\rho_{1n \cdot (2,3,\dots,n-1)}(t_1, \dots, t_n)$ есть коэффициент корреляции между $Z(t_1) = X(t_1) - \tilde{m}_{X_1}[X(t_2), \dots, X(t_{n-1})]$ и $Z(t_n) = X(t_n) - \tilde{m}_{X_n}[X(t_2), \dots, X(t_{n-1})]$, где $\tilde{m}_X(\cdot)$ СК – прямые регрессии.

5. По аналогии с парными и многомерными (множественными) характеристиками связи случайных величин можно ввести и другие множественные корреляционные характеристики. Их примеры можно найти в [20], а свойства и приложения исследовать самостоятельно.

4.3.6.3. Характеристики связи: спектральный анализ случайных функций

Цель спектрального анализа – исследование частотного состава СФ. Спектральный анализ СФ предполагает выполнение следующих этапов:

- 1) выбор модельного представления СФ и спектрального базиса;

2) нахождение эмпирических спектральных характеристик в выбранном базисе по отдельным реализациям СФ;

3) статистическая проверка гипотез о спектральном составе (в смысле выбранного базиса) СФ по ее реализациям.

Остановимся далее только на втором этапе и то только на разновидностях спектральных характеристик и их свойствах (вопросы оценивания СПМ см. в [5]).

Будем, как и ранее, рассматривать стационарные СФ. Обозначим через $A(\tau)$ временные собственные (авто) для $X(t)$ и взаимные (совместные) для $X(t)$, $Y(t)$ характеристики связи корреляционного типа, а именно КФ $R(\tau)$, $\rho(\tau)$, ККФ $K(\tau)$, $\chi(\tau)$, дисперсионные $Z(\tau)$, $\eta(\tau)$ и $\zeta(\tau)$ [5], полуконкорреляционные $H(\tau)$, частные ЧАКФ $\varphi(\tau)$ и иные и рассмотрим непрерывное и/или дискретное экспоненциальное¹ преобразование Фурье $\mathcal{F}\{\cdot\}$ $B(\omega)$ от них, т. е. (!)

$${}_A B(\omega) = \int_{-\infty}^{\infty} A(\tau) e^{-j\omega\tau} d\tau, \quad \omega = 2\pi f \in (-\infty, \infty); \quad (4.99)$$

$${}_A B(\omega) = \Delta\tau \sum_{i=-\infty}^{\infty} A(i\Delta\tau) e^{-j\omega i\Delta\tau}, \quad \omega = 2\pi f \in (-\infty, \infty); \quad (4.99a)$$

$${}_A B(\omega) = \Delta\tau \sum_{i=-\infty}^{\infty} A(i\Delta\tau) e^{-jiv}, \quad v = \omega\Delta\tau \in (-\pi, \pi), \quad (4.99б)$$

т. е. в общем виде (!)

$${}_A B(\lambda) = \mathcal{F} \left\{ \left[A(\tau) e^{-j\lambda\tau} \right]_{\tau} \right\}, \quad (4.100)$$

где $j = \sqrt{-1}$, $\lambda = \omega$ в непрерывном и $\lambda = v$ в дискретном случае, а оператор $\mathcal{F}\{\cdot\}$ есть интегрирование или суммирование по τ согласно (4.99), (4.99а, б).

¹ Можно рассматривать и другие базисы: Виленкина–Понтрягина, Хаара, Уолша... (см., например, [5, 20]). Экспоненциальный базис, как наиболее часто используемый на практике, выбран для конкретности.

Для конкретности ограничимся вначале, во-первых, ситуацией, когда $A(\tau) = R(\tau)$ или $\rho(\tau)$, т. е. рассмотрим только **спектральные плотности мощности** (СПМ) $S(\omega)$, $s(\omega)$, собственные для $X(t)$, т. е. при $R(\tau) = R_{XX}(\tau)$, $\rho(\tau) = \rho_{XX}(\tau)$, и взаимные при $R(\tau) = R_{XY}(\tau)$, $\rho(\tau) = \rho_{XY}(\tau)$; во-вторых, ограничимся только процессами, т. е. преобразованием (4.99).

Свойства получаемых при этом СПМ представлены в прил. 2, табл. П2.7. Из таблицы хорошо видны возможные приложения СПМ.

Так, помимо определения мощности в полосе частот (свойства 6 и 7), из равенства $S_{XY}(\omega) = 0$ следует линейная статистическая «независимость» $X(t)$ и $Y(t)$; из равенства $S_{XX}(\omega) \approx \text{const}$ следует δ -коррелированность $X(t)$, т. е. равенство $R(\tau) = 0$ при $\tau \neq 0$; из свойства 5 следует, что по пикам графика СПМ можно выделять периодические компоненты $X(t)$, $Y(t)$; из свойств 13 приходим к нахождению частотных характеристик $H(j\omega)$, $G(j\omega)$ через СПМ, причем проще, чем импульсных (весовых) переходных функций $h(\tau)$, $q(\tau)$ через КФ; свойства 8 и 9 можно использовать для определения идентичности каналов; СПМ – для определения звукозащитных свойств (частотных характеристик) различных материалов, их резонансных свойств, а также для проведения измерения гармонического искажения у систем со звуковыми частотами. По отношению $\theta_{XY}(\omega)/\omega$, где $\theta(\omega)$ – фазовый спектр, можно определять временную задержку τ_f на любой частоте $f = \omega/2\pi$, что невозможно при анализе временных сдвигов по КФ.

Следует иметь в виду, что при прохождении сигналов через нелинейные системы имеют место следующие эффекты: появление в экспоненциальном спектре выходного сигнала новых спектральных составляющих, которых не было в СПМ входного сигнала, а именно постоянной составляющей; гармоник основных частот входного сигнала, гармоник комбинационных частот в виде суммы и разности частот входного сигнала. Чем сильнее нелинейность отличается от прямой, тем более сложным может быть экспоненциальный спектр выходного сигнала нелинейных систем. Выходом из данной ситуации может быть переход к другим характеристикам, инвариантным к соответствующим нелинейным преобразованиям сигналов.

Учитывая пригодность СПМ только для описания линейных СФ и систем, необходимо для нелинейных ситуаций ввести аналоги СПМ типа *конкорреляционных* (КПМ) $C(\omega)$, $c(\omega)$ и *дисперсионных* (ДПМ) $\mathfrak{D}(\omega)$, $d(\omega)$, $D(\omega)$ спектральных плотностей «мощности»¹. Они получаются по (4.99), (4.99а, б), (4.100), если $A(\tau)$ есть конкорреляционная функция $K(\tau)$ или $\chi(\tau)$ либо дисперсионная функция $Z(\tau)$, $\eta(\tau)$ и $\zeta(\tau)$ [5, 20]. Их свойства вытекают из свойств ККФ и ДФ (см. табл. П2.6). Некоторые из них, специфичные, приведены в табл. П2.8 прил. 2.

Как видно из свойств табл. П2.8, КПМ и ДПМ пригодны для описания и исследования различных нелинейных систем. В частности, конкорреляционные (ККФ и КПМ) пригодны для нелинейных систем типа Гаммерштейна–Винера, содержащих последовательное соединение безынерционная нелинейность–линейное звено–безынерционная нелинейность, когда выходные характеристики нелинейностей описываются взаимно однозначными монотонными функциями.

По аналогии можно ввести матрицы парных СПМ, *частные* СПМ, *совокупные* и т. д. Это можно сделать самостоятельно. Например, многомерным преобразованием Фурье корреляционного момента

$$\begin{aligned} & \mu_{1,1,\dots,1}(\tau_1, \tau_2, \dots, \tau_{n-1}) = \\ & = R_{X_1 X_2 \dots X_n}(\tau_1, \dots, \tau_{n-1}) = \mathbf{M} \left\{ \overset{\circ}{X}_1(t) \overset{\circ}{X}_2(t + \tau_1) \dots \overset{\circ}{X}_n(t + \tau_{n-1}) \right\} \end{aligned}$$

получаем полиспектральную плотность мощности (!).

В частности, *биспектральная плотность мощности* получается как (!)

$$\begin{aligned} S_X(\omega_1, \omega_2) &= S_{XXX}(\omega_1, \omega_2) = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{XXX}(\tau_1, \tau_2) e^{-j(\omega_1 \tau_1 + \omega_2 \tau_2)} d\tau_1 d\tau_2; \end{aligned} \quad (4.101)$$

¹ Слово *мощность* здесь не совсем корректно, так как эти характеристики отражают распределение не мощности СФ, а ее аналога, либо мощности, но для других СФ: $F_X[X(t)]$, $F_Y[Y(t)]$ (для КПМ) или другой сущности (для ДПМ).

$$S_{XYZ}(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{XYZ}(\tau_1, \tau_2) e^{-j(\omega_1 \tau_1 + \omega_2 \tau_2)} d\tau_1 d\tau_2, \quad (4.102)$$

$$0 \leq \omega_1 \leq \infty, \quad 0 \leq \omega_2 \leq \omega_1.$$

Еще один класс важных спектральных характеристик – это функции когерентности.

Парная функция когерентности $\gamma^2(\omega)$ определяется как

$$\gamma_{XY}^2(\omega) = \frac{|S_{XY}(\omega)|^2}{S_{XX}(\omega)S_{YY}(\omega)}, \quad 0 \leq \gamma_{XY} \leq 1, \quad (4.103)$$

которая, согласно определению, аналогична квадрату нормированной корреляционной функции. Иногда ее называют квадратом когерентности, оставляя название функция когерентности за $\gamma(\omega)$. Заметим, что хотя по физическому смыслу $\gamma(\omega)$ аналогична $\rho(\tau)$, в приложениях она играет более значительную роль.

Следующая интересная характеристика **кепстр**¹ (от спектр). Он вводится соотношением (!)

$$\mathcal{R}(\tau) = \mathcal{F}^{-1} \left\{ \left[e^{j\omega\tau} \ln[S_{XX}(\omega)] \right]_{\omega} \right\}, \quad (4.104)$$

где $\mathcal{F}^{-1}(\cdot)$ – обратное преобразование Фурье, а основание логарифма может быть любым. Как видим, $\mathcal{R}(\tau)$ подобен корреляционной функции, поэтому τ имеет размерность времени и называется **сачтота** (от частота) или **кьюфренси**, а гармоники спектральной характеристики для кепстра называют **рахмоники**.

Являясь нелинейным преобразованием от спектра, кепстр обладает рядом замечательных свойств, выгодно отличающих его от КФ. Например, свойства 12 из прил. 2, табл. П2.5, для КФ (см., например, [19]) и 13 из табл. 2.7 переходят в свойство

$$\mathcal{R}_{YY}(\tau) = \mathcal{R}_{XX}(\tau) + \mathcal{R}_H(\tau),$$

¹ Вторая расшифровка Kestr – Kolmogorov Equation Power Series Time Response.

где $\mathcal{R}_{YY}(\tau)$, $\mathcal{R}_{XX}(\tau)$ – кепстры $Y(t)$ и $X(t)$; $R_H(\tau)$ – кепстр линейной системы, т. е. обратное преобразование Фурье от $\ln|H(j\omega)|^2$. Как видим, кепстр обладает аддитивностью при линейных инерционных преобразованиях. Например, при прохождении $X(t)$ через линейную систему удастся в $\mathcal{R}(\tau)$ получить аддитивный эффект, т. е. разделить сигнал возбуждения и путь его распространения, причем их проявление на разных частотах.

Наконец, обратим внимание еще на один аналог спектрального анализа – вейвлет¹-анализ.

4.3.6.4. Характеристики связи: вейвлет-анализ и динамический спектральный анализ

Вейвлет-анализ появился в конце 1980-х годов. Концептуально вейвлет-анализ родствен анализу Фурье, точнее его взвешенной разновидности. В их основе лежит одно и то же преобразование Фурье. Но в классическом и взвешенном анализе Фурье базовыми функциями являются периодические функции (волны), в частности синусы и косинусы в экспоненциальном преобразовании. Экспоненциальная функция $e^{j\omega t}$ при переменном ω является анализирующей в преобразовании Фурье. Во взвешенном преобразовании добавляется жестко фиксированная весовая функция $\mu(t)$ (см., например, [5]) или (4.99), (4.99а, б), если в нем $A(\tau)$ заменить на $A(\tau)\mu(\tau)$.

В отличие от преобразований Фурье вида (4.99) непрерывное вейвлет-преобразование описывается соотношением² (!)

$$\begin{aligned} [W_{\psi} f](\lambda, \tau) &= W_f(\tau; \lambda) = W_f(\lambda; \tau) = \\ &= \frac{1}{\sqrt{|\lambda|}} \int_{-\infty}^{\infty} f(t) \tilde{\psi}\left(\frac{t-\tau}{\lambda}\right) dt, \end{aligned} \quad (4.105)$$

¹ От *англ.* wavelet – вейвлет, волнушка, короткая волна, всплеск, рябь. Термин введен в 1984 г. А. Гроссманом и Дж. Морле при анализе акустических и сейсмических сигналов.

² В левой части (4.105) приведены обозначения, принятые в вейвлет-анализе, где к тому же чаще вместо λ используется символ a , а вместо τ – символ b . Мы будем использовать единые обозначения, принятые в пособии.

где $\psi(t)$ – *анализирующий (материнский) вейвлет*, или просто *вейвлет*, а сдвинутые (при изменении параметра сдвига τ , часто вместо τ пишут a) или сжатые-растянутые (при изменении параметра масштаба λ , часто вместо λ пишут b) копии вейвлета называются *вейвлетными функциями*, « \sim » – это знак комплексной сопряженности. Растяжение и сдвиг вейвлетных функций и отличает вейвлет-преобразование от преобразования Фурье. Второе отличие – сами вейвлеты являются затухающими или осциллирующе-затухающими функциями. Уменьшая масштабный параметр λ , т. е. выбирая его по правилу $0 \leq \lambda \leq 1$ (или вейвлет в виде узкого окна), мы будем выделять высокие частотные составляющие переходных процессов, имеющихся в сигнале, описываемом $X(t)$. При $|\lambda| \gg 1$, т. е. при очень широких вейвлет-окнах, будем выделять (регистрировать) медленные или длинноволновые колебательные процессы в $X(t)$. Принципиально важно подчеркнуть, что по оси масштабирования λ (частотной оси вейвлетного анализа, поэтому обозначается через λ) преимущественно используется логарифмический масштаб, как это часто делается в акустике и музыке. Это диктует шаг растяжения.

Обычно рассматривают прямое и обратное вейвлет-преобразование, непрерывное и дискретное, в том числе быстрое, а в качестве вейвлетов – вейвлеты Хаара, модулированную гауссову кривую, Добеши–Гроссмана–Мейера, Мейера, Добеши, Бэттла–Лемерье и др. (см., например, [19] и табл. П2.9 в прил. 2).

По локализации во временном и частотном представлении вейвлеты занимают промежуточное положение между гармоническими функциями, локализованными по частоте, и функцией Дирака, локализованной во времени. Основная область применения вейвлетных преобразований – анализ и обработка сигналов и функций, причем нестационарных во времени и неоднородных в пространстве, когда надо иметь не только общую частотную характеристику сигнала, т. е. распределение энергии по частотам, но и сведения о локальных координатах, на которых себя проявляют те или иные группы частотных составляющих или на которых происходят быстрые изменения частотных составляющих сигнала. Это обеспечивается тем, что в отличие от пары экспоненциальных преобразований Фурье в вейвлет-преобразовании можно использовать семейства функций, реализующих различные варианты соотношения неопределенности, т. е. гибко выбирать их и тем самым наиболее эффективно решать поставленные задачи, с гораздо

более высокой точностью представляя локальные особенности сигналов, вплоть до скачков (разрывов I рода). При этом развертка получается двумерной: частота–координата (задержка) как независимые переменные.

В заключение приведем еще основные понятия вейвлет-анализа. Функция (4.105) называется **вейвлет-спектром** функции $f(t)$ (а ее физический аналог – вейвлет-спектром сигнала).

Рассмотрим аналог (4.105) – *динамическое преобразование* Фурье с оконной (весовой) функцией $W(t)$ ¹:

$$F_f(\omega, \tau) = \int_{-\infty}^{\infty} f(t)W(t-\tau)e^{-j\omega t} dt, \quad j = \sqrt{-1}. \quad (4.106)$$

Сравнивая (4.105) и (4.106), видим, что в непрерывном вейвлет-преобразовании параметр масштаба $\lambda \in (-\infty, \infty)$ (классическое обозначение « a ») соответствует $\omega \in (-\infty, \infty)$ в динамическом фурье-преобразовании, а параметр сдвига τ (классическое обозначение b) – времени τ . Между этими параметрами существует простая связь: $a = \lambda = kT = 2k\pi/\omega$, где $k > 0$ – коэффициент пропорциональности, зависящий от выбранной вейвлет-функции $\psi(t)$; T – период, соответствующий ω : $T = 2\pi/\omega$.

Чтобы провести дальнейшую аналогию, заменим в (4.106) частоту ω на период T . Тогда (4.106) будет иметь вид (!)

$$F_f(T, \tau) = \int_{-\infty}^{\infty} f(t)W(t-\tau)e^{-j2\pi t/T} dt. \quad (4.107)$$

Квадрат модуля (4.107), т. е. неотрицательную функцию

$$P_F(T, \tau) = |F_F(T, \tau)|^2, \quad (4.108)$$

принято называть **спектрограммой** $f(t)$.

По аналогии рассмотрим плотность энергии для вейвлет-преобразования – **скэйлограмму**

$$P_W(\lambda; \tau) = |W_f(\lambda, \tau)|^2. \quad (4.109)$$

¹ От *англ.* window – окно.

Картина линий, соединяющих локальные экстремумы (отдельно минимумы и максимумы) поверхности $W_f(\lambda, \tau)$, $P_W(\lambda, \tau)$, $P_f(T, \tau)$, называется *скелетом* (хребтами – ridges) поверхности в пространстве. Многие полагают, что в скелетоне, построенном на основе вейвлет-спектра, содержится вся информация об анализируемой функции (сигнале) $f(t)$.

Функция $E_f(\lambda)$ вида

$$E_{W_f}(\lambda) = \int_{-\infty}^{\infty} |W_f(\lambda, \tau)|^2 d\tau = \int_{-\infty}^{\infty} P_W(\lambda; \tau) d\tau, \quad (4.110)$$

т. е. интеграл от скэйлограммы, называется *дисперсией вейвлет-преобразования*, или *скалограммой*, а интеграл

$$\int_{-\infty}^{\infty} E_W(\lambda) \frac{d\lambda}{\lambda^2} = E \quad (4.111)$$

есть *полная энергия* функции (сигнала) $f(t)$.

Как следует из (4.110) и (4.111), скалограмма представляет собой распределение энергии функции (или сигнала) $f(t)$ по различным масштабам λ . Положение максимума скалограммы можно интерпретировать как среднюю продолжительность некоторого элементарного события, описываемого $f(t)$, вносящего основной вклад в энергию анализируемого сигнала при параметре масштаба, равном λ . Аналогичное справедливо для $E_F(\lambda)$ или

$$E_F(T) = \int_{-\infty}^{\infty} |F_f(T, \tau)|^2 d\tau = \int_{-\infty}^{\infty} P_F(T; \tau) d\tau, \quad (4.112)$$

поскольку интеграл от скалограммы Фурье также дает энергию:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} E_F(\omega) d\omega = E. \quad (4.113)$$

Заметим, что указанная аналогия вейвлет-преобразования с динамическим преобразованием Фурье приводит к плодотворности их совместного использования как дополняющих друг друга. Например,

сравнение вейвлет-спектрограмм модульных сверхширокополосных сигналов (СШПС) с их спектрограммами Фурье подтверждает утверждение о принципиально большей локализованности спектральной плотности СШПС $f(t)$ при использовании вейвлетов.

Еще раз обратим внимание, что при вейвлет-преобразовании реального сигнала, как и при преобразовании Фурье, осуществляется разложение исследуемого сигнала в ряд базисных элементов, помноженных на определенные коэффициенты, т. е. осуществляется **декомпозиция сигнала**. При обратном же преобразовании происходит восстановление сигнала – его **реконструкция**.

При этом после прямого преобразования над спектрами осуществляются необходимые манипуляции для удаления тех компонентов, которые мешают реконструкции искомым особенностей сигнала. Преимуществом обработки в области спектров является то, что свертке во временной области соответствует произведение в спектральной.

Обратим попутно внимание на только что введенное новое понятие **динамического спектра** Фурье (4.106) и (4.107), которое в дополнение к специальным характеристикам нестационарных случайных функций (см., например, [20]) все более широко применяется в последнее время. Из (4.106), (4.107) и изложенного ранее следует, что идея динамического спектра в том, что, чем длиннее анализируемый сигнал, тем больше возможность реального появления новых нестационарностей, и они как следствие проявятся в итоге преобразования Фурье. Особенно это важно, когда короткие участки сигнала его слабо характеризуют, а сами участки можно считать квазистационарными. По каждому из таких квазистационарных участков и можно находить динамический спектр (4.106). Иными словами, если квазистационарные участки имеют протяженность $n\Delta t$, то K таких отрезков будут иметь длину $N\Delta t = Kn\Delta t$, а ЭХ-оценки спектра (4.108) можно определить как

$$\hat{P}_F(\omega; \tau) = \hat{P}_f(\omega; k) = \frac{1}{n} \left| \sum_{i=0}^{n-1} w(i) f[(i+(k-1)n)\Delta t] \exp\left\{ \frac{j2\pi kn}{n} \right\} \right|^2, \quad (4.114)$$

$$k = 1, 2, \dots, K, \quad i = 0, 1, \dots, N-1.$$

Тем самым динамический спектр нестационарной СФ не только позволяет выявить периодики по ω , присущие анализируемой СФ, но и дает общую картину изменения частотной структуры СФ (развертка по k). Заметим, что применение динамического спектра и вейвлет-анализа может обойтись без предположения о стационарности СФ, использования усреднения для уменьшения статистических погрешностей оценок спектров (см. [5]) и помогает обнаруживать те закономерности, которые таким усреднением затушевываются. Например, получать индивидуальные спектры электроэнцефалограмм человека, отражающие чередование низко- и высокоамплитудных колебаний в разных отделах полушарий мозга. Динамический спектр (и вейвлет-анализ) позволяют не только выявлять периодики, присущие анализируемому сигналу, но и устанавливать время смены одной периодики на другую (т. е. отслеживать частотную разладку сигнала!), а также длительность (время) «присутствия» той или иной периодики в анализируемом сигнале, выделять участки стационарности процесса и время возникновения нестационарности, отличать единичный всплеск от регулярного колебания.

Итак, завершая этот раздел спектрального, вейвлет-анализа и динамического спектрального анализа, отметим следующее.

1. Обычный спектральный фурье-анализ реальных сигналов имеет ряд слабых сторон:

- особенности сигналов, связанные с разрывами (скачками) и острыми пиками, в частотной области «размываются» по всей оси частот, что делает их трудно обнаруживаемыми, в частности из-за принципа неопределенности Гейзенберга, согласно которому нельзя одновременно измерить частоту и время со сколь угодно высокой точностью, т. е. чем лучше временное разрешение, тем хуже частотное, и наоборот;
- гармонические периодические базисные функции принципиально не способны отображать перепады сигналов с бесконечной крутизной (явления размывания и эффекта Гиббса);
- он отображает глобальные (протяженные) сведения о частотном составе и дает слабое представление о локальных свойствах сигналов при быстрых временных изменениях состава сигнала;
- не позволяет анализировать частотные характеристики сигнала в произвольные моменты времени;

- наблюдается неразделимость интергармоник (кратных гармоник), когда неверно обнаруживаются несуществующие интергармоники.

2. В отличие от обычного спектрального анализа вейвлет-анализ позволяет сконцентрировать внимание на тех особенностях сигналов, которые обычный СА не выявляет, причем нестационарных во времени и неоднородных в пространстве, когда результаты анализа должны содержать не только общую частотную характеристику сигнала (распределение энергии сигнала по его частотным составляющим), но и локальные сведения о координатах, на которых проявляются соответствующие частотные составляющие, например, происходят изменения разрывов первого рода (конечные скачки).

Двумерный вейвлет-образ обычно графически представляют так, что параметр масштаба λ откладывают по оси абсцисс, а параметр локализации τ – по оси ординат (координатной, независимой от переменной t сигнала).

Несомненным достоинством вейвлет-анализа перед динамическим (скользящим) оконным спектральным анализом является то, что при изменении масштаба времени вейвлет-преобразования сохраняется постоянная разрешающая способность, используется прежний объем данных, а мелкие детали поведения сигнала автоматически становятся несущественными. Наконец, еще раз подчеркнем, что вейвлет-анализ пригоден для обработки сигналов любой сложности.

Отметим, что форма вейвлета, его четность или нечетность, доминирующая частота и степень ее локализации существенно влияют на вейвлет-спектры анализируемых сигналов и на возможность его локальных особенностей. Поэтому использование разнотипных вейвлетов позволяет повысить достоверность выделения его локальных свойств: региональной функции тренда – средних значений по большим интервалам усреднения циклических компонент с определенным периодом повторения, как правило, гладких по форме (крупномасштабный анализ), и локальных особенностей (аномалий) разного вида – резких изменений в определенные моменты.

3. Динамический (скользящий) спектральный анализ подобен вейвлет-анализу и позволяет определять моменты частотной разладки сигнала, а также устанавливать периодические закономерности в этой «разладке».

Однако динамический спектр не различает локальные и глобальные свойства сигнала, затруднителен в выделении или исключении некоторых характерных свойств нестационарных сигналов.

4. Допустим, необходимо описать физический объект, работающий в течение интервала времени $\Gamma = [t_1, t_2]$, $t_2 - t_1 = T < \infty$, на основе модели стационарного случайного процесса $X(t)$. Возможны два принципиально отличных подхода к такому описанию. Первый связан с использованием стационарной модели $X(t) = X_\infty(t)$, для которой по определению $t \in (-\infty, \infty)$. Следовательно, данный подход основан на предпосылке, что если бы объект существовал при всех $t \in \Gamma = (-\infty, \infty)$, то он сохранил бы свойство стационарности и значит, мы вправе выбрать такое большое $T \rightarrow \infty$, при котором краевые эффекты не сказываются, т. е. перейти к пределам $t_1 \rightarrow -\infty$, $t_2 \rightarrow \infty$. Хотя об этом явно нигде не говорится, фактически почти во всех исследованиях по статистическим измерениям, а в большинстве случаев и в исследованиях по оцениванию, по статистике случайных процессов постулируется именно этот подход.

Второй подход основан на использовании финитных моделей $X_\Gamma(t)$, $\Gamma = [t_1, t_2]$, $|t_1|, |t_2| < \infty$. В дальнейшем в подобных ситуациях для определенности зачастую будем полагать $\Gamma = [0, T]$, где t – непрерывный (для процессов) или дискретный (для последовательностей) аргумент. Использование подобного описания не столь очевидно как предыдущего и может осуществляться в следующих вариантах.

Вариант 1. Использование «стационарных» финитных моделей $X_\Gamma(t)$, заданных на конечных интервалах, когда значение $t \notin \Gamma$ не имеет смысла. Стационарность в этом случае понимается как независимость рассматриваемых ансамблевых характеристик $Q(\lambda, t, \tau)$ от t для всех $t, t + \tau_1, t + \tau_2, \dots, t + \tau_{n-1} \in \Gamma$; $\tau = (\tau_1, \dots, \tau_{n-1})$. Трудность такого описания связана со сложностью учета краевых эффектов ансамблевых характеристик при $t + \tau \leq t_1$ и $t + \tau \geq t_2$.

Вариант 2. Учет нестационарности модели путем замены $x_\Gamma(t)$ на нестационарный процесс

$$X_n(t) = \begin{cases} X_\Gamma(t) & \text{при } t \in \Gamma; \\ 0 & \text{при } t \notin \Gamma. \end{cases}$$

Понятно, что ансамблевые характеристики такого процесса подчиняются условиям $Q(\lambda, t; \tau) = a$, где a – константа, если хотя бы одно из значений $t, t + \tau_1, t + \tau_2, \dots, t + \tau_{n-1}$ не принадлежит Γ .

Вариант 3. Представление $X_\Gamma(t)$ в виде произведения $X_\infty(t) \times 1(t; t_1, t_2)$, где $1(t; t_1, t_2)$ – финитная единичная весовая функция, равная единице при $t \in (t_1, t_2)$ и нулю при $t < t_1$ и $t > t_2$. Тогда согласно (3.20) спектральное представление $X_\Gamma(t)$ будет представлять собой свертку спектров $X_\infty(t)$ и частотной весовой функции $1(j\omega)$, соответствующей $1(t; t_1, t_2)$. Например, $S_{X_\Gamma}(\omega) = S_{X_\infty}(\omega) \otimes |1(j\omega)|^2$, где \otimes – символ операции свертки.

Вариант 4. Замена процесса на стационарный $X_{[T]}(t)$ полученный периодизацией $X_\Gamma(t)$.

$X_{[T]}(t) = X[t(\text{mod } T)] = X_\Gamma(t \pm kT)$, $k = 0, 1, 2, \dots$, где $t(\text{mod } T)$ – остаток от деления t на T .

Временные ансамблевые характеристики такого процесса будут периодическими по $t_1, \tau_i, i = 1, (n-1)$, с периодом T , т.е. для них выполняется соотношение

$$Q(\lambda, t, \tau) = Q(\lambda, t \pm kT; \tau \pm nT), \quad k, n = 0, 1, 2, \dots$$

Вариант 5. Замена процесса $X_\Gamma(t)$ на «эквивалентный» ему стационарный процесс $X_\Omega(t)$, ансамблевые характеристики которого $Q_{X_\Omega}(\lambda, \tau)$ получаются усреднением по траектории (по t) ансамблевых характеристик $Q_{X_\Gamma}(\lambda, t, \tau)$ на $(0, T)$:

$$Q_\Omega(\lambda, \tau) = \mu_T \left\{ Q_{X_\Gamma}(\lambda, t, \tau) \right\} = Q_{X_\Gamma}(\lambda, \tau),$$

где $\mu_T \{ \}$ – оператор усреднения по траектории на отрезке $[0, T]$.

Сходство и различие итоговых характеристик Q , получаемых при разных рассмотренных вариантах описания реальных сигналов стационарными моделями, рекомендуем исследовать самостоятельно, в том числе на практических занятиях.

Из изложенного ясно, что при описании реальных объектов вероятностными моделями в статистических измерениях, имитации и иных операциях необходимо учитывать способы задания вероятностных моделей, вид исследуемых характеристик (ансамблевые, траекторные или совокупные) и используемый подход к описанию объекта моделью. В частности, в большинстве практических измерений, применяющих усреднение по траектории, в явной или неявной форме присутствует замена измеряемых характеристик q на $q_{\mathcal{E}}$, как в варианте 5. Это результат того, что математическая стационарность, строго говоря, не может иметь место в реальных физических системах. Следовательно, какой бы из описанных выше вариантов мы ни взяли, обязательно используем его вместе с вариантом 5.

5. Как указано в подпараграфе 4.3.1, из вероятностного аппарата мы рассмотрели лишь раздел В1 – вероятностные модели. Разделы В2–В5 теории вероятностей рассматриваются в отдельной дисциплине. Для удобства восприятия материала пособия в прил. 3 приведены некоторые важные сведения по разделу В4, часть из которого упоминается по тексту пособия.

4.3.7. Апостериорный статистический аппарат (аппарат математической¹ статистики)

Согласно п. 4.3.1 статистический аппарат включает в себя три раздела, связанных с тремя типами статистических операций: оценивания характеристик (С1), проверки гипотез (С2) и формирования выводов (С3).

Рассмотрим их настолько кратко, насколько это удастся без потери понимания.

С1. Элементы теории оценивания

Прежде всего напомним некоторые введенные ранее термины (см., например, табл. 3.4).

¹ Поскольку прикладная статистика есть раздел математической статистики, далее всюду будем использовать только термин *математическая статистика*.

Под вероятностными понимают ансамблевые характеристики $Q(\mathfrak{Q})$ случайных элементов \aleph (величин, векторов и функций), т. е. характеристики вида

$$Q_I(\mathfrak{Q}) = \mathbf{M} \{ g_Q(\aleph; \mathfrak{Q}) \}; \quad (4.115)$$

$$Q_{II}(\mathfrak{Q}) = \mathbf{M} \{ g_Q(\aleph; Q_1, Q_2, \dots, Q_s; \mathfrak{Q}) \}, \quad (4.116)$$

или

$$Q_{III}(\mathfrak{Q}) = A \{ Q_1, Q_2, \dots, Q_s; \mathfrak{Q} \}, \quad (4.117)$$

где $Q(\mathfrak{Q})$ – некоторая функциональная или числовая характеристика случайного элемента \aleph ; \mathfrak{Q} – пустой (для числовых характеристик), скалярный или векторный аргумент характеристики $Q(\cdot)$; $\mathbf{M}\{\cdot\}$ – оператор усреднения по вероятностной мере случайного элемента \aleph (оператор математического ожидания); $g_Q(\cdot)$ – некоторая функция, определяемая видом характеристики Q ; $Q_1(\mathfrak{Q}), Q_2(\mathfrak{Q}), \dots, Q_s(\mathfrak{Q})$ – некоторые другие вероятностные характеристики, входящие по определению в $Q(\mathfrak{Q})$; $A\{\cdot\}$ – некоторая функция (явная или неявная), функционал или оператор (отличающийся от оператора $\mathbf{M}\{\cdot\}$ усреднения по вероятностной мере).

Примеры вероятностных характеристик (4.115)–(4.117) представлены ранее в табл. 4.4 и в прил. 3.

Генеральная совокупность и выборка из нее

Базовыми понятиями математической статистики являются генеральная совокупность, выборка из генеральной совокупности, статистическая совокупность, статистика, выборочная (статистическая) характеристика, эмпирическая характеристика и оценка.

Под статистической совокупностью (понятие прикладной статистики) понимают:

- а) первичный материал, подлежащий обработке;
- б) множество значений однородных признаков (показателей, характеристик) некоторого реального объекта, полученных в результате

эксперимента с объектом (или наблюдения за ним), обладающих качественной общностью.

Под генеральной совокупностью (ГС) (понятие математической статистики) понимают множество всех возможных «значений» качественного или количественного признака, характеризующего некоторый исследуемый объект, которые могут появиться при эксперименте с объектом. Следует различать теоретическую и практическую генеральную совокупности.

Теоретическая генеральная совокупность – объект теоретической (математической) статистики, допускающая два определения. Первое – это множество (абстрактное, мысленное) всех возможных реализаций (значений) случайного элемента в предположении случайного отбора их без возвращения в гипотетическом эксперименте. Второе – это множество всех возможных значений (реализаций) случайного элемента, которые могут быть получены при «разыгрывании», «опытном получении их» в процессе экспериментирования со случайным элементом, в стохастическом эксперименте.

Практическая генеральная совокупность (ПГС) является объектом прикладной статистики. Под ней обычно понимают следующее:

- любой набор рассматриваемых объектов как одушевленных, так и неодушевленных (точнее говорить не о самих объектах, а о значениях характеризующих их показателей);
- совокупность, из которой производится выборка, в частности вся статистическая совокупность;
- вся подлежащая изучению совокупность.

Генеральная совокупность может быть конечной или бесконечной, реальной или гипотетической. Количество элементов конечной генеральной совокупности N_T называется ее *объемом*. *Реальная* ГС – это та, что получается в процессе физических (реальных) экспериментов, *гипотетическая* – получаемая в мысленных (абстрактных, теоретических) экспериментах. Очевидно, что теоретическая ГС является гипотетической и бесконечной, если используются непрерывные математические модели случайных элементов. Заметим, что *всякая замена наблюдаемых* (эмпирических, выборочных, статистических (см. далее)) *характеристик их непрерывными моделями*, т. е. вероятностными характеристиками, *автоматически означает замену реальной конечной ГС на гипотетическую бесконечную*.

Гипотетическая бесконечная ГС может встречаться в двух случаях, которые будут описаны ниже.

Выборочной совокупностью, выборкой из генеральной совокупности или просто *выборкой*, называют совокупность случайно отобранных объектов ГС, т. е. ту часть ГС, которая попала (взята) на проверку, исследование и т.п. Число элементов выборки называется ее объемом и обозначается через N .

Для того чтобы выборка могла более или менее удачно характеризовать всю ГС, она должна удовлетворять некоторым требованиям (условиям, свойствам).

Приведем некоторые важнейшие из них.

1. **Репрезентативность** (представительность) – выборка должна, достаточно хорошо представлять ГС и все ее пропорции. Это требование касается следующих двух требований.

2. **Однородность**. Если ГС состоит из разнородных элементов, то выборка должна содержать элементы из каждой однородной подсовкупности ГС. В приложении к гипотетической теоретической ГС это означает, например, что в выборку включаются реализации (значения) χ_1, \dots, χ_N только одного и того же случайного элемента \aleph .

3. **Случайность** (рандомизованность) означает, что каждый элемент ГС (или каждый элемент выборки) имеет один и тот же шанс, одну и ту же вероятность попасть в выборку. Рандомизация (ослучаивание) выборки – один из способов обеспечения однородности выборки (вспомните барабан спортлото).

Если ГС является бесконечной, то случайность выборки обеспечивается выбором наудачу (случайно) любого ее элемента. Например, рассмотрим бросание игральной кости. В этом примере опыт (эксперимент), связанный с k -кратным бросанием кости, можно рассматривать тройко. Во-первых, как случайную выборку, т. е. как процесс выбора существующих элементов – конкретных комбинаций, состоящих из k цифр 1, 2, 3, 4, 5, 6 с возможными их повторениями из несуществующей (чисто гипотетической) бесконечной ГС, включающей в себя все мыслимые k -ричные комбинации выпаданий кости при сколь угодно большом (в пределах счетном) повторении извлечений. Во-вторых, как случайное извлечение цифр 1, 2, 3, 4, 5, 6 k раз из бесконечной гипотетической ГС, содержащей бесконечное множество цифр 1, 2, ..., 6. В-третьих, как случайное извлечение k раз из конечной совокупности шести цифр 1, ..., 6 с возвращением обратно извлеченной цифры и тщательным перемешиванием всех цифр после каждого

извлечения. С точки зрения получаемых при этом выборочных характеристик все варианты экспериментов (и, следовательно, разные варианты ГС) эквивалентны и выбор какого-то одного из вариантов для интерпретации результатов обуславливается лишь видом практической задачи, пристрастиями и интересами исследователя.

Прежде чем рассматривать понятие независимости выборки, отметим двойственный характер ее. Пусть ${}_N\chi = (\chi_1, \dots, \chi_N)$ – выборка объема N некоторого скалярного или векторного случайного элемента \aleph (теоретическая выборка) или показателя объекта (практическая выборка). Апостериори, после опыта (т. е. когда выборка уже получена) элементы χ_1, \dots, χ_N не являются случайными, так как их значения известны, объективно существуют и не зависят от воспринимающего их субъекта. Однако априори, до опыта, они случайны – ведь согласно третьему требованию в качестве значения χ_i , $i = \overline{1, N}$, с равной вероятностью может быть выбран любой элемент ГС. Для теоретической ГС это равносильно тому, что в качестве значения χ_i может выступать любое значение элемента \aleph . Именно это-то и обеспечивает случайность всего набора $\chi = (\chi_1, \dots, \chi_N)$ и невозможность предугадать, каков он будет в каждом конкретном эксперименте с N выборками из ГС. Чтобы отразить этот факт, элементы выборки априори, а также при нахождении по ней генеральных (в частности вероятностных) характеристик самой выборки, обозначают большими буквами, т. е. вместо (χ_1, \dots, χ_N) пишут $\aleph = (\aleph_1, \dots, \aleph_N)$. Набор $\aleph = (\aleph_1, \dots, \aleph_N)$ представляет собой N -мерный векторный элемент ${}_N\aleph$. Он может быть описан генеральными (вероятностными) характеристиками N -мерного векторного элемента ${}_N\aleph$ и отражает свойства выборки, в частности ее однородность, независимость, способ получения и т. д. Так, если выборка однородна и взята из ГС случайной величины X , то любой априорный элемент X_i , $i = \overline{1, N}$, имеет то же распределение, т. е. те же характеристики, что и величина X . Это условие можно отразить соответствием $X_i \sim X$ или $Q_{X_i}(\mathfrak{G}) = Q_i(\mathfrak{G}) = Q_X(\mathfrak{G}) = Q(\mathfrak{G})$. В частности, выборка X_1, \dots, X_N ГС случайной величины X будет иметь N -мерный совместный закон распределения ${}_NF(x_1, \dots, x_N)$ (вероятностный или для

конечной генеральной совокупности – генеральный), отражающий тот факт, что при отборе первого элемента выборки вероятность того, что соответствующее ему значение x_i не превзойдет x_1, \dots, N -го элемента x_N не превзойдет x_N , равно ${}_N F(x_1, \dots, x_N)$. Если ${}_N F(x_1, \dots, x_N) = F_1(x_1) \dots F_N(x_N)$, где $F_i(x_i)$ есть одномерное (маргинальное) распределение i -го элемента X_i выборки χ ($F_i(x_i) = F(x_i) = F_X(x_i)$ для однородной выборки), т. е. N -мерное распределение выборки равно произведению одномерных распределений ее элементов, то выборка называется *независимой* (простой), а способ выбора, которым получена такая выборка, *простым*. Таким образом, простой случайный выбор (отбор) из ГС случайного элемента \aleph обеспечивает получение однородной независимой выборки $\chi = (\chi_1, \dots, \chi_N)$.

Выборочные характеристики и оценки вероятностных характеристик

Любая функция $\psi(\chi) = \psi(\chi_1, \dots, \chi_N)$ от элементов выборки χ_i (от выборки χ) называется *статистикой*.

Важнейший класс статистик образуют *эмпирические характеристики*, аналогичные вероятностным характеристикам (ВХ) (4.115)–(4.117), но отличающиеся от них тем, что оператор усреднения по вероятностной мере $\mathbf{M}\{\cdot\}$ (оператор математического ожидания) в (4.115) и (4.116) заменяется оператором $\mu_\Gamma\{\cdot\}$ усреднения по выборке χ объема N (оператором выборочного среднего: ансамблевого среднего арифметического μ_N или траекторного среднего μ_T). Например, ансамблевого среднего

$$\mu_N\{f(\mathbf{x})\} = \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i) - \quad (4.118)$$

оператор ансамблевого усреднения ($\mathbf{x}_i = x_i$ для случайных величин, $\mathbf{x}_i = (x_{1,i}, \dots, x_{l,i})$ для l -мерного случайного вектора, $\mathbf{x}_i = (x_i(t_1), \dots, x_i(t_l))$ для случайных функций $X(t)$ и $\mathbf{x}_i = (x_{1,i}(t_1), \dots, x_{l,i}(t_l))$ для случайных

функций $X(t) = (X_1(t), \dots, X_I(t))$, а оператор траекторного среднего (4.119) вводится как (3.1), т. е.

$$\begin{aligned} \mu_{\Gamma} \{f[\mathbf{x}(t)]\} = \\ = \begin{cases} \mu_T \{f[\mathbf{x}(t)]\}, & \text{если } t \text{ – непрерывно;} \\ \mu_N \{f[\mathbf{x}(t)]\} = \mu_N \{f[\mathbf{x}(i\Delta t)]\}, & \text{если } t \text{ – дискретно.} \end{cases} \end{aligned} \quad (4.119)$$

Эмпирические характеристики будем обозначать рукописными буквами, т. е. через $\mathcal{Q}(\mathfrak{D})$, в отличие от обозначения генеральных (вероятностных) характеристик (4.115)–(4.117) печатными буквами $Q(\mathfrak{D})$. Эмпирические характеристики, таким образом, имеют вид (см. (3.2))

$$\mathcal{Q}_I(\mathfrak{D}) = \mu_{\Gamma} \{g_Q(\chi; \mathfrak{D})\}; \quad (4.120)$$

$$\mathcal{Q}_{II-1}(\mathfrak{D}) = \mu_{\Gamma} \{g_Q(\chi; Q_1, \dots, Q_s; \mathfrak{D})\}; \quad (4.120a)$$

$$\mathcal{Q}_{II-2}(\mathfrak{D}) = \mu_{\Gamma} \{\tilde{g}_Q(\chi; \mathfrak{D})\} = \mu_{\Gamma} \{g_Q(\chi; \mathcal{Q}_1, \dots, \mathcal{Q}_s; \mathfrak{D})\}; \quad (4.120б)$$

$$\mathcal{Q}_{III}(\mathfrak{D}) = A \{\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_s; \mathfrak{D}\}, \quad (4.121)$$

где $\tilde{g}_Q(\cdot)$ отличается от $g_Q(\cdot)$ тем, что $g_Q(\cdot)$ аналогично преобразованию в (4.115) и либо не содержит ансамблевых (теоретических) характеристик Q_1, \dots, Q_s , либо содержит ансамблевые характеристики Q_1, \dots, Q_s , значения которых известны, в то время как $\tilde{g}_Q(\cdot)$ содержит вместо ансамблевых характеристик их аналоги – эмпирические характеристики; оператор A в (4.121) аналогичен оператору A в (4.117) (см. подробнее в [5]).

Как видно из сопоставления (4.116), (4.117) и (4.120), (4.120a), (4.120б), (4.121), в (4.120), (4.116), (4.117) входят теоретические (вероятностные или генеральные для конечной ГС) характеристики, в то время как в (4.120б) и (4.121) входят эмпирические характеристики. Если ГС является конечной, то вместо вероятностных характеристик $Q(\mathfrak{D})$ рассматриваются генеральные $\mathcal{Q}(\mathfrak{D}) = Q_{\Gamma}(\mathfrak{D})$, совпадающие по определению с ЭХ (4.120)–(4.121) при $N = N_{\Gamma}$. В дальнейшем, как уже упоминалось, будет рассматриваться только бесконечная теоретиче-

ская (гипотетическая) ГС, т. е. в качестве генеральных характеристик будут рассматриваться вероятностные $Q(\mathfrak{B})$.

Эмпирические характеристики являются разновидностью (образуют подкласс) *выборочных (статистических) характеристик*, т. е. характеристик, полученных по выборке, описывающих, характеризующих выборку.

Другие выборочные характеристики также являются аналогами $\mathcal{Q}(\mathfrak{B})$, но получаются по формулам, отличным от (4.120)–(4.121). Это отличие легко уясняется из рассмотрения приводимых ниже выборочных характеристик. (Формула (4.121) общая для всех выборочных характеристик этого типа. Отличие от таких характеристик сводится лишь к отличию в операторе $A\{\cdot\}$.)

Согласно определению понятия *статистика* все выборочные характеристики являются статистиками. Они могут иметь самостоятельное применение, а также использоваться для приближенного определения значений вероятностных (генеральных) характеристик $Q(\cdot)$, для описания всей генеральной совокупности, в качестве оценок ее характеристик $\hat{Q}(\mathfrak{B})$. Именно этот класс выборочных характеристик и будет в основном рассматриваться в дальнейшем.

Важный класс статистик, в том числе выборочных характеристик, составляют оценки характеристик $Q(\mathfrak{B})$ или $Q_{\Gamma}(\mathfrak{B})$.

Под *оценкой* характеристики $Q(\mathfrak{B})$ случайного элемента \mathfrak{N} в точке \mathfrak{B} понимается приближенное значение характеристики $Q(\mathfrak{B})$, полученное по выборке χ из ГС случайного элемента \mathfrak{N} . Оценку будем обозначать некоторым значком над Q , например $\hat{Q}(\mathfrak{B})$, $\check{Q}(\mathfrak{B})$, $Q^*(\mathfrak{B})$ и т. д. Очевидно, что понятие *оценка* относится к теоретической (математической) статистике, а сам термин *оценивание* характеристики $Q(\mathfrak{B})$ случайного элемента \mathfrak{N} означает нахождение значений $Q(\mathfrak{B})$ по случайной выборке χ из ГС случайного элемента \mathfrak{N} . Обратим в связи с этим внимание на различие и сходство понятий математической статистики *оценивание*, *оценка* (в смысле значение оценки) с понятиями измерительной техники *измерение*, *результат измерения*. Измерение есть *физическое действие*, связанное с установлением (постановкой) соответствия между *измеряемым* значением некоторой *физической* величины (длина, сила тока, масса и т. п.) и символом (например, коли-

чеством единиц измерения, содержащихся в этой физической величине) из шкалы измерения. Например, между длиной удава и количеством единиц измерения (сантиметров, спичечных коробок, попугаев и т. п.), которые этой длине соответствуют, в нее укладываются. Это количество и есть результат измерения. В то же время оценивание есть абстрактная (а не физическая) *математическая* операция, связанная, правда, с физическими действиями по формулам (4.118)–(4.121) над числами (объектами математики), а не над физическими величинами. Результат оценивания есть число, полученное по итогам выполнения численных преобразований (4.118)–(4.121) с указанием погрешности его получения, т. е. с учетом приближения его к действительности. Сходство же оценивания с некоторыми видами измерений (статистические измерения) заключается в том, что при измерениях для получения итоговых значений результаты первичных измерений (это уже числа!) могут преобразовываться по формулам типа (4.118)–(4.121), но результат преобразований по ним будет означать «количество косвенных единиц измерения», соответствующих «реальной физической величине» (в рамках принятой модели) $Q(\mathfrak{Q})$ с указанием погрешности измерения.

Учитывая двойственный характер выборки (т. е. случайный – до ее получения и неслучайный – после получения), очевидно, что апостериори любая оценка $\hat{Q}(\mathfrak{Q})$ характеристики $Q(\mathfrak{Q})$ является неслучайной функцией аргумента \mathfrak{Q} , в то время как априори ее значения случайны. Следовательно, чтобы охарактеризовать (описать, исследовать) свойства оценки, можно применить аппарат теории вероятностей. Например, оценку $\hat{Q}(\mathfrak{Q})$ характеризовать как случайную функцию аргумента \mathfrak{Q} с помощью одномерного или многомерного (для разных наборов \mathfrak{Q}) закона распределения, математического ожидания $\mathbf{M}\{\hat{Q}(\mathfrak{Q})\}$, дисперсии $\mathbf{D}\{\hat{Q}(\mathfrak{Q})\} = D_{\hat{Q}}(\mathfrak{Q}) = \sigma_{\hat{Q}}^2(\mathfrak{Q})$ или лучше среднего квадратического отклонения $\sigma_{\hat{Q}}(\mathfrak{Q})$, доверительного интервала, корреляции $\text{cor}\{\hat{Q}(\mathfrak{Q}_1), \hat{Q}(\mathfrak{Q}_2)\}$ и т. д. Такое описание важно, с одной стороны – с точки зрения получения точностных показателей и характеристик

достоверности оценки, отражающих ее качество в этом смысле, с другой – с точки зрения выработки требований, которые должны быть предъявлены к какой-либо статистике, чтобы она могла претендовать или быть использованной в качестве оценки $\hat{Q}(\mathfrak{B})$ характеристики $Q(\mathfrak{B})$. Ясно, что не всякая статистика может быть оценкой $\hat{Q}(\mathfrak{B})$ характеристики $Q(\mathfrak{B})$, а лишь такая, которая удовлетворяет определенным требованиям, свойствам оценки. Что это за требования, свойства?

1. **Несмещённость оценки.** Оценка $\hat{Q}(\mathfrak{B})$ называется *несмещенной*, если ее математическое ожидание равно оцениваемой характеристике $Q(\mathfrak{B})$, т. е.

$$\mathbf{M}\{\hat{Q}(\mathfrak{B})\}=Q(\mathfrak{B}). \quad (4.122)$$

Разность $\varepsilon_{\hat{Q}}(\mathfrak{B})=\mathbf{M}\{\hat{Q}(\mathfrak{B})-Q(\mathfrak{B})\}$ называется *смещением оценки*.

Для несмещенной оценки $\varepsilon_{\hat{Q}}(\mathfrak{B})=0$. Если $\varepsilon_{\hat{Q}}(\mathfrak{B})\neq 0$ при конечном объеме выборки N , но $\varepsilon_{\hat{Q}}(\mathfrak{B})\rightarrow 0$ при $N\rightarrow\infty$ (или $N\rightarrow N_{\Gamma}$), то такая оценка $\hat{Q}(\mathfrak{B})$ называется *асимптотически несмещенной*.

Рассмотрим с этой точки зрения пригодность эмпирических характеристик $\mathcal{Q}_1(\mathfrak{B})$ (4.120) в качестве оценок $\hat{Q}(\mathfrak{B})$ (ЭХ-оценок) вероятностных характеристик $Q(\mathfrak{B})$. Имеем

$$\begin{aligned} \mathbf{M}\{\hat{Q}_N(\mathfrak{B})\} &= \mathbf{M}\{\mathcal{Q}_1(\mathfrak{B})\} = \\ &= \mathbf{M}\left\{\frac{1}{N}\sum_{i=1}^N g_Q(\aleph_i, \mathfrak{B})\right\} = \frac{1}{N}\sum_{i=1}^N Q_i(\mathfrak{B}), \end{aligned} \quad (4.123)$$

где $Q_i(\mathfrak{B})$ – характеристика $Q(\mathfrak{B})$ для элемента χ_i выборки χ . Заметим, во-первых, что здесь заменой χ в (4.119) на \aleph в (4.123) под знаком оператора $\mathbf{M}\{\cdot\}$ подчеркивается факт априорной случайности $\hat{Q}(\mathfrak{B})$ и усреднения именно по вероятностной мере всех возможных значений χ элемента \aleph .

Во-вторых, в (4.123) учитывается факт возможной неоднородности выборки, когда i -й элемент выборки χ_i выбирается из ГС случайного элемента \aleph_i . Если выборка однородна, то $Q_i(\mathfrak{B}) = Q(\mathfrak{B})$ и (4.123) переходит в (4.122), т. е. ЭХ-оценки (4.120) являются несмещенными оценками ВХ (4.115), если выборка однородна.

В-третьих, зависимость членов выборки между собой не влияет на математическое ожидание и, следовательно, на смещение ЭХ-оценок. Можно убедиться, что $\mathcal{L}(\mathfrak{B})$ в (4.120) и (4.120а) является несмещенной оценкой $\hat{Q}_2(\mathfrak{B})$ ВХ (4.116), в то время как ЭХ $\mathcal{L}(\mathfrak{B})$ (4.120б) – в общем случае *смещенная оценка* (обозначим ее через $\check{Q}_2(\mathfrak{B})$, чтобы отличить от $\hat{Q}_2(\mathfrak{B})$ (4.120а)), когда Q_1, \dots, Q_s известны, но, как правило, асимптотически несмещенная, если $\mathcal{L}_1, \dots, \mathcal{L}_s$ есть несмещенные оценки Q_1, \dots, Q_s .

Аналогично, можно убедиться, что ЭХ $\mathcal{L}_j(\mathfrak{B})$, $j = I, II-2, III$, могут использоваться в качестве, как правило, асимптотически несмещенных оценок $Q^*_{j_i}(\mathfrak{B})$ ВХ $Q(\mathfrak{B})$ из (4.116) по так называемому *непараметрическому методу базовых характеристик* оценивания $Q(\mathfrak{B})$ (БХ-оценки).

2. Состоятельность оценки. Оценка $\hat{Q}(\mathfrak{B})$ называется состоятельной, если при $N \rightarrow \infty$ (или $N \rightarrow N_T$) она сходится по вероятности к $Q(\mathfrak{B})$. Согласно неравенствам Чебышева (см. прил. 3) достаточным условием состоятельности оценки является стремление к нулю среднего квадрата отклонения ее $\Delta_{\hat{Q}}(\mathfrak{B}) = \sqrt{\varepsilon_{\hat{Q}}^2(\mathfrak{B}) + \sigma_{\hat{Q}}^2(\mathfrak{B})} \rightarrow 0$ при $N \rightarrow \infty$ (или $N \rightarrow N_T$), т. е. условие

$$\begin{aligned} \lim_{N \rightarrow \infty} \Delta_{\hat{Q}}^2(\mathfrak{B}) &= \lim_{N \rightarrow \infty} \mathbf{M} \left\{ \left[\hat{Q}(\mathfrak{B}) - Q(\mathfrak{B}) \right]^2 \right\} = \\ &= \lim_{N \rightarrow \infty} \left[\varepsilon_{\hat{Q}}^2(\mathfrak{B}) + \sigma_{\hat{Q}}^2(\mathfrak{B}) \right] = 0. \end{aligned} \quad (4.124)$$

Для ЭХ-оценок $\hat{Q}(\mathfrak{B})$ (4.120) и (4.120а) как оценок ВХ (4.115) и (4.116) смещение $\varepsilon_{\hat{Q}}(\mathfrak{B})$ равно нулю, и, следовательно, для них достаточным условием состоятельности ЭХ $\mathcal{L}_I(\mathfrak{B})$ и $\mathcal{L}_{II-1}(\mathfrak{B})$ как ЭХ-оценок $\hat{Q}_I(\mathfrak{B})$ и $\hat{Q}_{II}(\mathfrak{B})$ является условие $\mathbf{D}\{\hat{Q}(\mathfrak{B})\} = 0$ при $N \rightarrow \infty$.

В свою очередь дисперсии оценок (4.120) и (4.120а) для однородных выборок равны

$$\mathbf{D}\{\hat{Q}(\mathfrak{B})\} = \frac{1}{N} \mathbf{D}\{g_Q(\mathfrak{N}; \mathfrak{B})\} \left[1 + \frac{1}{N} \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N \rho_{ij} \right], \quad (4.125)$$

где $\mathbf{D}\{g_Q(\mathfrak{N}; \mathfrak{B})\}$ – дисперсия $g_Q(\mathfrak{N}; \mathfrak{B})$ для оценок (4.120) характеристик (4.115) или дисперсия $g_Q(\mathfrak{N}; \mathfrak{B}; Q_1, \dots, Q_S)$ для оценок (4.120а) характеристик (4.116); ρ_{ij} – коэффициент корреляции между $g_Q(\mathfrak{N}_i; \mathfrak{B})$ и $g_Q(\mathfrak{N}_j; \mathfrak{B})$ или $g_Q(\mathfrak{N}_i; \mathfrak{B}; Q_1, \dots, Q_S)$ и $g_Q(\mathfrak{N}_j; \mathfrak{B}; Q_1, \dots, Q_S)$ соответственно. При простом выборе, т. е. для независимой выборки χ , $\rho_{ij} = 0$ при $i \neq j$ и (4.125) переходит в

$$\mathbf{D}\{\hat{Q}(\mathfrak{B})\} = \frac{1}{N} \mathbf{D}\{g_Q(\mathfrak{N}; \mathfrak{B})\}. \quad (4.126)$$

Отсюда видна *желательность требования независимости* выборки, поскольку чаще всего двойная сумма в (4.125) неотрицательна и имеет тем большие значения, чем больше зависимы члены выборки. Из (4.126) также следует состоятельность ЭХ (4.120) и (4.120а) как несмещенных ЭХ-оценок характеристик (4.115) и (4.116) соответственно.

3. Оценка $\hat{Q}_1(\mathfrak{B})$ называется *эффективной* оценки $\hat{Q}_2(\mathfrak{B})$ той же характеристики $Q(\mathfrak{B})$, если $\Delta_{\hat{Q}_1}(\mathfrak{B}) \leq \Delta_{\hat{Q}_2}(\mathfrak{B})$, что для несмещенных оценок равносильно неравенству $\mathbf{D}\{\hat{Q}_1(\mathfrak{B})\} \leq \mathbf{D}\{\hat{Q}_2(\mathfrak{B})\}$ или $\sigma_{\hat{Q}_1}(\mathfrak{B}) \leq \sigma_{\hat{Q}_2}(\mathfrak{B})$.

Методы оценивания

Подробно с основными методами и алгоритмами оценивания различных характеристик $Q(\mathfrak{B})$ случайных элементов \aleph можно ознакомиться по монографиям [5, 23], а более детально – по многочисленным литературным источникам по математической и прикладной статистике. Рассмотрим лишь некоторые из них и только на уровне идей методов. Они условно делятся на непараметрические, параметрические и смешанные.

Непараметрические методы оценивания (Н-оценки) отличаются тем, что они не требуют знания аналитического описания измеряемой характеристики $Q(\mathfrak{B})$. Поэтому иногда их называют методами, свободными от аналитического описания оцениваемых характеристик и/или моделей случайного элемента \aleph , характеристики которого оцениваются. Из них наиболее часто используются и исследуются следующие.

Метод эмпирических характеристик (ЭХ-оценки). В качестве оценок $\hat{Q}(\mathfrak{B})$ характеристики $Q(\mathfrak{B})$ берутся их статистические аналоги – эмпирические характеристики (4.120)–(4.121) (см. табл. 4.4) или их поправленные варианты, обеспечивающие несмещенность оценок.

При выполнении соответствующих условий (см., например, [5]) ЭХ-оценки являются несмещенными или асимптотически несмещенными, состоятельными, в ряде случаев асимптотически эффективными.

Помимо ЭХ-оценок для характеристик типа плотностей (распределения вероятностей, спектральных) используются ядерные (Я-оценки) или дельтаобразных функций (ДОФ-оценки), когда в (4.118) и (4.119) выборочные данные перед операцией усреднения подвергаются соответствующему ядерному или дельтаобразному функциональному преобразованию. Это такое преобразование каждого мгновенного значения x и $x(t)$, которое имеет единичный объем и стремится (стягивается) к δ -функции при неограниченном объеме выборки G . При определенных условиях такие оценки в отличие от ЭХ-оценок плотностей являются асимптотически несмещенными, состоятельными.

Следующую группу методов непараметрического оценивания составляют **методы проекций**. Они основаны на разложении оцениваемых функциональных характеристик в ряды на базе ортогональных (ОР-оценки) и неортогональных (НР-оценки) функций.

Разнообразие подобных методов оценивания, реализующих их алгоритмов и свойств оценок рассмотрено в [5].

Параметрические методы оценивания (П-оценки) предполагают использование для оценивания характеристики $Q(\mathfrak{S})$ априори известной (с точностью до неизвестных значений параметров θ) параметрической модели $q(\mathfrak{S};\theta)$ оцениваемой характеристики $Q(\mathfrak{S})$ и/или модели случайного элемента \aleph , характеристику которого $Q(\mathfrak{S})$ мы оцениваем.

П-оценивание может осуществляться двумя способами. Первый, наиболее часто применяемый, основан на использовании модели $q(\mathfrak{S};\theta)$, второй – на использовании модели элемента \aleph и так называемых достаточных статистик параметра θ характеристики $Q(\mathfrak{S};\theta)$. Вторым способом основан на непосредственном определении значений $Q(\mathfrak{S})$ для разных \mathfrak{S} по выборке χ через достаточные статистики, если, конечно, они существуют (!).

Рассмотрим в качестве примера первый способ. Он обычно реализуется в три этапа.

На первом этапе характеристика $Q(\mathfrak{S})$ представляется функцией $q(\mathfrak{S};\theta)$ n -мерного аргумента $\mathfrak{S} = (\mathfrak{S}_1, \dots, \mathfrak{S}_n)$ и k -мерного неизвестного (оцениваемого на втором этапе) параметра $\theta = (\theta_1, \dots, \theta_k)^1$.

На втором этапе по выборке χ находится оценка $\hat{\theta}$ с помощью алгоритма, зависящего от характеристики Q аналитического описания q и числа $k \leq m$ оцениваемых параметров¹.

Наконец, на третьем этапе для всех интересующих исследователя значений \mathfrak{S} по соотношению

$$\hat{Q}(\mathfrak{S}) = q(\mathfrak{S}; \hat{\theta}; \theta_{k+1}, \dots, \theta_m) \quad (4.127)$$

рассчитываются (см. косвенное измерение) искомые значения оценки $\hat{Q}(\mathfrak{S})$.

¹ Заметим, что число параметров, входящих в аналитическое выражение $q(\mathfrak{S};\theta)$, может быть m , где $m \geq k$, т. е. содержать помимо k неизвестных (оцениваемых) еще $m - k$ известных параметров (положения, масштаба, формы, связи и т. д.).

Заметим, что здесь, как и всюду до этого и в дальнейшем, мы опускаем этап, связанный с нахождением значений метрологических и других показателей оценок (см. § 3.5).

Для оценивания параметров θ в этом способе П-оценивания используется много методов и алгоритмов. Рассмотрим некоторые из них [5, 23].

Метод значений характеристик (ЗХ-оценки)¹ (!) На первом этапе находятся² k каких-либо непараметрических оценок $\hat{Q}(\mathfrak{D})$ исследуемой или простейшей содержащей параметра θ характеристики $Q(\mathfrak{D})$ для фиксированных значений \mathfrak{D} , равных $\mathfrak{D}_1, \mathfrak{D}_2, \dots, \mathfrak{D}_k$. На втором этапе составляется система из k уравнений, в левой части каждого l -го уравнения которой записывается известная (найденная или выбранная по справочнику [20]) аналитическая зависимость $q(\mathfrak{D}; \theta)$ от \mathfrak{D}_l и θ , а в правой – найденные на первом этапе по выборке χ оценки $\hat{Q}(\mathfrak{D})$ для $\mathfrak{D} = \mathfrak{D}_l$, $l = \overline{1, k}$, т. е. уравнений вида $q(\mathfrak{D}_l; \theta) = \hat{Q}(\mathfrak{D}_l)$. Заметим, что первый и второй этапы могут быть переставлены местами. На третьем этапе проводится аналитическое или численное решение системы уравнений относительно неизвестных $\theta_1, \dots, \theta_k$. Полученные в итоге решения значения $\theta_1, \dots, \theta_k$ принимаются за ЗХ-оценки $\hat{\theta}_1, \dots, \hat{\theta}_k$.

Метод числовых характеристик (ЧХ-оценки). Этот метод аналогичен ЗХ-методу. Отличие в следующем. Первое отличие: вместо непараметрических оценок $\hat{Q}(\mathfrak{D})$ для разных $\mathfrak{D} = \mathfrak{D}_1, \dots, \mathfrak{D}_k$ используются непараметрические оценки совокупности заранее отобранных на нулевом этапе числовых характеристик b_1, \dots, b_k . Характеристики b_1, \dots, b_k выбираются так, чтобы все последующие операции были как можно проще, а получаемые оценки – как можно более эффективны. Второе отличие: вместо уравнений $q(\mathfrak{D}_l; \theta) = \hat{Q}(\mathfrak{D}_l)$, $l = \overline{1, k}$ ЗХ-оценивания составляется система $b_l(\theta) = \hat{b}_l$, $l = \overline{1, k}$, где $b_l(\theta)$ – аналитическая зави-

¹ Условия существования, свойства оценок, рекомендации по их использованию, в частности по выбору $\mathfrak{D}_1, \mathfrak{D}_2, \dots, \mathfrak{D}_k$, опускаются (см. [5, 23] и другие работы по математической и прикладной статистике).

² По числу оцениваемых параметров $\theta_1, \dots, \theta_k$.

симось числовой характеристики b_l от параметров $\theta_1, \dots, \theta_k$. Идеальный, но трудно достижимый случай, когда b_l зависит только от одного параметра θ_s , а в итоге все b_l , $l = \overline{1, k}$, охватывают зависимость от всех θ_s из $\theta_1, \dots, \theta_k$. Здесь \hat{b}_l – некоторая (желательно лучшая, эффективная) непараметрическая оценка.

Если b_l выбирается из класса моментных характеристик, то такой метод оценивания параметров называется *методом моментных характеристик* (МХ-оценки) или, компактнее, *методом моментов*. Если b_l – это квантили, то метод ЧХ-оценки называется методом *квантильных характеристик* (КХ-оценки) или, компактнее, *метод квантилей*.

Метод максимального правдоподобия (МП-оценки). Методы ЗХ- и ЧХ-оценивания являются идейно и технически просто реализуемыми, но зачастую не позволяют получить эффективные оценки. В математической статистике наиболее популярным для оценивания параметров θ характеристик $Q(\mathfrak{F}; \theta)$ считается метод максимального правдоподобия (МП-оценивания). Дело в том, что при достаточно общих необходимых и достаточных условиях (!) МП-оценки являются асимптотически несмещенными, состоятельными, эффективными и к тому же обладают рядом других замечательных свойств. Но (конечно!) только в том случае, если эти условия выполняются. Свойство утрачиваются, если условия нарушаются.

МП-оценки находятся как такое значение векторного параметра θ , при котором функция правдоподобия (ФП) выборки χ $L(\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_N | \theta)$ принимает максимальное значение¹, т. е. МП-оценка $\hat{\theta}$ равна

$$\hat{\theta} = \arg \max_{\theta} L(\chi | \theta), \quad (4.128)$$

где ФП $L(\chi | \theta) = L(\mathfrak{x}_1, \mathfrak{x}_2, \dots, \mathfrak{x}_N | \theta)$ есть рассматриваемый как функция аргумента θ *отсчет* N -мерной плотности распределения вероятностей выборки χ , когда аргументы плотности равны (заменены на

¹ Если ФП многоэкстремальная, то выбирается такое θ , которое соответствует наибольшему из ее максимумов (принцип максимум максиморе).

значения) значениям элементов выборки. Например, для выборки $\chi' = (x'_1, x'_2, \dots, x'_N)$ из генеральной совокупности случайной величины X ФП будет $W(x'_1, \dots, x'_N | \theta)$, а для $\chi = (x_1, \dots, x_N)$ ФП есть отсчет $W(x_1, \dots, x_N | \theta)$, когда x_1, \dots, x_N – не аргументы плотности, а их конкретные значения, полученные в выборке χ (не путаться!).

Если выборка χ независимая в совокупности, то согласно определению независимости¹

$$L_N(\chi | \theta_0) = L(x_1, \dots, x_N | \theta_0) = \prod_{i=1}^N W_X(x_i | \theta_0), \quad (4.129)$$

где $W_X(x_i | \theta_0)$ – отсчет плотности распределения случайной величины X , когда аргумент плотности равен значению x_i i -го элемента выборки при условии, что неизвестный параметр θ принимает значение θ_0 .

Для нахождения $\hat{\theta}$ согласно (4.128) удобнее прологарифмировать (4.129) и перейти от поиска $\hat{\theta}$ по (4.128) к решению системы уравнений

$$\frac{\partial}{\partial \theta_i} \ln L(\chi | \theta) = 0, \quad i = \overline{1, k}. \quad (4.130)$$

Решение системы и есть МП-оценка $\hat{\theta}$.

Метод минимума меры близости (расстояния) (ММБ-оценки). Идея метода сводится к введению меры близости (расстояния) $d\{\cdot\}$ между непараметрической оценкой $\hat{Q}(\mathfrak{S})$ характеристики $Q(\mathfrak{S})$ и ее известным параметрическим описанием $q(\mathfrak{S}; \theta)$ с неизвестным θ и последующим нахождением такого значения $\hat{\theta}$, при котором $d\{\hat{Q}(\mathfrak{S}); q(\mathfrak{S}; \theta)\}$ будет минимальным. Такое значение $\hat{\theta}$ и будет оценкой неизвестного параметра θ :

¹ Заметим, что если выборка неоднородная, то $W_X(x_i | \theta_0)$ в (4.129) надо заменить на $W_{X_i}(x_i | \theta_0)$, где W_{X_i} – плотность распределения случайной величины X_i , приписываемой i -му элементу выборки χ .

$$\hat{\theta} = \arg \min_{\theta} d \left\{ \hat{Q}(\theta); q(\theta; \theta) \right\}. \quad (4.131)$$

Конкретные разновидности ММБ-оценок зависят от вида меры $d\{\cdot\}$ и алгоритмов реализации метода [5].

Введем оператор

$$\mathfrak{I}\{f(x)|_{\mathcal{X}}\} = \begin{cases} \int f(x)dx, & \text{если } x \in \mathcal{X}, \mathcal{X} \text{ – непрерывно;} \\ \mathcal{X} \\ \sum_{i \in I} f(x_i), & \text{если } x_i \in \mathcal{X} \quad \forall i \in I; \mathcal{X} \text{ – дискретно.} \end{cases} \quad (4.132)$$

Если

$$d \left\{ \hat{Q}(\theta); q(\theta; \theta) \right\} = \mathfrak{I} \left\{ \left[\hat{Q}(\theta) - q(\theta; \theta) \right]^2 \omega(\theta; \theta) \Big|_{\mathcal{V}} \right\}, \quad (4.133)$$

где \mathcal{V} – область значений аргумента θ (для дискретных θ – число значений θ не менее k), а $\omega(\theta; \theta)$ – заданная детерминированная весовая функция (!), то метод ММБ-оценивания называется **методом обобщенных наименьших квадратов** (ОНК-оценки). При $\omega(\theta; \theta) = 1$ метод получил название **метода наименьших квадратов** (НК-оценки). Если $\omega(\theta; \theta) = \omega(\theta)/q(\theta; \theta)$, то получаем **обобщенный метод минимума хи-квадрат** (ОМХК-оценки), а при $\omega(\theta) = 1$ – **метод минимума хи-квадрат** (МХК). При $\omega(\theta; \theta) = \omega(\theta)/\hat{Q}(\theta)$ имеем **обобщенный модифицированный МХК** и т. д. Если же

$$d \left\{ \hat{Q}(\theta); q(\theta; \theta) \right\} = \mathfrak{I} \left\{ \omega(\theta; \theta) \left| \hat{Q}(\theta) - q(\theta; \theta) \right| \right\}, \quad (4.134)$$

то получаем **обобщенный метод наименьших модулей** (ОНМ-оценки), из которого при $\omega(\theta; \theta) = 1$ получаем **метод наименьших модулей** (НМ-оценки).

Подробно об этих и других методах оценки параметров θ изложено в [5].

Смешанные методы оценивания. Согласно названию эта группа методов основана на комбинации непараметрических и параметрических приемов оценивания [5]. Они реализуются двумя способами. Первый основан на апостериорной аппроксимации модели случайного элемента \aleph , второй – модели оцениваемой характеристики $Q(\mathfrak{B})$. Первый способ в настоящее время бурно развивается в приложении к оцениванию спектральных плотностей мощности. Второй способ чаще всего реализуется в два этапа, когда на первом этапе находится непараметрическая оценка искомой или базовой¹ характеристики, а на втором этапе осуществляется ее аппроксимация путем подбора подходящей математической модели и оценивания ее параметров, а затем, в случае необходимости, параметрическое оценивание всех других искомых характеристик на основе формулы (4.127). В обоих способах смешанное оценивание может осуществляться методом, основанном на априорном упорядочении и апостериорном выборе модели случайного элемента \aleph или характеристики $Q(\mathfrak{B})$, с последующим параметрическим оцениванием всех требуемых характеристик. Особый интерес в этом случае представляет метод вероятностной моделетеки, рассматриваемый во второй части учебного пособия (см. [5, 20]).

С2. Статистическая проверка гипотез

Теория оценивания решает количественные задачи статистики, т. е. разрабатывает аппарат для нахождения по выборке χ значений статистических характеристик как оценок характеристик случайного элемента \aleph , модель которого приписывается генеральной совокупности. В приложении к идентификации стохастических объектов слово *статистический* означает следующее. Вначале мы убеждаемся в пригодности для описания исследуемого объекта стохастических моделей.

¹ Под базовой понимается характеристика, во-первых, существующая для всех практических случаев; во-вторых, содержащая все необходимые параметры $\theta_1, \dots, \theta_k$; в-третьих, имеющая эффективные оценки как ее, так и ее параметров $\theta_1, \dots, \theta_k$ по ней; в-четвертых, позволяющая по ней получать функциональные зависимости любых других искомых характеристик Q_1, \dots, Q_s от параметров $\theta_1, \dots, \theta_m$, чтобы по оценкам $\theta_1, \dots, \hat{\theta}_k$ и известным $\theta_{k+1}, \dots, \theta_m$ найти параметрические оценки $\hat{Q}_1, \dots, \hat{Q}_s$, воспользовавшись (4.127) [5].

Затем проводим пассивное или активное экспериментирование, набирая «статистику» – измеряя и накапливая как можно больше данных об объекте, удовлетворяющих требованиям к генеральной совокупности (ГС). После этого по этим данным, как по выборке из ГС, находим статистические характеристики. Они могут далее использоваться для решения задач исследования объекта как апостериорные самостоятельные «статистические модели» (характеристики) или как оценки характеристик вероятностных моделей.

В отличие от оценивания проверка гипотез связана с получением качественного (категорийного, бинарного) ответа: да, гипотеза «верна» (лучше утверждать: «не отвергается данными») или «не верна» (отвергается). Слово *статистическая* в сочетании *статистическая проверка гипотез* здесь имеет тот же смысл, что и в статистическом оценивании. Вначале выдвигается гипотеза, затем обосновывается пригодность для ее проверки стохастических моделей, подходов. После этого проводится серия опытов (набирается «статистика») и на их основании проверяется справедливость выдвигаемых гипотез. В связи с изложенным ответьте на вопрос: «Какой смысл вкладывается в словосочетания *статистическая проверка гипотез*, *проверка статистических гипотез* и *статистическая проверка статистических гипотез*? Можно ли отождествлять эти словосочетания? Какие еще методы проверки гипотез, кроме статистических, вы знаете?

Внимание! Если вы не сможете ответить на последний вопрос сейчас, попробуйте ответить на него перед защитой выпускной работы.

Рассмотрим простейшие варианты статистической проверки гипотез (СПГ). Предположим, что некоторый исследуемый объект может находиться в $m + 1$ возможных непересекающихся состояниях S_0, S_1, \dots, S_m , каждому S_i , $i = \overline{0, m}$, из которых мы можем поставить в соответствие случайное событие A_i , $i = \overline{0, m}$. Допустим далее, что возможные состояния S_0, S_1, \dots, S_m взаимоисключающие и охватывают все возможные состояния объекта, которые наступают с априорной вероятностью p_0, p_1, \dots, p_m . На языке случайных событий это означает, что A_i – несовместные события, образующие полную группу, т. е. $p(A_0) + p(A_1) + \dots + p(A_m) = 1$.

Например, состояние S_0 – кровяное давление в норме, S_1 – ниже нормы, S_2 – выше нормы; урожайность первого сорта яблок U_0 ,

а второго сорта – U_1 на том же участке при той же погоде; плотность распределения $W_X(x)$ показателя X объекта, находящегося в состоянии S_0 , есть $W_0(x)$ или $W_X(x; \theta_0)$, находящегося в состоянии S_1 , – $W_1(x)$ или $W_X(x; \theta_1)$. Необходимо проверить справедливость этих утверждений и выявить (принять одно из решений $\gamma_0, \gamma_1, \dots$ или γ_m) о том, в каком состоянии находится объект. Например, решение γ_i , что справедлива гипотеза H_i о том, что имеет место событие A_i – объект находится в состоянии S_i . Статистическая проверка гипотез означает, что нужно провести стохастический эксперимент из N опытов, получить по нему данные (выборку χ), отражающие состояние объекта, т. е. получить эмпирические значения x_1, x_2, \dots, x_N отражающего эти состояния показателя X объекта, описываемого случайной величиной X с плотностью распределения $W_X(x)$. Затем проверить гипотезу, принять решение γ_i о справедливости гипотезы $H_i, i = \overline{0, m}$.

Одну из гипотез назовем *основной*, например H_0 , а другие – *конкурирующими*, или *альтернативными*. Для простоты положим, что $m = 1$, т. е. рассмотрим только бинарный случай справедливости одной из двух гипотез H_0 или H_1 . Их априорные вероятности равны $P(H_0) = P(A_0) = p_0, P(H_1) = P(A_1) = p_1, p_0 + p_1 = 1$.

Уже при планировании эксперимента возникает несколько вопросов (задач), на которые надо найти ответ.

Первый: «Какую гипотезу следует взять в качестве основной, проверяемой: H_0 или H_1 ?» Проблему выбора поясним простейшим примером. Некто утверждает, что он обладает каким-то уникальным даром. Какую гипотезу следует проверять: H_0 – он говорит правду или H_1 – он лжет? Понятно, что во втором случае достаточно проводить опыты с данным человеком до первого опыта, когда будет доказано, что он лжет, т. е. гипотеза H_1 верна. Тогда как в первом случае даже подтверждение его правоты в большом числе опытов позволит сделать только вывод о том, что эксперимент не опровергает гипотезу H_0 .

Второй вопрос: «Как назначить необходимое количество N опытов?» Из приведенного примера проверки гипотезы о наличии (или

отсутствии!) дара ясно, что иногда число опытов может быть неограниченным, если их искусственно не остановить. Так как же выбирать N при планировании эксперимента?

Третий вопрос: «Как организовать технологический процесс экспериментирования?» Пусть, например, рассматриваемый при обсуждении первого вопроса дар человека – это его телепатические способности или умение угадать, что вначале наливается в кружку чая: заварка или кипяток. Как лучше организовать эксперименты? Вот примеры их реализации:

а) по жесткой программе в каком-то порядке (например, по очереди) в каждом опыте эксперимента даются разные образы при телепатии или кружки с разным способом приготовления чая: сначала заварка, потом кипяток или наоборот;

б) постоянно дается один и тот же образ или кружка с тем же способом приготовления;

в) случайным образом выбираются образцы из некоторого набора или способы приготовления чая.

Внимание! Придумайте сами другие варианты.

Отстранимся от индивидуальных особенностей, психологической устойчивости и других важных факторов, влияющих на итоги нашего опыта и эксперимента в целом. Оставим только формальную математическую постановку задачи, позволяющую ответить на поставленный вопрос¹.

Наконец, еще один, четвертый, важнейший вопрос: «Как принять решение о подтверждении, опровержении или допустимости, справедливости (непротиворечивости Данным) проверяемой гипотезы?» Иначе: «Каков должен быть формальный критерий, позволяющий принимать однозначные решения после каждого опыта или по окончании эксперимента из N опытов?».

Рассмотрим базовые, используемые в статистической проверке гипотез, решения, как ответы на эти вопросы.

Ответ на первый вопрос в общих чертах таков: выбор основной гипотезы осуществляют вместе два специалиста – прикладник и математик-статистик. Первый определяет гипотезу содержательно, второй разрабатывает формализованную постановку задачи проверки гипотез,

¹ Здесь, кстати, наглядно видна практическая пригодность формализации только отдельных аспектов исследуемого объекта, явления.

учитывая изложенное при формулировке первого вопроса и излагаемое далее.

Второй вопрос о назначении объема выборки N касается не только проверки гипотез, но и оценивания характеристик. При оценивании он уместен и при активных экспериментах, и для пассивных, если объем данных велик, а для решаемой задачи можно обойтись меньшим объемом выборки, чем имеется в наличии. В общем виде ответ на него прост: объем N выбирается как минимальное значение числа опытов (элементов выборки), при которых обеспечивается требуемое качество решения итоговой задачи исследования объекта по принятым показателям. Здесь и далее для упрощения предполагаем, что все опыты (выборка) являются независимыми в вероятностном смысле. Следовательно, вся конкретика в ответе сводится к показателям качества, во-первых, итогового результата достижения поставленной цели исследования объекта; во-вторых, получаемых на базе итогового показателя допустимых значений показателей качества оценивания и проверки гипотез. Кстати заметим, что при проверке гипотез мы обязательно имеем дело со статистиками, которые можно рассматривать как оценки соответствующих показателей. Однако зачастую рассмотрение их и их показателей качества как оценок не производится, поскольку рассматриваются сразу только показатели качества результатов проверки гипотез. Именно допустимые (граничные) значения этих показателей качества определяются, исходя из требований к качеству результатов решения итоговой задачи (см. текст далее).

Необходимость ответа на второй вопрос обусловлена тем, что при использовании статистического подхода всегда трудно, а иногда просто невозможно делать безошибочные выводы по ограниченному объему данных N . Из-за ограниченного объема опытов (выборки) возможны два рода (типа, вида) ошибок (*именно ошибок, неверных решений, а не погрешностей!*). **Ошибки первого рода** состоят в том, что отвергается основная гипотеза H_0 , хотя на самом деле она верна. **Ошибки второго рода** – принятие решения о том, что отвергается конкурирующая (альтернативная) гипотеза H_1 , когда она верна.

Обратите внимание! В ошибках речь идет не об ошибках приема гипотез. Основной акцент делается на приеме решения об отклонении гипотез, на доказательстве, что они неверны, вместо доказательства их истинности.

Поскольку мы рассматриваем статистические методы проверки гипотез, количественно описать возможность появления ошибок в экспериментах можно с помощью их вероятностей (или частотей), т. е.

вероятностей p_I ошибок I рода и p_{II} – II рода. В связи с этим вернемся к выбору проверяемых гипотез.

Примеры гипотез. По отметке на экране радиолокационной станции (РЛС) о наличии в контролируемом регионе неба цели (самолета, ракеты) необходимо решить – «своя» она (гипотеза H_0) или «чужая» (гипотеза H_1). Тогда ошибки I рода называются *ложной тревогой*, а ошибки II рода – *пропуском цели* (сигнала). Вероятность ошибки I рода равна вероятности приема решения γ_1 о том, что верна гипотеза H_1 : это чужой самолет, когда на самом деле это свой самолет, т. е. верна гипотеза H_0 (имеет место состояние S_0):

$$p_I = P(\gamma_1 | \gamma_0) = P(H_1 | H_0) = P(S_1 | S_0).$$

Вероятность же ошибки II рода:

$$p_{II} = P(\gamma_0 | \gamma_1) = P(H_0 | H_1) = P(S_0 | S_1).$$

Внимание! Подумайте, какую гипотезу здесь лучше взять за основную, отвергаемую, подтверждаемую, если при ложной тревоге будут одни затраты для страны, а при пропуске цели (с мощным боезарядом) – другие.

Рассмотренные задачи проверки гипотез характерны тем, что в них считаются известными m гипотез. Однако это не всегда так. В ряде случаев нет оснований выдвигать явно конкурирующие с основной гипотезы. Достаточно в качестве альтернативной подразумевать невыполнимость основной гипотезы. Иными словами, формулировать задачу следует не так, как это делалось ранее: какая из m гипотез верна, т. е. в каком из m состояний находится объект, а в ином виде: согласуются ли результаты эксперимента с высказанной гипотезой, предположением о соответствующих особенностях, свойствах, характеристиках Данных (объекта). Например указать, применим ли для их описания конкретный закон распределения; является ли выборка однородной; являются ли величины X и Y некоррелированными (вместо гипотез, что коэффициент корреляции равен $0(H_0)$; $0,2(H_1)$; ...; $0,7(H_m)$) и т. п. Используемые для проверки подобных гипотез правила называются *критериями согласия*. Сами же гипотезы являются чаще всего непараметрическими, сложными. Рассмотрим эти понятия.

Гипотеза, не разложимая на составные, называется *простой*. Иными словами, простая – это гипотеза, состоящая из одной единственной проверяемой гипотезы (типа равенства). Если проверяемую или альтернативную гипотезу можно разбить на несколько (в том числе счетное или континуальное бесконечное множество) простых гипотез, то она называется *сложной*.

На формальном языке математической статистики проверка гипотез означает, что надо:

а) либо проверить, принадлежит ли выборка χ_X генеральной совокупности (ГС) величины X с плотностью $W_0(x)$ (гипотеза H_0 о том, что объект находится в состоянии S_0) или она принадлежит ГС с другой плотностью $W_1(x)$, отличающейся от $W_0(x)$ видом (сложная гипотеза H_1) или только значениями одного или более параметров (простая гипотеза H_1);

б) либо проверить принадлежность двух выборок, соответствующих двум измерениям состояния объекта, одной и той же ГС с одним и тем же распределением по типу (форме, сложная гипотеза) или с распределением одного и того же типа с одинаковыми параметрами (простая гипотеза, разное проявление одного состояния объекта в результатах двух опытов за счет случайностей при измерении его).

Это, например, сложные гипотезы о законах распределения, параметрические и непараметрические о равенстве математических ожиданий или дисперсий, некоррелированности величин X и Y (равенстве нулю коэффициента корреляции ρ_{XY}) и т. п. Например, при проверке гипотез о распределении *простая* это гипотеза H_0 о том, что выборка χ соответствует конкретному распределению $W_0(x; \theta_0)$ (нормальному, гамма-распределению, экспоненциальному, равномерному) с точным значением параметра $\theta = \theta_0$. Альтернативная ей простая гипотеза H_1 :

$$W(x) = W_0(x; \theta_1),$$

а сложные H_1 :

$$W(x) = W_0(x; \theta) \text{ при } \theta \in [\theta_1; \theta_2],$$

где $\theta_1 \neq \theta_0$, $\theta_2 \neq \theta_0$, $\theta_1 \neq \theta_2$ либо $\theta \leq \theta_0$ или $\theta > \theta_0$.

Простая гипотеза H_0 : средняя урожайность культуры $m_X = a_0$. Простая альтернативная гипотеза H_1 : $m_X = b$. Сложная гипотеза

$m_X > a$. Простая гипотеза H_0 о равенстве разбросов двух выборок χ_1 и χ_2 той же ГС, выражаемых через дисперсии, т. е. $H_0: \sigma_1^2 = \sigma_2^2$, против сложной $H_1: \sigma_1^2 \neq \sigma_2^2$.

Если гипотеза касается проверки параметров при известном распределении ГС, она называется *параметрической*. Если же речь идет о проверке гипотез при неизвестном распределении или же о самом виде распределения, то такие гипотезы называются *непараметрическими* (см. параметрическое, непараметрическое и смешанное оценивание). Понятно, что можно рассматривать и смешанные гипотезы.

Внимание! Предложите их вариант (варианты) самостоятельно.

Вернемся ко второму и третьему вопросам после рассмотрения четвертого. Ограничимся при этом только проверкой гипотез, для которых достаточно описания объекта на уровне случайных величин: X_0 , соответствующей состоянию S_0 (верна гипотеза H_0), и X_1 (состояние S_1 , верна гипотеза H_1).

Пусть проводится статистический эксперимент с объектом из N опытов (говоря математически – осуществляется выборка из ГС объема N). В результате получается выборка $\chi = (x_1, x_2, \dots, x_N)$ значений X . Ее можно изобразить либо в виде точки в N -мерном пространстве, либо в виде эмпирического распределения $\hat{W}_X(x|\chi)$. Поскольку выборка случайная, в другой раз, в другом эксперименте получится другая выборка (последовательность) x_1, x_2, \dots, x_N . Все повторные точки χ в N -мерном *выборочном пространстве* будут попадать в область, соответствующую распределению $W_X(x|S_0)$, если объект находится в состоянии S_0 , или в альтернативную область, соответствующую состоянию S_1 , т. е. распределению $W_X(x|S_1)$ (рис. 4.7, а). На рис. 4.7, б изображены четыре варианта выборок χ'_0, χ''_0 и χ'_1, χ''_1 , две из которых (χ'_0, χ''_0) получаем для ситуации, когда объект находился в состоянии S_0 , и две (χ'_1, χ''_1) – когда объект был в состоянии S_1 . Аналогичное можно сказать о расположении и повторяемости значений x_1, x_2, \dots, x_N от выборки к выборке того же объема N на оси x (рис. 4.7, б).

По расположению точек выборки χ в N -мерном выборочном пространстве можно предположить, в каком состоянии S_0 или S_1 был

объект, с которым проводился эксперимент, т. е. какая гипотеза – H_0 или H_1 – наиболее правдоподобна. Из рис. 4.7 видна причина появления ошибочных решений γ_1 вместо γ_0 или γ_0 вместо γ_1 , т. е. появление ошибок I и II родов: выбросы точек в альтернативную область выборочного пространства (рис. 4.7, а) и весьма вероятное появление таких значений (рис. 4.7, б).

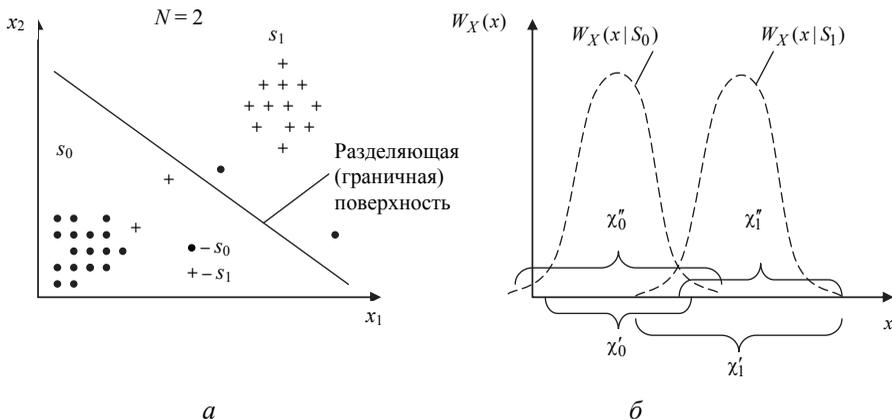


Рис. 4.7. Пример расположения значений выборок χ_0 и χ_1 при состоянии объекта S_0 и S_1

Иными словами, из рис. 4.7 видно, что N -мерное выборочное пространство можно разбить на две области, одна из которых s_0 соответствует состоянию S_0 (верна гипотеза H_0), а вторая s_1 – состоянию S_1 (верна гипотеза H_1). Тогда при попадании выборочной точки χ в соответствующую область s_0 или s_1 можно принять решение γ_0 или γ_1 , что выборка χ соответствует гипотезе H_0 или H_1 . Область s_1 , при попадании в которую выборочной точки χ отвергается основная гипотеза H_0 , называется **критической**.

Понятно, что работать в N -мерном пространстве, особенно при больших N , непросто. Поэтому рассматриваются разные способы сведения N -мерного выборочного пространства к одномерному. Это делается с помощью различных статистик – одномерных функций φ от элементов выборки χ и заменой разделяющих поверхностей

в N -мерном пространстве (рис. 4.7, *a*) граничными (разделяющими) значениями φ_0 статистики φ по различным правилам (критериям).

Правило принятия по выборке χ решения γ_i , $i = \overline{0, m}$, из множества $\gamma_0, \gamma_1, \dots, \gamma_m$, т. е. правило, по которому устанавливается соответствие набора решений $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_m)$ и возможных результатов эксперимента (возможных значений (x_1, x_2, \dots, x_N) выборки χ), называется **критерием проверки гипотез**. Вероятность ошибки первого рода p_I есть вероятность $P(\gamma_1 | S_0) = P(\chi \in s_1 | S_0) = P(H_1 | H_0)$, т. е. вероятность принять конкурирующую гипотезу H_1 , когда верна основная гипотеза H_0 . Она называется **уровнем значимости критерия** и обычно задается некоторым значением α , которое нежелательно превышать. Вероятность p_{II} ошибки II рода, обычно обозначаемая через β , есть $p_{II} = P(\gamma_0 | S_1) = P(\chi \in s_0 | S_1) = P(H_0 | H_1)$, так как это вероятность отвергнуть верную конкурирующую гипотезу H_1 . Обычно H_0 и H_1 выбирают так, чтобы желательным свойством «хорошести» критерия их проверки была минимальность вероятности β ошибки II рода при фиксированной ошибке α I рода. Тогда вероятность $P(H_1 | H_1) = P(\gamma_1 | S_1) = P(\chi \in s_1 | S_1) = 1 - \beta$ того, что значения выборки (точка χ в N -мерном пространстве) попадут в критическую область s_1 , когда объект находится в состоянии S_1 , т. е. когда действительно верна конкурирующая гипотеза H_1 , будет максимальной. Эта вероятность $1 - \beta = P(\chi \in s_1 | S_1)$ называется **мощностью критерия**. Поэтому ясно, что нужно выбирать такой критерий, для которого при заданном α мощность $1 - \beta$ будет наибольшей из всех других возможных критериев. Это один из вариантов назначения (выбора) критериев – правил выбора решений γ_0 или γ_1 при формализации постановки задачи статистической проверки гипотез и, следовательно, выбора вида статистики φ и граничного значения φ_0 .

А какие варианты могут быть еще?

Оказывается есть всего три «рычага» управления выбором критериев.

Первый «рычаг» связан с выбором вида критериальной статистики – функции φ , по которой конкретная выборка $\chi = (x_1, x_2, \dots, x_N)$ преобразуется в одно число – значение статистики φ . Дело в том, что определять границу, разделяющую основную зону s_0 N -мерного выборочного пространства от критической s_1 тем сложнее, чем больше объем выборки N . Поэтому, как уже упоминалось, желательно все N -мерное пространство R^N выборок χ преобразовать в одномерное R^1 с помощью критериальной функции $\phi(\chi)$:

$$\varphi = \phi(\chi) = \phi(x_1, x_2, \dots, x_N). \quad (4.135)$$

Первая причина разнообразия критериев заключается в разнообразии видов функции (4.135).

Второй «рычаг» связан с назначением на прямой значений φ такой граничной точки φ_0 , которая будет играть роль границы между основной s_0 и критической s_1 областями N -мерного выборочного пространства. Например, если объект находится в состоянии S_0 , т. е. верна гипотеза H_0 , и для подавляющего числа экспериментов выборка χ попадает в основную зону s_0 , то должно выполняться условие: для таких выборок χ φ_0 должно быть таковым, чтобы выполнялось неравенство $\varphi(\chi) \leq \varphi_0$. Вследствие этого принимается решение γ_0 о том, что гипотеза H_0 (система находится в состоянии s_0) не отвергается, и, следовательно, отвергается гипотеза H_1 . В противном случае, если при других χ $\varphi(\chi) > \varphi_0$, то гипотеза H_0 отвергается и, следовательно, не отвергается конкурирующая гипотеза H_1 . Тогда вероятности ошибок I и II родов будут иметь вид

$$p_I = \mathbf{P} \{ [\varphi(\chi) > \varphi_0 | H_0] \} = \mathbf{P}(\gamma_1 | S_0); \quad (4.136)$$

$$p_{II} = \mathbf{P} \{ [\varphi(\chi) \leq \varphi_0 | H_1] \} = \mathbf{P}(\gamma_0 | S_1). \quad (4.137)$$

Именно выбор вида критериальной функции $\varphi(\chi)$ и граничного значения φ_0 и определяет вид и наименование критериев проверки гипотез. Как любая статистика, φ (4.135) является апостериори случайной (детерминированной) величиной, а априори случайной величиной.

ной Φ . Следовательно, для нее можно найти плотность (или функцию) распределения вероятностей $W_{\Phi}(\varphi)$ (рис. 4.8).

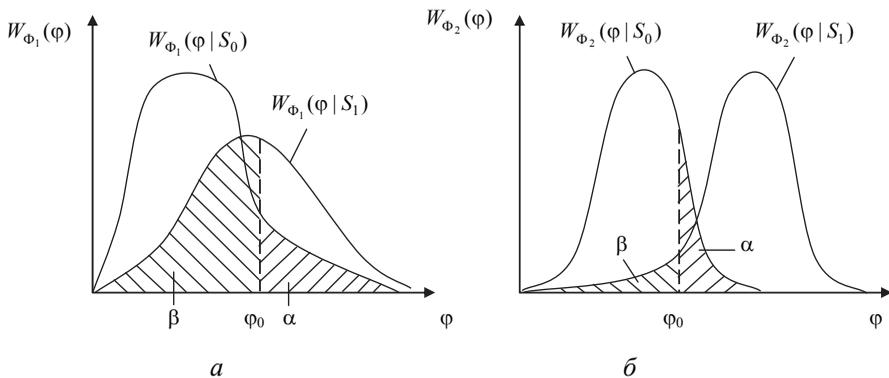


Рис. 4.8. Плотности распределения $W_{\Phi}(\varphi)$ критериальных статистик φ_1 (а) и φ_2 (б) и соответствующих им вероятностей ошибок I рода α и II рода β при фиксированном φ_0 (заштрихованные площади)

Затем по $W_{\Phi}(\varphi)$ при необходимости можно найти вероятности ошибки I и II рода. По ним при построении критериев можно найти такие φ и φ_0 , которые будут в каком-то смысле «хорошими» с точки зрения пользователя и специалиста по проверке гипотез. При применении же выбранных критериев, заданных φ и φ_0 , можно осуществлять проверку гипотез, рассчитав конкретное значение $\varphi(\chi)$ по полученной выборке χ и сравнивая его с φ_0 . Как правило, φ_0 находится по таблицам значений вероятностей попадания φ в зоны $\varphi(\chi) \leq \varphi_0$ и $\varphi(\chi) > \varphi_0$ для каждого распределения $W_{\Phi}(\varphi)$, соответствующего виду функции φ (наименованию критерия). Изложенное поясняется рис. 4.8.

Из рисунка хорошо видно, как, варьируя видом статистики $\varphi(\chi)$ и граничным значением φ_0 , можно влиять на значения вероятностей ошибок α и β .

Рассмотрим несколько критериев проверки гипотез.

Вначале рассмотрим бинарную ситуацию проверки простой гипотезы H_0 против простой альтернативы H_1 . Критерии проверки таких гипотез были введены разными авторами в разное время, исходя из

разных подходов (правил) и понимания, что такое «хороший» критерий. Как оказалось позже, большинство из них используют в качестве критериальной функции $\varphi(\chi)$ различные частные случаи статистики $l_N(\chi)$, называемой *отношением правдоподобия* (отсюда обобщенное название таких правил – *критерии отношения правдоподобия*):

$$\varphi(\chi) = l_N(\chi) = l(x_1, \dots, x_N) = \frac{L_N(\chi|S_1)}{L_N(\chi|S_0)}, \quad (4.138)$$

где $L_N(\chi|S_j)$ – функция правдоподобия (4.129), когда выборка χ получается при условии $S_j, j = 0$ или 1 .

Если при такой $\varphi(\chi)$ граничное значение $\varphi_0 = 1$, то критерий называется *критерием максимального правдоподобия*. Согласно этому критерию по выборке χ не отвергается (условно говорим: «принимается») та гипотеза H_j из H_0, H_1, \dots, H_m , для которой значение функции правдоподобия $L_N(\chi|S_j)$ будет больше значений других функций правдоподобия. Это означает, что при $m = 2$ гипотеза H_0 «принимается» при $\varphi(\chi) < 1$, а H_1 – при $\varphi(\chi) \geq 1$.

Ранее мы указывали на подход (правило) назначения критерия проверки гипотез, когда «наилучшим» (оптимальным, хорошим) считается такой, при котором $\varphi(\chi)$ и φ_0 обеспечивают наибольшую мощность критерия, т. е. минимальное значение вероятности ошибки II рода β при условии, что значение вероятности ошибки I рода не превышает заданный уровень значимости α . Оказывается, этому правилу также соответствует критериальная функция (4.138), а значение φ_0 при этом определяется неравенством (!)

$$\mathbf{P}\{\ln(\chi) \geq \varphi_0 | S_0\} = \alpha. \quad (4.139)$$

Такое правило называется *критерием Неймана–Пирсона*.

Рассмотренные два критерия проверки простой гипотезы против простых альтернатив не требовали знания какой-либо априорной информации о решаемой исследовательской задаче: априорных вероятностях $P(H_0), P(H_1), \dots, P(H_m)$ гипотез H_0, H_1, \dots, H_m , о потерях от принятия соответствующих решений $\gamma_0, \gamma_1, \dots, \gamma_m$ и т. п. А если мы

имеем такие сведения? Можно ли их использовать? Оказалось, что можно, и было введено несколько критериев для таких ситуаций. Они могут быть объединены единым подходом байесовского правила выбора критерия, основанного на минимизации среднего риска.

Рассмотрим опять проверку простой гипотезы H_0 против простой альтернативы H_1 . Предположим, что на основе проверки гипотез мы решаем практическую задачу. Например, принимаем решение «свой»–«чужой» по отметке на РЛС, на основании которого ничего не делаем ($\gamma_0 | s_0$), правильно ($\gamma_1 | s_1$) или ошибочно ($\gamma_1 | s_0$) посылаем навстречу «цели» самолеты или сбиваем «цель» ракетой, либо пропускаем «цель», не обнаружив ее ($\gamma_0 | s_1$). Назначим плату за такие решения Π_{jk} , где j соответствует состоянию S_j , а k – решению γ_k , так что¹

$$\Pi_{01} > \Pi_{00} \geq 0; \quad \Pi_{10} > \Pi_{11} \geq 0. \quad (4.140)$$

Предположим также, что нам известны априорные вероятности $p = P(H_0)$ и $q = 1 - p = P(H_1)$.

Тогда *среднее значение потерь (средний байесовский риск) R* от принятия решений по проверяемым гипотезам будет равен

$$R = pr_0 + qr_1 = pr_0 + (1-p)r_1, \quad (4.141)$$

где

$$r_0 = (1-\alpha)\Pi_{00} + \alpha\Pi_{01}; \quad (4.142)$$

$$r_1 = \beta\Pi_{10} + (1-\beta)\Pi_{11} \quad (4.142a)$$

есть условные риски, соответствующие состояниям S_0 и S_1 объекта.

В качестве «хорошего» критерия выбираем такой, при котором R будет иметь минимальное значение из всех возможных решений.

Оказывается, такому критерию опять-таки соответствует критериальная функция отношения правдоподобия (4.138) при

$$\varphi_0 = \frac{(\Pi_{01} - \Pi_{00})p}{(\Pi_{10} - \Pi_{11})q}. \quad (4.143)$$

¹ Если (4.140) не выполняется в конкретной задаче, гипотезы H_0 и H_1 можно поменять местами.

Если Π_{jk} не известны, то порог $\varphi_0 = p/q$ соответствует критерию **максимума апостериорной вероятности**. Если же известны платы, но не известны p и q , то «хорошим» критерием считается критерий **минимакса** – минимаксного правила выбора решений. Согласно ему оптимальным (наилучшим по этому критерию) считается решение, которому соответствует минимальное среди максимальных значений r^* условного риска (4.142), определенных при любых других правилах, т. е. $r^* = \min_{\gamma} \max_j r_j(\gamma)$. Как было показано А. Вальдом, такое решение является специальным случаем байесовского для наименее благоприятного априорного распределения вероятностей состояний S_0 и S_1 , т. е. для наименее благоприятных p и q .

Теперь рассмотрим примеры критериев проверки непараметрических гипотез, когда вид функций распределения $W(x)$ не известен. Это означает, что рассмотренные методы проверки параметрических гипотез не приемлемы и применяются *критерии согласия*.

В этом случае критериальная функция, во-первых, в большей степени зависит от сути проверяемой гипотезы: о распределении, о равенстве среднего или дисперсии ГС и т. п., во-вторых, распределение ее статистики $\varphi(\chi)$ является приближенным (в расчетах, как правило, асимптотическим при $N \rightarrow \infty$), зависящим от ряда предположений (условий) и существенно – от объема выборки N , особенно если N мало (см., например, [24]). Это может привести к дополнительным неточностям проверки гипотезы в связи с неточностью (в отличие от параметрических ситуаций, когда распределение точно соответствует выборке) нахождения α и β , даже если предполагаемое распределение точно соответствует выборке (!) (см. рис. 4.8, в предположении, что статистики $\varphi_1(\chi)$ и $\varphi_2(\chi)$ одинаковые, но получены при разных N_1 и $N_2 \gg N_1$).

Рассмотрим несколько примеров.

1. Проверка гипотезы о виде функции распределения

Имеется независимая однородная выборка $\chi = (x_1, \dots, x_N)$, принадлежащая ГС с неизвестным распределением (функцией $F(x)$ или плотностью $W(x)$). Необходимо проверить гипотезу H (или H_0) о том, что $F(x)$ (или $W(x)$) совпадает с заданной (гипотетической, т. е. определяемой гипотезой H , H_0) непрерывной функцией $F_0(x)$ ($W(x)$). Можно добавить: против альтернативы, что $F(x) \neq F_0(x)$ или

$F(x) > F_0(x)$ либо $F(x) < F_0(x)$. Как уже упоминалось, для проверки таких гипотез используются соответствующие критерии согласия. Рассмотрим некоторые из них, опуская, как и раньше, необходимые и достаточные условия, при которых они применимы, и напоминая о необходимости обращать на это внимание символом (!).

Критерий согласия Колмагорова (!). Для него критериальная статистика $\varphi(\chi)$ имеет вид

$$\varphi(\chi) = \sqrt{N} d_N = \sqrt{N} \sup_x |F_0(x) - F^*(x)|, \quad (4.144)$$

\sup_x – верхняя грань по всем значениям x разности функций; $F^*(x)$ – эмпирическая функция распределения

$$F^*(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(x - x_{(i)}) = \frac{1}{N} \sum_{i=1}^N \frac{N_{(x_{(i)} \leq x)}}{N}; \quad (4.145)$$

$\mathbf{1}(x - a)$ – функция единичного скачка, равная 0 при $x < a$ и 1 при $x \geq a$; $x_{(i)}$ – i -й элемент вариационного ряда $x_{(1)}, \dots, x_{(i)}, \dots, x_{(N)}$; $N_{(x_{(i)} \leq x)}$ – количество элементов выборки χ , значения которых не превышают x .

А.Н. Колмогоровым было показано, что если $F_0(x)$ непрерывна, то асимптотически функция распределения статистики d_N (4.144) не зависит от вида $F_0(x)$. Следовательно, вид $W_\Phi(\varphi)$ (рис. 4.8) асимптотически однозначно определен и получил название распределения Колмогорова [20, 24]. Его таблицы для расчета квантилей $x_{1-\alpha}$, принимаемых за φ_0 , табулированы для разных N и находятся во всех книгах по математической статистике. Заметим, что при малых $N < 20$ $W_\Phi(\varphi)$ отличается от асимптотического даже при проверке простых гипотез $H_0: F(x) = F_0(x; \theta_0)$. При проверке сложных гипотез $F(x) = F_0(x)$ при неизвестном параметре θ , оцениваемом по той же выборке χ , закон распределения статистики φ зависит еще и от метода оценивания параметра θ (см. примеры в [24]). Это необходимо учитывать при расчете $\varphi_0 = x_{1-\alpha}$.

Аналогичное справедливо для других критериев согласия, используемых для проверки гипотез и вида распределения. В частности для следующих.

Критерий Реньи (!)

$$\varphi(\chi) = \sqrt{N} d_N = \sqrt{N} \sup_x \left| \frac{F^*(x) - F_0(x)}{F_0(x)} \right|, \quad F_0(x) > 0; \quad (4.146)$$

$W_\Phi(\varphi)$ – асимптотическое распределение Реньи, φ_0 определяется как квантиль $x_{1-\alpha}$ этого распределения.

Критерий ω^2 Мизеса (!)

$$\varphi(\chi) = N \omega_N^2 = \frac{1}{12N} + \sum_{i=1}^N \left[F^*(x_{(i)}) - \frac{2i-1}{N} \right]^2; \quad (4.147)$$

$W_\Phi(\varphi)$ – асимптотически, если гипотеза H_0 верна, имеет распределение Мизеса–Смирнова; φ_0 есть квантиль $x_{1-\alpha}$ этого распределения.

Критерий хи-квадрат (!). Вся область значений x , где определена гипотетическая функция распределения $F_0(x)$, разбивается на конечное число n интервалов группирования одинаковой Δ или разной Δ_l длины, $l = 1, 2, \dots, n$. Размер Δ должен быть таким, чтобы в него попадало не менее $p_l N \geq 10$ моментов выборки, где $p_l = p_l(\theta)$ – априорная вероятность попадания выборочного значения x_i , $i = \overline{1, N}$, в интервал Δ_l ($p_1 + p_2 + \dots + p_n = 1$) при условии, что выборка имеет распределение $F(x) = F_0(x; \theta)$. Иначе закон $W_\Phi(\varphi)$ может сильно отличаться от рекомендуемого табличного представления. Тогда статистика $\varphi(\chi)$ имеет вид (ср. с (4.133))

$$\varphi(\chi) = \sum_{l=1}^n \frac{(N_l - N p_l(\theta))^2}{N p_l(\theta)} = N \sum_{l=1}^n \frac{(N_l/N - p_l(\theta))^2}{p_l(\theta)}, \quad (4.148)$$

где N_l – количество элементов выборки, значения которых попали в интервал Δ_l .

Если проверяется простая гипотеза $H_0: F(x) = F_0(x; \theta_0)$, т. е. выборка принадлежит параметрически заданному распределению

$F_0(x; \theta_0)$ с параметром $\theta = \theta_0$, то, как показал К. Пирсон, при определенных условиях (!) (в частности, если $p_l(\theta)$ гладкая функция по θ) число интервалов $n = \log_2 N$, то статистика $\varphi(\chi)$ (4.148) имеет асимптотически распределение хи-квадрат ([20], отсюда название критерия) с $(n - 1)$ -й степенью свободы независимо от вида гипотетического распределения $F_0(x)$ ¹.

А как быть, если проверяется гипотеза на соответствие выборки χ распределению $F_0(x; \theta)$ при k неизвестных параметрах θ ? Тогда параметры θ для распределения $F_0(x; \theta)$ можно оценить по той же выборке χ . Затем воспользоваться модификацией статистики (4.148), в которой вместо неизвестных θ взять значения оценок $\hat{\theta}$. Если для оценивания используется один из методов (МП, МХК и др.), обеспечивающих (хотя бы асимптотически) получение эффективных оценок θ , то в случае справедливости гипотезы H_0 распределение модифицированной статистики типа (4.148) будет также асимптотически хи-квадрат, но с числом степеней свободы $r = (n - k - 1)$.

Заметим, что для данного критерия вид распределения $W_{\Phi}(\varphi)$ для конечных N зависит не только от вида распределения выборки χ , объема выборки N , метода оценивания θ , но и от того, как будет произведено группирование, т. е. назначены Δ_l . Желаящие могут с этим ознакомиться подробнее по работе [24].

2. Проверка гипотез о принадлежности двух выборок одному и тому же распределению

Предположим, что имеются две независимые однородные выборки χ_1 объема N_1 и χ_2 объема N_2 , принадлежащие распределениям $F_1(x)$ и $F_2(x)$. Необходимо проверить гипотезу H_0 о том, что $F_1(x) = F_2(x)$ (можно добавить при альтернативе $H_1: F_1(x) \neq F_2(x)$).

¹ Если только гипотеза H_0 верна. Если на самом деле выборка принадлежит другому распределению $F_1(x; \theta_1)$, то асимптотическое распределение статистики (4.148) будет нецентральным хи-квадрат с $n - 1$ степенью свободы и параметром нецентральности, определяемом по (4.148), если в (4.148) заменить N_l/N на вероятность $p_l(\theta_1)$, определяемую по $F_1(x; \theta_1)$, а $p_l(\theta)$, как и ранее, определять по $F_0(x; \theta_0)$ (см. рис. 4.8).

В этом случае можно использовать модификации предыдущих критериев согласия. Например, аналог критерия Колмагорова – двухвыборочный критерий Колмагорова–Смирнова, критериальная функция $\varphi(\chi_1, \chi_2)$ которого имеет вид

$$\varphi(\chi_1, \chi_2) = d_{N_1, N_2} = \sup_x \left| F_1^*(x) - F_2^*(x) \right|, \quad (4.149)$$

асимптотически также имеет распределение Колмогорова.

3. Проверка гипотез о равенстве числовых характеристик

Это очень часто используемый класс проверяемых на практике параметрических и непараметрических гипотез. Сюда относятся параметрические гипотезы (как правило, в предположении, что выборки имеют нормальное распределение) о равенстве средних двух независимых однородных выборок из ГС с равными и неравными, известными и неизвестными дисперсиями; о равенстве дисперсий двух ГС при известных и неизвестных средних и т. д. Понятно, что критериальная функция $\varphi(\chi)$ и распределение $W_\varphi(\varphi)$ статистики φ при этом зависят как от вида гипотезы, так и от априорных сведений. При проверке гипотезы о равенстве нулю коэффициента корреляции ρ_{XY} величин X, Y в качестве статистики $\varphi(\chi)$ используется статистика

$$\varphi(\chi) = t = \frac{\hat{\rho}_{XY} \sqrt{N-2}}{\sqrt{1 - (\hat{\rho}_{XY})^2}}, \text{ где } \hat{\rho}_{XY} \text{ – эмпирический (либо выборочный)}$$

коэффициент корреляции.

Если X и Y имеют совместное нормальное распределение и парные выборки $(x_1, y_1), \dots, (x_N, y_N)$ независимы, то распределение $W_\varphi(\varphi)$ такой статистики при больших N хорошо аппроксимируется распределением Стьюдента с $N - 2$ степенями свободы (см. [20]) и критическое значение φ_0 определяется по нему как квантиль $t_{1-\varepsilon}$. Иными словами, если $t > t_{1-\alpha}$, то считается, что коэффициент корреляции ρ_{XY} значимо отличается от нуля. При других распределениях X, Y и проверяемых значениях ρ_{XY} плотность $W_\varphi(\varphi)$ будет другой. Иногда в этом случае используют статистику Фишера (!)

$$\varphi(\chi) = \frac{1}{2} \ln \frac{1 + \hat{\rho}_{XY}}{1 - \hat{\rho}_{XY}}, \hat{\rho} \neq 1, \quad (4.150)$$

которая уже при небольших N имеет распределение, хорошо аппроксимируемое нормальным с математическим ожиданием, равным

$$M\{\varphi\} = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(N-1)}, \text{ и дисперсией, равной } 1/(N-3).$$

4. Примеры других задач проверки гипотез

Из других задач, рассматриваемых в математической статистике, наиболее часто встречаются проверки гипотез на однородность выборки, средних и дисперсий; отклонения закона от нормального; независимости или случайности выборки; отсутствие тренда во временных рядах; наличие аномальных значений в выборке, а также гипотез о спектральных плотностях (см., например, [5, 24]).

Последовательные правила проверки гипотез

Отличительная особенность рассмотренных правил проверки гипотез – их одношаговость. До начала эксперимента задается число N опытов и затем по полученной выборке χ объема N осуществляется проверка гипотезы. Другой подход (последовательный) заключается в том, чтобы определять N по ходу экспериментирования. Для этого применяются последовательные многошаговые процедуры проверки гипотез, использующие уже два правила (критерия). Одно – правило остановки опытов (наблюдений), т. е. выбора текущего шага N , на котором надо остановиться. Второе, аналогичное уже рассмотренным, – это правило выбора решения γ после остановки опытов, наблюдений.

Для этого на оси γ назначается не одно критическое (пороговое) значение φ_0 , как в рассмотренных ранее критериях, а два: φ_1 и φ_2 . Решения же принимаются после каждого очередного опыта эксперимента, начиная с $N=1$. Они выбираются так, что, если при очередном N $\varphi(N) \leq \varphi_1$, то выбирается решение γ_0 (не отвергается гипотеза H_0). Если же $\varphi(N) \geq \varphi_2$, то принимается решение γ_1 , что не отвергается («подтверждается») конкурирующая гипотеза H_1 . Если же $\varphi_1 < \varphi(N) < \varphi_2$, то принимается промежуточное решение γ' – провести следующий опыт. Так продолжается до тех пор, пока не будет принято одно из двух решений γ_0 или γ_1 . Ясно, что при этом число опытов N будет случайной величиной. Следовательно, можно поставить вопрос о минимальной средней стоимости всего эксперимента из N опытов, т. е. об оптимальном выборе φ_1 и φ_2 .

Для уяснения сути этих правил рассмотрим критерий Вальда на примере проверки простой параметрической гипотезы против простой альтернативы, когда решение γ принимается на базе критерия отношения правдоподобия (4.138).

А. Вальд положил, что если стоимость эксперимента пропорциональна числу его опытов, т. е. N , то тогда ее средний минимум будет пропорционален среднему минимуму объема выборки N при условии, что уровень значимости не превышает α , а мощность не менее $1 - \beta$. Поскольку средние значения \bar{N} при справедливости гипотез H_0 или H_1 в общем случае не равны, требуется минимизация обеих величин (\bar{N}_0 и \bar{N}_1). Тогда в указанных условиях оптимальные φ_1 и φ_2 можно назначить из условий

$$\varphi_2 \leq \max\left(\frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\right). \quad (4.151)$$

$$\varphi_1 \geq \min\left(\frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\right); \quad (4.151a)$$

Ясно, что в последовательном правиле N может достигать больших величин без принятия решений γ_0 или γ_1 . Для исключения этой ситуации применяется **усеченное последовательное правило**. Его суть сводится к тому, что последовательное правило действует до $N < N_{\max}$. Если же N достигает значения $N = N_{\max}$ и ни одно из решений γ_0 или γ_1 по итогам N_{\max} -го опыта не принято, то устанавливается один порог φ_0 по обычным правилам проверки гипотез и решение принимается согласно этим правилам. Понятно, что при этом значения вероятностей $\alpha_{\text{ус}}$ и $\beta_{\text{ус}}$ ошибок I и II рода могут оказаться большими, чем заданные при последовательном правиле. Ошибочные решения могли бы не появиться, если бы опыты продолжались при $N > N_{\max}$. Также очевидно, что, если N_{\max} назначать согласно обычным непоследовательным правилам, то при использовании усеченных правил мы можем выиграть, если реальное N в эксперименте меньше N_{\max} .

Наконец рассмотрим **третий «рычаг»** управления критериями.

Ранее мы полагали, что принятие решений $\gamma_0, \gamma_1, \dots, \gamma_m$ при проверке гипотез осуществлялось по жесткому детерминированному алгоритму. Но ведь выбор можно рандомизировать. Это означает, что при попадании выборки χ или статистики φ в область, соответствующую гипотезе H_i , $i = \overline{0, m}$, допускается вместо выбора решения γ_i выбор любого другого с вероятностью p_{ki} , $i = \overline{0, m}$, $k = \overline{0, m}$, когда $p_{0i} + p_{1i} + \dots + p_{mi} = 1$. При соответствующих условиях такие правила могут дать более эффективные в определенном смысле решения. Напомним, что еще одна рандомизация может быть на этапе постановки и проведения эксперимента.

В заключение приведем обобщенный технологический процесс проверки гипотез. Он представлен на рис. 4.9.

С3. Статистические выводы

Этот раздел математической статистики связан с формализацией процедур получения выводов, заключений о свойствах, связях, закономерностях случайных элементов, их моделей, характеристик по их выборкам из генеральной совокупности. В приложении к исследованию физических объектов это означает использование формализованных статистических процедур для получения выводов о свойствах и особенностях исследуемых стохастических объектов, их строения и поведения. Поскольку аппарат статистических выводов базируется на теориях оценивания и проверки гипотез, отдельно его рассматривать не будем.

4.3.8. Аппарат статистической имитации случайных элементов. Метод Монте-Карло

Как уже указывалось, теория статистической имитации случайных элементов есть раздел стохастической математики, связанный с построением аппарата искусственного получения обладающих требуемыми свойствами последовательностей выборочных значений x_1, x_2, \dots, x_N объема N случайных элементов \aleph в следующих ситуациях:

А – по априори заданной модели \mathcal{M} случайного элемента \aleph (т. е. по вероятностной модели);

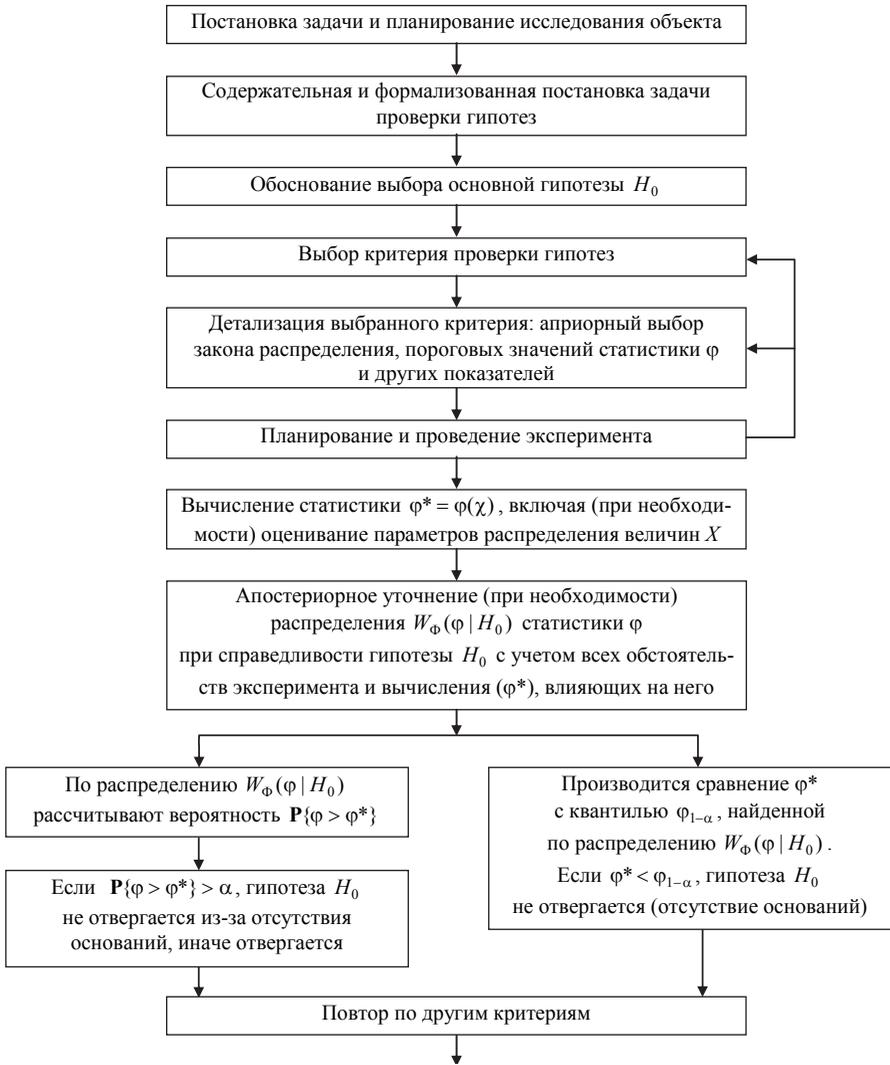


Рис. 4.9. Обобщенный технологический процесс одношаговой статистической проверки гипотез

В – по апостериори полученной модели $\hat{\mathcal{M}}$ случайного элемента \aleph (т. е. по статистической модели);

С – по ранее полученным выборочным (или приравненным к ним эмпирическим, экспериментальным) значениям y_1, \dots, y_N объема N_1 из генеральной совокупности случайного элемента \aleph , описываемого той же моделью. При этом свойства новой выборки могут совпадать со всеми свойствами исходной выборки y_1, \dots, y_N либо частично отличаться заданным образом в рамках изменения некоторых параметров модели (положения, масштаба или связи).

Остановимся только на основных идеях, используемых в данном аппарате для простейших ситуаций, а также на методе статистических испытаний (Монте-Карло), основа которого – имитация последовательностей реализации выборочных значений случайных величин и векторов.

Прежде всего обратим внимание на то, что мы ранее оперировали термином *статистическая имитация*, а теперь даже в подзаголовке конкретизировали *статистическая имитация случайных элементов*¹. Если быть более точным, то это имитация последовательностей реализаций случайных элементов или выборок (последовательностей выборочных значений) из генеральной совокупности «имитируемого» случайного элемента. Причем последовательность должна обладать требуемыми свойствами. Например, она должна быть аналогом случайной выборки либо иметь заданный одномерный закон распределения и корреляционную функцию (спектральную плотность мощности), либо заданный n -мерный закон распределения и т. д.

Дело в том, что часто термин *статистическая имитация* используется в качестве аналога термина *статистическое моделирование* или *статистическое имитационное моделирование*. Однако эти термины охватывают все этапы имитационного моделирования (см., например, рис. 2.3: стадия исследования объекта по модели). Мы же остановимся только на одном подэтапе 13-го этапа технологического процесса моделирования (рис. 2.3), касающегося генерации последовательностей выборок с заданными свойствами. Связь трех рассматриваемых аппаратов: вероятностного, статистического и части имитационного (генерация выборки) представлены на рис. 4.10 (см. также табл. 4.1).

¹ Событий, величин, векторов, функций и т. д.

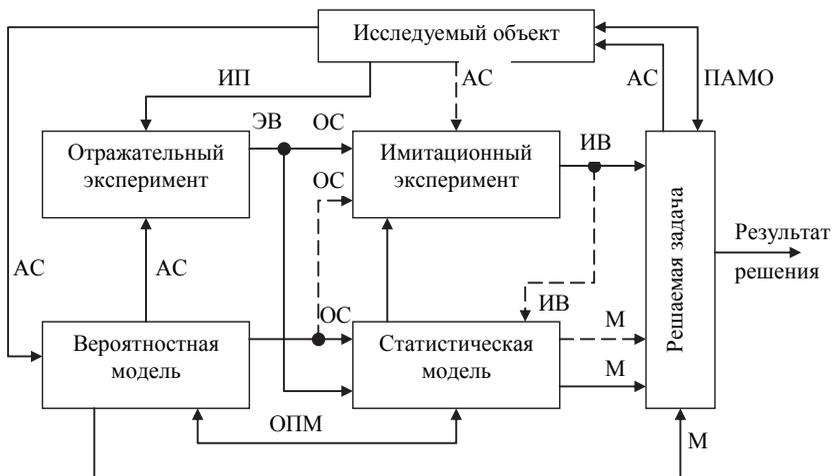


Рис. 4.10. Условная схема связи отражательного и имитационного экспериментов с вероятностно-статистическим модельным аппаратом и статистической имитацией (генерацией):

АС – априорные сведения; ИВ – имитационная (имитируемая, генерируемая) выборка; ИП – измеряемые показатели (состояния); М – модели; ОПМ – оценивание и проверка модели; ОС – отправные сведения; ПАМО – проверка адекватности модели объекту; ЭВ – эмпирическая выборка

Поэтому рассмотрим только вопросы имитации (генерирования) выборок с заданными свойствами. Поскольку в основе современных методов имитации разных случайных элементов на цифровой вычислительной машине (ЦВМ) лежат методы имитации случайных величин, рассмотрим прежде всего их, дополняя особенностями методов имитации случайных векторов и процессов.

Основные методы имитации выборок случайных величин можно разделить на три группы: физические, табличные случайных цифр и алгоритмическо-программных псевдослучайных чисел.

Физические методы основаны на использовании естественных или искусственных приборных, технических (аппаратных) генераторов «шумовых», флуктуационных сигналов. Флуктуационные – это сигналы, которые можно описать моделью некоторого случайного процесса, с дальнейшими линейным и нелинейным преобразованиями их для получения требуемых свойств. Методы их преобразований будут описаны далее. Первый вариант использования таких сигналов – непо-

средственное применение при натуральных имитационных экспериментах, а также для получения по ним последовательностей реализаций случайных величин, векторов, функций. Для этого они дискретизируются по времени и квантуются по уровню с помощью АЦП. Полученные при этом цифровые отсчеты и будут представлять собой элементы выборок случайных величин или векторов. Основное достоинство таких методов – получение выборок, элементы которых наиболее близко соответствуют понятию выборки из ГС случайных величин, векторов, процессов в силу отсутствия соответствующих свойств «детерминизма» в них (см. далее). Основные недостатки методов:

а) их малая приемлемость из-за чувствительности свойств генерируемых сигналов к изменениям внешних условий работы приборов и, как следствие, малая стабильность статистических свойств сигнала;

б) сложность их воспроизводимости, повторного генерирования той же выборки при необходимости повторения эксперимента, например в контрольных целях.

Второй вариант применения подобных генераторов – получение с их помощью сигналов с равномерными на интервале $[0, 1]$ или сегменте $(0, 1)$ распределениями значений. Он основан на использовании в качестве таких сигналов (или их отсчетов) *шумов квантования* – разности между текущим значением сигнала и его цифровым аналогом (см. рис. 3.5). Однако эти методы в настоящее время редко используются в связи с широким применением алгоритмических программных средств.

Метод таблиц случайных цифр (чисел). Суть его сводится к предварительному составлению N' заведомо случайных наборов цифр, из которых потом можно легко формировать случайные n -разрядные числа $\xi_i = 0, \varepsilon_i \dots \varepsilon_{i+n-1}$, взяв в качестве $\varepsilon_i, \dots, \varepsilon_{i+n-1}$ n очередных цифр из таблицы. Этот метод связан с трудностью формирования больших таблиц, с «законностью» многократного обращения к таблице, а также с возможным наличием в таблицах «плохих» участков. Поэтому он используется чаще всего в учебных задачах, а также при решении независимых задач, не требующих большого объема выборок.

Метод псевдослучайных чисел (ПСЧ). Он основывается на следующих четырех идеях.

Первая – в качестве базовой используется последовательность независимых (в вероятностном смысле) равномерно распределенных на

(0, 1) чисел. Вторая – из этой последовательности формируются n -мерные базовые последовательности для имитации n -мерных случайных векторов или векторных последовательностей (временных рядов), значения которых равномерно распределены в n -мерном единичном гиперкубе. Например, путем выбора n неповторяющихся элементов («отсчетов»), в частности n соседних отсчетов базовой последовательности как одного элемента n -мерной выборки. Третья – формирование из одно- или n -мерных базовых последовательностей новых последовательностей с заданными распределениями, корреляционно-спектральными свойствами, типами стационарности или нестационарности, операторного модельного представления и т. д. Наконец, основная, четвертая, идея, послужившая причиной наименования метода, – использование для получения элементов последовательности рекуррентных детерминированных формул вида

$$x_i = f(x_{i-1}, x_{i-2}, \dots, x_{i-k}), \quad i = k, k+1, \dots, \quad (4.152)$$

когда начальные значения x_0, x_1, \dots, x_{k-1} фиксируются.

1. Из (4.152) следует, что получаемая по этой формуле последовательность всегда будет имитировать только дискретные случайные величины, которые в лучшем случае будут ближе к сингулярным, нежели к абсолютно непрерывным. Тем не менее они часто используются как заменители последовательностей выборок из ГС с абсолютно непрерывными законами распределения.

2. Если мы имеем дело с теоретическими выборками из ГС n -мерных случайных векторов, например (X, Y) или (X, Y, Z) , с абсолютно непрерывными равномерными на (0, 1) законами распределения, то их значения будут равномерно распределены в единичном n -мерном гиперкубе (в квадрате $0 < x < 1, 0 < y < 1$ для (X, Y) или кубе $0 < x < 1, 0 < y < 1, 0 < z < 1$ для (X, Y, Z)), в то время как точки последовательности (4.152) будут лежать на кривой $y = \varphi(x)$ или $z = f(x, y)$ (рис. 4.11). Следовательно, «хорошую» (близкую к непрерывно распределенной случайной) последовательность чисел можно получить по (4.152), если функция $f(\cdot)$ и входящие в нее составные функции $\varphi(\cdot)$, связывающие $x_{i-1}, x_{i-2}, \dots, x_{i-k}$ между собой, как можно более плотно будут заполнять базовые квадрат, куб, k -мерный гиперкуб.

Этого можно достичь, например, если на рис. 4.11, б использовать не пять линий, а настолько много, насколько это возможно на компьютере.

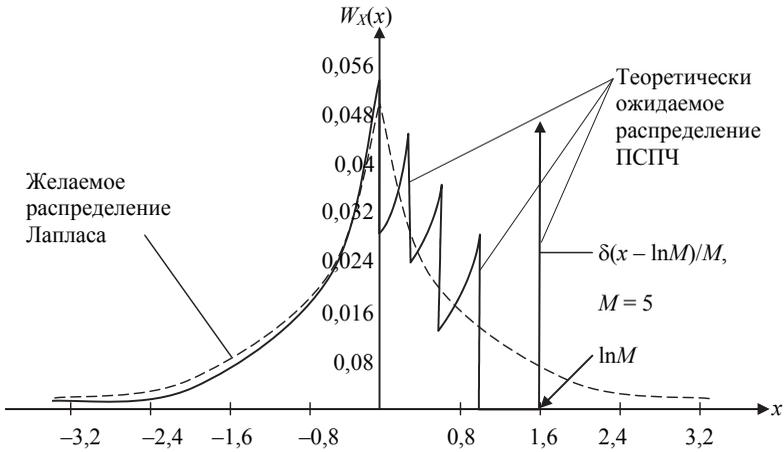
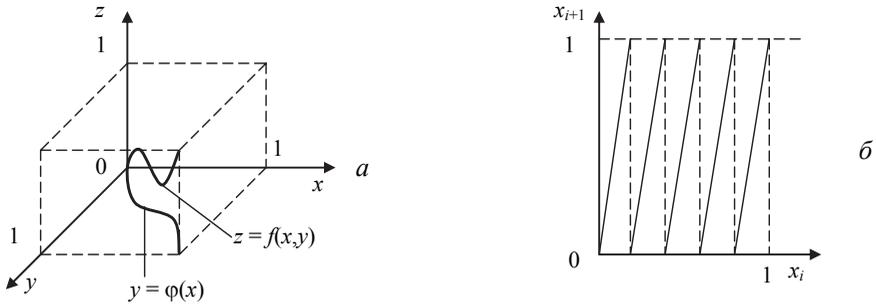


Рис. 4.11. Пояснение расположения точек в последовательности (4.152) (а, б) и плотности распределения псевдослучайных чисел X с ожидаемым распределением Лапласа (в)

3. Нетрудно убедиться, что использование рекуррентных формул типа (4.152) на ЦВМ всегда порождает периодические последовательности. Действительно, во-первых, на любой ЦВМ мы можем оперировать лишь конечным числом M чисел, заключенных между 0 и 1 (рис. 4.11, б). Поэтому рано или поздно текущее значение последовательности совпадет с одним из предшествующих и начнется повторная рекурсия в базовой последовательности. Понятно, что для увеличения длины периода L можно использовать рандомизацию (ослучивание) начальных значений, но это позволяет всего лишь частично решить проблему. Выход из положения – делать период L настолько большим,

насколько это возможно, чтобы при решении практических задач периодичность не успевала проявиться. Во-вторых, поскольку между соседними n числами последовательности существует своя зависимость $\varphi(\cdot)$, через определенное число элементов они могут повторяться чаще, чем две соседние пары. Это надо учитывать при имитации n -мерных векторов и временных рядов.

4. Поскольку соседние элементы базовой последовательности связаны рекуррентной зависимостью (см., например, связь соседних элементов базовой последовательности на рис. 4.11, *а, б*), они в принципе не могут быть независимыми. Это необходимо учитывать, когда имитируется одномерная последовательность (например, временной ряд), но при этом используются несколько соседних элементов базовой последовательности. На рис. 4.11, *в* представлена теоретическая плотность вероятностей ПСПЧ, получаемых по формуле $x_i = \ln \xi_{2i-1} - \ln \xi_{2i}^1$, $i = 1, 2, 3, \dots$, где ξ_i – элементы базовой равномерно распределенной ППСЧ, полученной по рекуррентной формуле (4.152) по методу вычетов (при $M = 5$), когда соседние элементы ее x_i и x_{i-1} связаны зависимостью рис. 4.11, *б*. Если бы ξ_{2i-1} и ξ_{2i} были независимы, то распределение x_i было бы лапласовским [20].

Именно в силу этих четырех особенностей получаемые таким образом числа называются *псевдослучайными*.

Опуская различные варианты выбора функций $f(\cdot)$, остановимся на наиболее часто используемом *конгруэнтном методе вычетов*. В его основе лежит рекуррентное соотношение

$$x_{i+k}^* = \sum_{j=0}^{k-1} a_j x_{i-j}^* + b \pmod{M}, \quad (4.153)$$

где множители a_j , приращение b и x_i^* – целые числа, а суммирование осуществляется по модулю M .

Напомним, что выполнение операций в (4.153) по модулю M показывает, что x_{i+k}^* получается как остаток от деления правой части на M . Это означает, что $x_{i+k}^* \in [0, M-1]$ и поэтому для получения x_i в от-

¹ Замена $x_i = \ln \xi_{2i-1} - \ln \xi_{2i}$ на $x_i = \ln(\xi_{2i-1}/\xi_{2i})$ нежелательна!

крытом интервале $(0, 1)$ необходимо выполнить преобразование $x_i = x_i^* / M$.

При реализации (4.153) в двоичной системе обычно $M = 2^n$, где n – количество двоичных разрядов, используемых для представления целого числа.

При $k=1$ и $b=0$ имеем *линейную мультипликативную форму* метода вычетов, а при $k=1$ и $b \neq 0$ – *линейную смешанную форму*.

Показано, что (4.153) при $k=1$ обеспечивает получение полного периода L последовательности тогда и только тогда:

1) если b является взаимно простым по отношению к M (т. е. оба являются простыми);

2) если b является простым числом, на которое делится M , то $a_0 - 1$ делится на b ;

3) если M делится на 4, то $a_0 - 1$ тоже делится на 4.

При $M = 2^n$ максимально возможный отрезок аperiodичности L_{\max} последовательности (4.153) при $k=1$ и $b=0$ равен 2^{n-2} , если x_0^* взаимно просто с 2^n (т. е. x_0^* – нечетное) и a_0 удовлетворяет сравнениям $a_0 \equiv 3 \pmod{8}$ или $a_0 \equiv 5 \pmod{8}$. Последнее условие означает, что младшая цифра восьмеричной записи множителя a_0 равна 3 или 5. Это выполняется, например, при $a_0 = 5^{2q+1}$, $q = 0, 1, 2, \dots$, или $a_0 = 2^l + 3$, $l = 3, 4, 5, \dots$ Работа на ЦВМ с последними множителями сводится к сдвигу и трем сложениям. Для последовательности же (4.153) при $k=1$ и $b \neq 0$ $L_{\max} = 2^n$ и не зависит от x_0^* . Максимальная длина последовательности при этом достигается при $a_0 \equiv 1 \pmod{4}$ и $b_0 \equiv 1 \pmod{2}$.

В настоящее время существуют десятки алгоритмов генерирования псевдослучайных последовательностей чисел (ПСПЧ). Желающие могут с многими из них и с результатами их тестирования на случайность и периодичность ознакомиться, воспользовавшись электронными ресурсами «Statistical Testing of Random Number Generators [электронный ресурс], URL: csrc.nist.gov/groups/ST/toolkit/rng/documents/nissc-paper.pdf; The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness [электронный ресурс], URL: <http://stat.fsu.edu/pub/diehard>; Dieharder: A Random Number Test Suite [электронный

ресурс], URL: <http://www.phy.duke.edu/~rgb/General/dieharder.php>. Заметим, что разработано множество тестов на проверку разных свойств генераторов псевдослучайных последовательностей. Простейшие из них связаны с проверкой параметрических гипотез (см. рис. 4.10) о распределениях отрезков аperiodичности, независимости элементов, равномерности заполнения областей n -гиперкуба и т. п. Сейчас эта работа для многих генераторов уже выполнена.

Итак, базовые последовательности имеют равномерное распределение. *Имитация по ним случайных событий* из полной группы H_1, H_2, \dots, H_m несовместных событий с вероятностями появления p_1, p_2, \dots, p_n сводится:

а) к делению интервала $(0,1)$ на m подынтервалов, длина каждого из которых равна вероятности p_l ;

б) к определению граничных точек полуинтервалов;

в) к выявлению факта попадания или непадения очередного сгенерированного числа x_i в s -й полуинтервал, соответствующий имитируемому событию H_s . Попадание в s -й полуинтервал соответствует появлению события H_s .

Для *имитации случайных величин* с распределениями, отличными от равномерного, используются различные нелинейные преобразования базовых равномерно распределенных последовательностей. При этом учитываются следующие правила.

1. Из базовых последовательностей обычно генерируются последовательности величин \tilde{X} со стандартными распределениями $F_{\tilde{X}}(x) = F_{\tilde{X}}(x; a=0, \lambda=1)$, $W_{\tilde{X}}(x) = (1/\lambda)W_X(x; a=0, \lambda=1)$, т. е. такими, которые имеют нулевое значение параметра положения (сдвига) $a=0$ и единичное значение параметра масштаба $\lambda=1$. Если имитации подлежит величина X с другими значениями параметров a и λ , то она получается из \tilde{X} преобразованием

$$X = a + \lambda \tilde{X}. \quad (4.154)$$

2. Если X имеет абсолютно непрерывное распределение $F_X(x; a, \lambda; \mathbf{\alpha})$, где a, λ и $\mathbf{\alpha}$ – параметры положения, масштаба и формы ($\mathbf{\alpha} = \alpha_1, \dots, \alpha_k$), то случайная величина

$$\Xi = F_X(X; a, \lambda, \mathbf{\alpha}) \quad (4.155)$$

всегда имеет равномерное на $(0,1)$ распределение. Верно и обратное утверждение: если Ξ имеет равномерное абсолютно непрерывное на $(0,1)$ распределение и $F_X(x; a, \lambda; \alpha)$ – абсолютно непрерывная функция распределения, то, подвергая Ξ обратному к $F_X(x)$ преобразованию $F^{(-1)}(\xi)$, мы получим величину X с распределением $F_X(x; a, \lambda, \alpha)$. На этом основан метод обратных функций имитации случайных величин X с любым абсолютно непрерывным распределением $F_X(x)$. Для этого базовую равномерно распределенную на $(0,1)$ псевдослучайную последовательность необходимо поэлементно подвергнуть обратному $F^{(-1)}(x)$ к функции распределения $F_X(x)$ преобразованию. Сложность реализации метода при априорном задании $F_X(x)$ связана со сложностью математического описания функций распределения $F_X(x)$. Однако метод может успешно использоваться при апостериорной имитации в ситуациях В и С (см. начало подпараграфа 4.3.8), когда вместо $F_X(x)$ используется ее оценка (например, ЭХ-оценка) и обратное преобразование делается по ней.

3. В разделе теории вероятностей, связанной с функциональными преобразованиями случайных величин, векторов и функций, разработаны правила получения законов распределения детерминированных функций от случайных элементов при известных законах распределения их случайных аргументов. В частности, имеется много примеров получения случайных величин Y и векторов $\mathbf{Y} = (Y_1, \dots, Y_n)$ с разными распределениями из независимых или зависимых случайных величин X_1, X_2, \dots, X_n путем их элементарных функциональных преобразований. Например, преобразований вида

$$\begin{aligned} Y_1 &= f_1(X_1, X_2, \dots, X_n); \\ Y_2 &= f_2(X_1, X_2, \dots, X_n); \\ &\vdots \\ Y_m &= f_m(X_1, X_2, \dots, X_n), \end{aligned} \tag{4.156}$$

где f – элементарные функции.

В [20] приведено более 140 таких преобразований для случайных величин и показано, что 60 часто используемых распределений

случайных величин получаются с помощью элементарных преобразований величин с другими известными распределениями.

Подобный метод имитации случайных элементов называется *методом элементарных функциональных преобразований*. Его идея: простыми способами генерируются случайные величины с одними распределениями, чтобы потом по имеющимся соотношениям типа (4.156) получить последовательности с другими распределениями. Следует иметь в виду, что из-за указанных ранее «неидеальностей» последовательностей псевдослучайных чисел (ППСЧ) как заменителей случайных выборок свойства последовательностей после преобразований (4.156) могут отличаться от ожидаемых. В частности, из-за наличия детерминированной связи пар, троек, n рядом стоящих в последовательности псевдослучайных чисел в последовательностях чисел Y могут иметь место искажения распределений, появиться корреляция между ними и т. д. Правда, при больших модулях M искажениями можно пренебречь. Пример таких искажений приведен на рис. 4.11, в.

На нем изображено реальное распределение величины X , получаемой по формуле $X = \ln \Xi_1 - \ln \Xi_2 = \ln(\Xi_1/\Xi_2)$, где Ξ_1 и Ξ_2 – равномерно распределенные на $(0,1)$ величины, зависимость между которыми изображена на рис. 4.11, б.

Особый интерес имеет имитация случайных величин с нормальным (К.Ф. Гаусса) распределением [20]. Разные методы их имитации представлены во многих программных системах. Это, например, методы, основанные на обратных функциях; на преобразованиях пары независимых, равномерных на $(0,1)$ случайных величин Ξ_1, Ξ_2 в пару независимых нормальных величин X_1, X_2 :

$$\left. \begin{aligned} X_1 &= \sqrt{-2 \ln \Xi_1} \cos(2\pi \Xi_2), \\ X_2 &= \sqrt{-2 \ln \Xi_1} \sin(2\pi \Xi_2) \end{aligned} \right\}; \quad (4.157)$$

на суммировании независимых величин Ξ_i , равномерно распределенных на $(0,1)$, по формуле

$$X = \sqrt{\frac{12}{n}} \left[\sum_{l=1}^n \Xi_l - \frac{n}{2} \right], \quad (4.158)$$

которая уже при $n = 12$ дает очень хорошее приближение распределения X к нормальному.

Теперь коротко об *имитации случайных векторов и временных рядов*. Прежде всего обратим внимание на три метода генерирования n -мерных базовых последовательностей псевдослучайных чисел.

Первый метод. В качестве имитируемой реализации равномерной случайной величины Z_i , входящей в n -мерный вектор ${}_n Z = (Z_1, \dots, Z_l, \dots, Z_n)$, $l = \overline{1, n}$, включаются очередные значения i -го независимого генератора ППСЧ. Этот метод особенно подходит для параллельных алгоритмов решения задач исследования объектов.

Второй метод. Используется nN чисел одной ППСЧ. При этом первые N чисел берутся в качестве выборочных значений величины X_1 , следующие N – величины X_2 и т. д. Ясно, что при практической реализации метода необходимо либо сгенерировать nN чисел, запомнить их и потом использовать для имитации n -мерного вектора (что не всегда желательно или приемлемо), либо после однократного генерирования определять последние числа предшествующих N чисел, которые являются начальными для первых чисел следующих ППСЧ из N чисел и т. д. Тем самым второй метод сведется к первому методу.

Третий метод (наиболее часто используемый). В качестве первых сгенерированных значений имитируемого n -мерного вектора $\mathbf{x}_1 = (x_{1,1}, x_{2,1}, \dots, x_{n,1})$ берутся первые n чисел генерируемой ППСЧ, в качестве значений $\mathbf{x}_2 = (x_{1,2}, x_{2,2}, \dots, x_{n,2})$ – следующие n чисел ППСЧ и т. д.

Понятно, что во всех методах предполагаются «независимость» чисел в ППСЧ и выполнение условия $L > N$ для первого и $L > nN$ для второго методов, где L – длина периода последовательности.

Для получения из равномерно распределенных базовых ППСЧ n -мерных ППСЧ пригодны все рассмотренные ранее методы. В частности, если величины X_1, X_2, \dots, X_n , входящие в вектор ${}_n \mathbf{X}$, независимы, то каждую величину X_l можно генерировать независимо. Если же элементы вектора ${}_n \mathbf{X}$ зависимы, то общий метод генерирования имитации таких векторов основывается на представлении n -мерного распределения вектора ${}_n \mathbf{X}$ через условные распределения

$$\begin{aligned} W_X(x_1, x_2, \dots, x_n) &= W_{X_1}(x_1) W_{X_2}(x_2 | x_1) \times \\ &\times W_{X_3}(x_3 | x_1, x_2) \dots W_{X_n}(x_n | x_1, \dots, x_{n-1}). \end{aligned} \quad (4.159)$$

Поэтому имитация вектора \mathbf{nX} сводится к имитации последовательностей «условных» случайных величин:

$$X_1; X'_2 = X_2 | (X_1 = x_1); X'_3 = X_3 | (X_1 = x_1) \cap (X_2 = x_2), \dots;$$

$$X'_n = X_n | (X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_{n-1} = x_{n-1}),$$

где $X'_l = X_l | A$ есть «условная» величина X_l при условии, что до ее появления имело место событие A ; x_l – конкретное значение, которое приняла величина X_l .

Для отдельных распределений векторов имеется множество более простых методов. Например, для устойчивых¹ распределений, куда относятся векторы с n -мерным нормальным распределением, имитируются текущие значения вектора $\mathbf{nX} = (X_1, X_2, \dots, X_n)$ независимых величин, которые затем подвергаются линейным функциональным преобразованиям, т. е. преобразованиям (4.156) при $m = n$ и

$$Y_i = a_i + \sum_{s=1}^n \alpha_{i,s} X_s. \text{ В результате имитируется } n\text{-мерный вектор } \mathbf{nY} \text{ с}$$

тем же законом распределения и требуемой корреляционной матрицей \mathbf{R} , по которой определяются коэффициенты $\alpha_{i,s}$.

Необходимо при этом всегда помнить, что задаваемая матрица \mathbf{R} должна быть именно корреляционной. Это означает, что она должна удовлетворять всем свойствам корреляционных матриц. В частности, быть *неотрицательно определенной, теплицевой*. Вы никогда не сгенерируете вектор с заданной матрицей \mathbf{R} , не являющейся корреляционной. Заметим также, что для негауссовых векторов на вид корреляционной матрицы накладываются другие ограничения(!).

В случае, если имитируемые распределения отличаются от устойчивых, используются функциональные преобразования (4.156) векторов с n -мерным равномерным, нормальным, экспоненциальным и другими известными, легко имитируемыми векторами с соответствующей корреляционной матрицей. Сделаем в связи с этим несколько замечаний.

¹ *Устойчивые* – это распределения, не меняющие свой вид при взвешенном суммировании независимых величин. Для устойчивых нормальных векторов из некоррелированности следует независимость (см. примеры в [5]).

1. В настоящее время довольно хорошо проработана теория и имеется алгоритмическое и программное обеспечение имитации нормальных (с n -мерным распределением Гаусса) n -мерных векторов с произвольной корреляционной матрицей. Поэтому случайные вектора с другими распределениями получаются преобразованиями нормальных векторов. Кстати, n -мерные векторы с коррелированными компонентами и с одномерными равномерными распределениями можно получить из n -мерного нормального вектора с зависимыми компонентами, если каждый l -й ($l = \overline{1, n}$) элемент нормального вектора подвергнуть преобразованию (4.155) l -го элемента нормального вектора через его функцию распределения. Таким образом можно из гауссовой коррелированной векторной ППСЧ получить базовую одномерно равномерно распределенную коррелированную векторную ППСЧ. С ней можно поступать далее так же, как при имитации по ней случайных величин с разными законами распределения. Получаемые при этом последовательности будут иметь требуемый закон распределения и будут коррелированными. Заметим, что есть множество других методов имитации равномерно распределенных векторов.

2. При преобразовании \mathbf{nX} с плотностями распределения $W_{\mathbf{X}}(x_1, \dots, x_n)$ ¹ или с одномерными плотностями $W_{X_l}(x_l)$ и корреляционными матрицами $\mathbf{R}_{\mathbf{X}}$ в n -мерный вектор \mathbf{nY} с плотностью $W_{\mathbf{Y}}(y_l)$ и корреляционной матрицей $\mathbf{R}_{\mathbf{Y}}$ по (4.156) необходимо решать (как правило, итерационно) следующие сильно связанные задачи.

А. Преобразование $W_{\mathbf{X}}(x)$ в $W_{\mathbf{Y}}(y)$ при заданном f или нахождение при требуемом $W_{\mathbf{Y}}(y)$ таких распределений $W_{\mathbf{X}}(x)$ и преобразования f , которые обеспечат получение этого распределения $W_{\mathbf{Y}}(y)$.

В. Пересчет корреляционной матрицы (КМ) $\mathbf{R}_{\mathbf{X}}$ в матрицу $\mathbf{R}_{\mathbf{Y}}$ или пересчет требуемой (желаемой) матрицы $\mathbf{R}_{\mathbf{Y}}$ в $\mathbf{R}_{\mathbf{X}}$ при выбранном по пункту А преобразовании f . Помимо того, что необходима проверка задаваемой матрицы $\mathbf{R}_{\mathbf{Y}}$ на ее принадлежность к корреляционным матрицам, должна быть проверка $\mathbf{R}_{\mathbf{X}}$ не только на

¹ В многомерном случае правила преобразования функций распределения \mathbf{X} напрямую в функции распределения \mathbf{Y} очень сложны в отличие от правил для плотностей распределения.

корреляционность – принадлежность к КМ, но и на совместимость ее с распределением $W_X(x)$. Дело в следующем. Во-первых, как указывалось, корреляция отражает лишь линейную статистическую связь, а здесь мы имеем дело с нелинейными преобразованиями. Поэтому для получения желаемой КМ R_Y с помощью преобразования f типа (4.156) вектора X будет иметь место нелинейная связь между компонентами векторов X и Y . Во-вторых, может быть много векторов X с тем же самым одномерным распределением, позволяющих при разных функциональных преобразованиях (4.156) получить те же одномерные распределения вектора Y , но с разными многомерными распределениями $W_X(x)$, т. е. с разными КМ R_X . Правильнее было бы находить n -мерное распределение вектора X по n -мерному распределению вектора Y с требуемыми одномерными распределениями $W_Y(y)$ и КМ R_Y . Подобная задача довольно сложная и имеет множество решений. Поэтому можно ставить задачу о поиске наилучшего в каком-то смысле варианта преобразования f и базового распределения $W_X(x)$. Во многих случаях удастся избежать этой проблемы, если вместо корреляционной матрицы использовать нормированную конкорреляционную [20]. Она сохраняет свои значения при любых взаимно однозначных монотонных преобразованиях и не зависит от одномерного закона распределения. Это позволяет независимо управлять преобразованием законов при сохранении конкорреляции.

3. Для решения многих задач может оказаться предпочтительным имитировать не вектор с заданным n -мерным распределением или одномерным распределением и корреляционной (КМ) либо конкорреляционной (ККМ) матрицей, а вектор с близкими к желаемому одномерными распределениями и КМ, ККМ, но проще генерируемыми ППСЧ.

4. Удобным инструментом при имитации случайных величин и векторов и их аппроксимации другими может оказаться метод моделирования [5, 20].

Наконец, несколько слов об *имитации случайных процессов*.

Прежде всего отметим, что на ЦВМ обязательно происходит замена процесса последовательностью, например временным рядом, $x(0), x(\Delta t), \dots, x[(N-1)\Delta t]$. Шаг дискретизации Δt при этом выбирается исходя из решаемой задачи (см. частично об этом во второй части учебного пособия).

Для имитации скалярных и векторных процессов пригодны (с соответствующими модификациями) рассмотренные методы имитации

случайных величин и векторов. Например, при небольших объемах N можно имитировать процессы $X(t)$ или $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_n(t))$ в виде последовательности их отсчетов как результатов имитации n -мерного вектора $\mathbf{X} = (X_1, X_2, \dots, X_n)$, полагая $X(0) = X_1$, $X(\Delta t) = X_2, \dots$, $X[(N-1)\Delta t] = X_{N-1}$ или $X_1(0) = X_1$, $X_2(i_2\Delta t) = X_2, \dots$, $X_n[(i_{n-1})\Delta t] = X_{N-1} \quad \forall i_j = \overline{0, N-1}, \quad j = \overline{1, n}$, используя имитацию по методу условных распределений. При этом вместо n -мерных распределений или одномерных распределений и корреляционных и конкорреляционных матриц чаще всего ограничиваются заданием одномерных плотностей распределения и корреляционных и конкорреляционных функций или спектров. При больших и переменных N более предпочтительны специальные рекуррентные алгоритмы имитации базовых $X(t)$ и $\mathbf{X}(t)$, имеющих одномерное или двумерное нормальное или равномерное распределение и заданные корреляционно- или конкорреляционно-спектральные характеристики (КСХ). Если необходимо имитировать случайные процессы с другими требуемыми одномерными или двумерными распределениями и КСХ, используют функциональное преобразование базовых последовательностей по типу тех, о которых упоминалось при рассмотрении имитации случайных величин и векторов. При этом не следует забывать об указанных для имитации векторов проблемах 1–4, заменяя в них КМ на КСХ. Рассмотрим, например, стационарные последовательности с дробно-рациональными спектральными плотностями мощностей [5](!):

$$S_{XX}(v) = \sigma_{\Xi}^2 |A(z)|^2 / |B(z)|^2, \quad (4.159)$$

где $z = e^{jv}$, $j = \sqrt{-1}$, e – основание натурального логарифма,

$$A(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \dots + \alpha_p z^p; \quad (4.160)$$

$$B(z) = \beta_0 + \beta_1 z + \beta_2 z^2 + \dots + \beta_q z^q. \quad (4.161)$$

Их можно имитировать как последовательность авторегрессии и скользящего среднего АРСС $_{p,q}$ порядка $p, q \geq 0$:

$$X(i\Delta t) = \varepsilon_0 + \sum_{k=1}^p \alpha_k X[(i-k)\Delta t] + \Xi(i\Delta t) - \sum_{l=1}^q \beta_l \Xi[(i-l)\Delta t], \quad (4.162)$$

где $\Xi(i\Delta t)$ – δ -коррелированная случайная последовательность, а соответствующее слагаемое в (4.162) отсутствует, если p или q равно нулю. Если $\Xi(i\Delta t)$ имеет нормальное распределение, то $X(i\Delta t)$ является нормальной гауссовой последовательностью. Связь КСХ-последовательности (4.162) с ее параметрами α и β можно найти, например, в [5]. Заметим, что не при всех α_k , $k = \overline{1, p}$, и β_l , $l = \overline{1, q}$, последовательность (4.162) является стационарной (!).

В литературе описано множество методов имитации стационарных случайных процессов и последовательностей, описываемых другими конкретными моделями.

Для *имитации нестационарных случайных процессов и последовательностей* можно использовать либо методы имитации их как отсчетов соответствующих N -мерных векторов, либо машинные генераторы стационарных случайных процессов, выходную последовательность которых затем подвергать дополнительным преобразованиям в зависимости от вида нестационарности. Например, добавлять к стационарной последовательности дополнительно последовательности отсчетов, имитирующих аддитивный $a(t)$ или мультипликативный $\lambda(t)$ детерминированный тренд, сезонные $s(t)$ или периодические компоненты и т. д. (см. пункт 4.6.3.1), вводить нестационарность (как и коррелированность) в $\Xi(i\Delta t)$, а также в параметры $\alpha_0, \alpha_1, \dots, \alpha_p$; $\beta_0, \beta_1, \dots, \beta_q$.

Ранее мы рассмотрели ситуацию А – имитация по априорным данным из описанных в начале § 4.3. Теперь несколько слов о ситуациях В и С – имитация по апостериорным данным. Первый вариант имитации случайных величин, векторов и процессов по апостериорным данным сводится к построению по ним вероятностной модели методами, рассмотренными в методах оценивания и проверки гипотез. Второй – имитация либо непосредственно по статистическим моделям (ситуация В, аналогично ситуации А), либо по специальным алгоритмам (ситуация С). В случае ситуации В имитация осуществляется так же, как в ситуации А с той лишь разницей, что вместо априорной вероятностной используется апостериорная статистическая модель. В ситуации С используются специализированные алгоритмы, описанные в специальной литературе. Заметим, что при имитации по апостериорным данным случайных векторов и процессов во многих случаях предпочтение

следует отдать замене корреляционно-спектральных характеристик на конкорреляционно-спектральные, определяемые непосредственно по выборке [5, 20].

Метод статистических испытаний (метод Монте-Карло)

Важным приложением статистической имитации является решение многих прикладных задач методом статистических испытаний, часто называемых по известному «игорному» городу методом Монте-Карло. Поясним его суть на примере численного вычисления интегралов вида (!)¹:

$$I_1 = \int_a^b f(x) dx; \quad (4.163)$$

$$I_n = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_1 \dots dx_n, \quad (4.163a)$$

где $f(\cdot)$ – детерминированная функция; a и b принимают конечные или бесконечные ($a = -\infty$, $b = \infty$) значения.

Рассмотрим для конкретности одномерный случай (4.163). Представим I в обобщенном виде (!)

$$I_1 = \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \frac{f(x)}{W(x)} W(x) dx = \int_{-\infty}^{\infty} \varphi(x) W(x) dx, \quad (4.164)$$

где $W(x)$ – такая плотность распределения некоторой легко имитируемой случайной величины, что $W(x) \neq 0$ при $f(x) \neq 0$. Согласно определению оператор усреднения по вероятностной мере (оператор математического ожидания² $\mathbf{M}\{\cdot\}$) (4.164) можно представить в виде

$$I_1 = \mathbf{M}\{\varphi(X)\} \approx \frac{1}{N} \sum_{i=1}^N \varphi(x_i), \quad (4.165)$$

¹ Полагается, что необходимые условия существования и сходимости соответствующих интегралов и приводимых далее соотношений выполняются.

² Это означает, что математическое ожидание случайной величины $Y = \varphi(X)$ существует и конечно.

где x_i – реализация случайной величины с плотностью распределения $W(x)$. Первая часть (4.165) представляет собой ЭХ-оценку математического ожидания величины $Y = \varphi(X)$ по выборочным данным величины X . Если она состоятельна, то сумма в (4.165) будет тем ближе к левой части, чем больше N . Это означает, что точность нахождения I_1 определяется (при идеальных реализации преобразования φ и выборки x_i из ГС величины X) только объемом выборки N . Это же справедливо и для (4.163а) в отличие от численного интегрирования по методам кусочно-линейных аппроксимаций функций $f(\cdot)$, когда точность зависит от качества аппроксимации, кратности интеграла и т. п.

Понятно, что если $f(\cdot)$ задана на конечном интервале $[a, b]$ или гиперпараллелепипеда $\mathbf{A} = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$, то (4.163) можно представить в виде

$$I_1 = (b - a) \mathbf{M}\{f(X)\} \approx \frac{1}{N} \sum_{i=1}^N f(x_i); \quad (4.166)$$

$$\begin{aligned} I_n &= \prod_{i=1}^n (b_i - a_i) \mathbf{M}\{f(X_1, X_2, \dots, X_n)\} \approx \\ &\approx \frac{1}{N} \sum_{i=1}^N f(x_{1,i}, x_{2,i}, \dots, x_{n,i}), \end{aligned} \quad (4.166a)$$

где X равномерно распределена на $[a, b]$, а \mathbf{X} – в гиперпараллелепипеде \mathbf{A} .

Из (4.165) и (4.166) видно, что для нахождения I_1, I_n можно вместо выборочных значений x_i использовать значения сгенерированных ППСЧ, имитирующих случайную величину X с распределением $W(x)$ для (4.165) или равномерно распределенный на \mathbf{A} независимый вектор \mathbf{X} для (4.166), (4.166a) и вычисления ЭХ-оценок согласно правым частям (4.165) и (4.166), (4.166a) по таким ППСЧ. Отсюда ясно требование равномерности заполнения значениями ППСЧ интервала $[a, b]$ или гиперпараллелепипеда \mathbf{A} .

В связи с изложенным сделаем два замечания.

1. Вместо генераторов ППСЧ для вычислений по (4.166) можно использовать генераторы ЛП_r-последовательностей Соболя [27], точки которых равномерно заполняют многомерные кубы.

2. Близким к методу Монте-Карло является нахождение значений интегралов I_1, I_N с использованием *квазислучайных чисел*. Это такие неслучайные числа x_i , которые гарантируют сходимость правых частей (4.166), (4.166а) к левым не по вероятности или в среднем квадратическом, как в методе Монте-Карло, а в обычном детерминированном смысле сходимости численного решения к аналитическому.

Интересно отметить, что применение квазислучайных чисел может обеспечить сходимость правой части в (4.166), (4.166а) к левой с ошибкой порядка $1/N^{1-\varepsilon}$, где $\varepsilon > 0$ как угодно мало, в отличие от $1/\sqrt{N}$ сходимости в статистических методах Монте-Карло.

§ 4.4. Другие разновидности формального аппарата описания и исследования объектов

4.4.1. Вводные замечания

Описанный в предыдущих параграфах формальный аппарат может быть эффективно использован, если выполняются условия его применимости. Коротко они сводились к следующим.

1. Возможно многократное воспроизведение экспериментов в допустимо сходных условиях.

2. Наличие законов или закономерностей в исходах опытов экспериментов в тех же условиях.

3. Множество A элементов исходов опытов является канторовским (по фамилии автора их теории Георга Кантора). Канторовские множества отличает то, что они имеют бинарную характеристическую функцию (индикатор исходов x , функцию $\mu_A(x)$ принадлежности элемента x множеству A), которая может принимать только два значения: 0 или 1, т. е. $\mu_A(x) = \mathbf{I}_A(x) = 1$, если x принадлежит множеству A , или 0, если x не принадлежит множеству A (см. индикатор $\mathbf{I}(A')$ события A' в стохастическом аппарате).

4. Предполагается, что параметрическая модель, используемая для идентификации объекта или результатов опыта, эксперимента, имеет априори неизвестные параметры, которые надо определить по итогам

опыта. В частности, в модели измерения таким параметром считается сама измеряемая физическая величина в количественных измерениях или состояние объекта в категориальных (качественных) измерениях. И хотя значения параметра априори не известно, оно стабильно, постоянно, является «константой».

5. Используется точечное представление результата измерения.

6. Аналитически и графически модели могут быть полностью представлены в соответствующем многомерном евклидовом пространстве целой размерности: одномерном (прямая, числовая ось), двумерном (плоскость), трехмерном x , y , z и т. д.

А как быть, если хотя бы одно из этих условий не выполняется? Ниже очень коротко рассматриваются разработанные, в основном в последнее столетие, варианты формального аппарата, пригодные в этих случаях.

4.4.2. Байесовский стохастический аппарат

Предположим, что выполняются условия применимости стохастического аппарата, в частности 1, 2. Но при этом не выполняется условие 4. Иными словами, неизвестный оцениваемый (в математической статистике) или измеримый (при измерениях) скалярный θ или векторный θ параметр модели $\mathcal{M}(\theta)$ (например, какой-либо ее функциональной характеристики: закона распределения, корреляционной функции, спектральной плотности мощности) не является постоянной величиной (константой). Например, рост человека, значение его артериального давления и т. д. Допустим также, что по отношению к нему выполняются условия стохастичности. Тогда для его идентификации, а следовательно, и всей модели $\mathcal{M}(\theta)$, можно применить байесовский стохастический аппарат. Его идея основывается на формуле Тóмаса Бáйеса, опубликованной через два года после смерти ученого. Рассмотрим два зависимых случайных события A и B . Событие $C = A \cup B$ называется *объединением событий* A и B , если оно происходит тогда, когда происходит хотя бы одно из событий A или B , т. е. либо A , либо B , либо оба они вместе. Например, событие A – выпадание «орла» при бросании одной монеты, B – выпадание «орла» при бросании второй монеты или повторно той же монеты. Тогда событие C – появление «орла» при бросании двух монет или дважды одной монеты.

Событие $D = A \cap B$, состоящее в совместном осуществлении событий A и B , называется пересечением событий A и B . Например, D – появление «орла» два раза при бросании двух монет. Это пример независимых событий, для которых вероятность $P(D) = P(A)P(B)$. Заметим, что пересечение независимых событий, т. е. таких, для которых $P(D) = P(A)P(B)$, называется произведением этих событий и записывается в виде $D = AB$ или $D = A \cdot B$. Второй пример – зависимые события. Допустим, что в урне содержатся 14 шаров: 5 красных, 6 желтых и 3 белых. Событие A – извлечение красного шара, B – желтого, а C – извлечение подряд двух красных шаров в двух попытках. Пусть A_1 – событие, связанное с извлечением шара в первой попытке (первом опыте), а A_2 – во второй. Тогда вероятность¹ $P(A_1) = 5/14$, а вот вторая вероятность $P(A_2)$ зависит от условия – какой шар красный (К), желтый (Ж) или белый (Б) будет извлечен в первой попытке, а именно $P(A_2/K_1) = 4/13$, $P(A_2/Ж_1) = P(A_2/Б_1) = 5/13$. Вероятности $P(A_2/K_1)$, $P(A_2/Ж_1)$ и $P(A_2/Б_1)$ называются **условными**. Интересующее нас событие D есть не что иное, как пересечение $D = A_1 \cap A_2$. Следовательно

$$P(D) = P(A_1 \cap A_2) = P(A_1)P(A_2/A_1). \quad (4.167a)$$

Рассматривая две попытки извлечения шара как один опыт, нетрудно убедиться, что справедливо соотношение²

$$P(D) = P(A_1)P(A_2/A_1) = P(A_2)P(A_1/A_2), \quad (4.167б)$$

¹ Все вероятности $P(A)$ рассчитывались по правилу подсчета вероятностей в «классической» схеме статистических испытаний (опытов) [21] как отношение числа n_A исходов опыта, благоприятствующих появлению события A (в данном случае красного шара), к общему числу n равновероятных (равновозможных) исходов опытов (у нас $n_1 = 14$, а $n_2 = 13$ – один шар был извлечен и не возвращен в урну). Равновозможность исходов обеспечивается тщательным перемешиванием (рандомизацией) шаров в урне перед каждым опытом.

² Вероятность $P(A_2)$ может быть найдена либо по «классическому» правилу, либо по формуле полной вероятности [21]. Попробуйте найти $P(A_2)$ самостоятельно.

которое является частным случаем проявления правила подсчета вероятностей для пересечения двух событий

$$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B). \quad (4.168)$$

На этих равенствах и основана идея байесовских методов. По своей сути, по отношению к рассматриваемым опытам, вероятность $P(A_1/A_2)$ является апостериорной по отношению к первому опыту. Ведь она отвечает на следующий возникающий вопрос. Проведены два опыта и в результате второго опыта произошло событие A_2 – извлечен красный шар. Какова вероятность, что в первом опыте ему предшествовало извлечение также красного шара, т. е. имело место событие A_1 ? Или какова вероятность того, что появлению события A_2 предшествовало событие A_1 ? Ответ на вопрос о значении этой апостериорной вероятности и дает формула Байеса

$$P(A_1/A_2) = \frac{P(A_1)P(A_2/A_1)}{P(A_2)}; \quad P(A/B) = \frac{P(A)P(B/A)}{P(B)}. \quad (4.169)$$

Чаще всего эта формула связывается с формулой полной вероятности.

Пусть некоторое событие A может произойти только с одним из событий H (их принято называть гипотезами) H_1, \dots, H_n , образующих полную группу несовместных (несовместимых) событий, т. е. таких, что для их априорных вероятностей $P(H_i)$, $i = \overline{1, n}$, выполняется равенство $P(H_1) + P(H_2) + \dots + P(H_n) = 1^1$. Тогда $P(A)$ находится по формуле полной вероятности

$$P(A) = \sum_{i=1}^n P(H_i)P(A/H_i). \quad (4.170)$$

Пусть далее опыт проведен и в результате случилось событие B (либо A , либо противоположное ему событие \bar{A}). Тогда вероятность того, что событию B предшествовало в опыте появление события (верна гипотеза) H_i , $i = \overline{1, n}$, может быть определена по формуле Байеса (апостериорной проверки гипотез), принимающей вид

¹ Напомним, что всегда $0 \leq P(A) \leq 1$.

$$P(H_i/B) = \frac{P(H_i \cap B)}{P(B)} = \frac{P(H_i)P(B/H_i)}{P(B)}, \quad (4.171)$$

где $P(B)$ находится по (4.170) при $B = A$ или по (4.170) с заменой A на \bar{A} , если $B = \bar{A}$. Именно на идеях апостериорности и соотношениях (4.169), (4.171) базируются байесовские стохастические методы. Один из примеров их применения мы уже фактически рассмотрели. А именно, до опыта определяется или задается априорная вероятность $P(A_1)$, которая затем уточняется заменой на апостериорную $P(A_1/A_2)$ или $P(A_1/\bar{A}_2)$, когда в ходе второго опыта оказалось, что имело место событие A_2 либо противоположное ему \bar{A}_2 (второй шар не красный)¹.

Теперь рассмотрим байесовский подход к оцениванию скалярного параметра θ модели $\mathcal{M}(\theta)$ в предположении, что условие 4 не выполняется. У нас есть основание считать, что его можно представить случайной величиной Ξ с априорным распределением, например плотностью $W_{\Xi}(\theta)$, или его средним значением m_{Ξ} , СКО σ_{Ξ} и т. п. Априорные сведения о распределении параметра Ξ , его среднем значении и других характеристиках могут быть получены на основании предыстории функционирования объекта (процесса, сигнала), профессиональных теоретических соображениях о нем, из опыта решения подобных задач ранее и т. д. Эти сведения отражают степень уверенности субъекта исследователя в том, что параметр θ примет то или иное значение еще до начала опыта, эксперимента, т. е. до начала сбора исходных для исследования данных.

¹ Попробуйте применить этот подход к делению добычи (убитого кабана или доли прибыли) по итогам опыта. Например, в такой задаче. Два охотника A и B идут вместе охотиться на кабана. Охотник A более опытный и вероятность поражения им движущейся цели $P(A) > P(B)$. На охоте кабан был убит одной пулей (событие C), т. е. либо только охотником A (события A, \bar{B}), либо B (события \bar{A}, B). Как справедливее разделить тушу между охотниками: а) по априорным вероятностям $P(A)$ и $P(B)$ (например, в пропорциях от их суммы) или б) по апостериорным вероятностям $P(A/C)$ и $P(B/C)$? Заметим, что $C = (A \cap \bar{B}) \cup (\bar{A} \cap B)$.

Проводим N опытов, в результате которых получаем N значений x_1, x_2, \dots, x_N исследуемой величины X .

Составляем функцию псевдоподобия¹ $L(\cdot)$

$$L(x_1, x_2, \dots, x_N | \theta) = \prod_{i=1}^N W_X(x_i | \theta), \quad (4.172)$$

где x_1, x_2, \dots, x_N – значения X , полученные в i -м опыте, $i = \overline{1, N}$; $W_X(x_i | \theta)$ – плотность распределения вероятностей величины X в предположении (при условии), что значение параметра Ξ равно θ ; Π – символ произведения. После этого находим апостериорную плотность распределения параметра Ξ по модификации формулы Байеса для непрерывных плотностей распределения:

$$W_{\Xi}(\theta | x_1, x_2, \dots, x_N) = \frac{W_{\Xi}(\theta) L(x_1, \dots, x_N | \theta)}{\int_{-\infty}^{\infty} L(x_1, \dots, x_N | \theta) W_{\Xi}(\theta) d\theta}. \quad (4.173)$$

Теперь по апостериорной плотности $W_{\Xi}(\theta | x_1, x_2, \dots, x_N)$ в качестве апостериорной оценки, т. е. приближенного значения θ , можно использовать апостериорное среднее (математическое ожидание), моду, медиану, а также апостериорное СКО и другие характеристики параметра θ . Методы байесовского оценивания характеристик случайных элементов приведены в [5].

Как видим, здесь опять все свелось к построению стохастической модели величины X , когда выполняются условия 1–3, но не выполняется условие 4.

В тех случаях, когда априорная вероятность не может быть объективно обусловлена, вместо нее вводят субъективные вероятности и действуют так же, как только что было описано. Желаящие могут с этим ознакомиться самостоятельно. Байесовский подход широко применяется на практике: в статистической теории связи, при измерениях и т. п.

¹ Предполагается, что выборка x_1, \dots, x_N однородна, случайна, независима.

4.4.3. Аппарат нечетких (размытых) множеств

Теперь предположим, что не выполняется условие 3, т. е. само построение множества и его подмножеств является размытым, нечетко определенным. В качестве примера рассмотрим множество теплых дней в году и некоторый день со средней суточной температурой $+10^\circ\text{C}$. Зададимся вопросом: «Можно ли его однозначно отнести к этому множеству?» Добавим к этому несколько условий: это январский или июльский день; солнечный или пасмурный; сухой или влажный, сильно влажный; тихий или ветренный, сильно ветренный и т. д. Ясно, что в зависимости от этих условий, а тем более, их комбинаций, наш ответ будет меняться. Подобные множества получили название нечетких (размытых, туманных, пушистых, ... – *англ.* fuzzy set).

Нечеткое множество A на универсальном множестве U есть пара $(U, \mu_A(u))$, где $u \in U$ – элементы множества; $\mu_A(U)$ – функция принадлежности элемента $u \in U$ множеству A , т. е. функция, характеризующая степень принадлежности u к A . Это понятие в 1965 году ввел Лотфи Заде [1], расширив канторовское понятие множества допущением, что его характеристическая функция (функция принадлежности элемента $u \in U$, где U – универсальное множество) может принимать любое значение в интервале $[0,1]$, а не только значение 0 или 1, как в канторовском. Например, $U = X$ – универсальное множество всех возможных значений некоторой рассматриваемой физической величины, а $u = x$ – конкретное значение из этого множества. Заметим, что если выполняются условия 1, 2, то в качестве функции принадлежности можно взять вероятностную меру $P_A(\omega)$ распределения элементов ω в A вдоль всех элементов множества (см. табл. 4.2). Такого типа нечеткие множества можно рассматривать и называть их вероятностными или стохастическими, статистическими нечеткими множествами. Однако при этом не следует отождествлять понятие вероятности $P(A)$ события A с функцией принадлежности $P_A(u)$. Пример табличного представления нечеткого множества представлен в табл. 4.5².

¹ Заметим, что введенное понятие нечетких множеств само является строгим, абсолютно четким и не содержит никакой неточности, нечеткости.

² В отличие от ряда распределения вероятностей сумма $\mu_1 + \mu_2 + \dots + \mu_n$ не обязана равняться единице.

Табличное представление нечеткого множества X

u_i	u_1	u_2	...	u_i	...	u_n
$\mu_A(u_i)$	μ_1	μ_2		μ_i		μ_n

Пример функции принадлежности представлен на рис. 4.12.

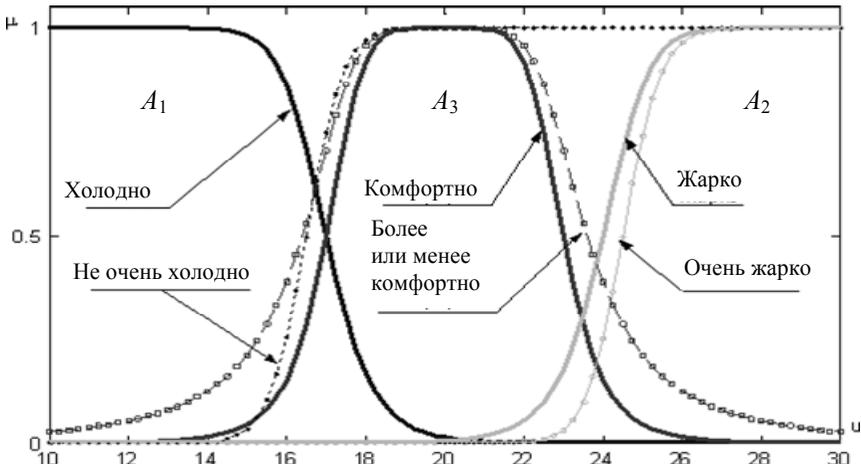


Рис. 4.12. Примеры графического представления функции $\mu_A(u)$ принадлежности температур $u = t$ к разным нечетким множествам:

A_1 – множество холодных; A_2 – жарких; A_3 – комфортных для человека температур

Чаще всего предполагается, что функция принадлежности выбирается субъективно, исходя из личного и общественного опыта решения задачи. Множество всех рассматриваемых элементов u нечеткого множества называется универсальным множеством и обозначается U .

Как следует из рис. 4.12, одна и та же температура, т. е. один и тот же элемент $u = t$ множества из множества температур t от 10 до 30° С, может одновременно с разной степенью принадлежности относиться к разным нечетким множествам: A_1 , A_2 , A_3 . В этом и проявляется нечеткость, размытость нечетких множеств U : A_1 , A_2 , A_3 .

Заметим, что нечеткие множества A_1 , A_2 , A_3 не есть просто подмножества значений температур из диапазона $t \in T = [10, 30^\circ \text{C}]$. Они

представляют собой именно пары этих подмножеств и функций принадлежности значений температур к соответствующему нечеткому множеству, а именно

$$A_1 = \{t, \mu_{A_1}(t)\}; \quad A_2 = \{t, \mu_{A_2}(t)\}.$$

Это отличие легко усваивается на примере дискретных нечетких множеств. Предположим, что универсальное множество U состоит из конечного количества элементов $U = \{u_1, u_2, \dots, u_n\}$. Тогда нечеткое множество A можно представить в табличном виде (по типу ряда распределения, см. табл. 4.5), где $\mu_i = \mu_A(u_i)$, либо в аналитическом виде¹:

$$A = \sum_{i=1}^n \mu_A(u_i)/u_i = \mu_1/u_1 + \mu_2/u_2 + \dots + \mu_n/u_n. \quad (4.174)$$

Например, нечеткое множество «экзаменационная оценка» по пяти- (четырёх-) балльной системе можно представить в виде

$$A = 0,1/2 + 0,3/3 + 1/4 + 0,7/5. \quad (4.175)$$

Для нечетких множеств по аналогии с канторовскими, в частности с множествами элементарных случайных событий, для которых аналогом универсального множества U является полное множество Ω элементарных событий ω (см. табл. 4.2), введены и изучены разные операции, отношения (сравнения) и свойства. Например, считается, что A содержится в B , если для любого u из U (пишется $\forall u \in U$) $\mu_A(u) \leq \mu_B(u)$. Множество A является равным B (пишется $A = B$), если $\mu_A(u) = \mu_B(u) \quad \forall u \in U$.

Объединением $C = A \cup B$ нечетких множеств A и B называется их наименьшее нечеткое подмножество, содержащее элементы A или B . При этом²

$$\mu_{A \cup B}(u) = \max \{ \mu_A(u), \mu_B(u) \} \quad \forall u \in U, \quad (4.176)$$

$$\text{т. е. } \mu_{A \cup B}(u) = \mu_A(u) \cup \mu_B(u).$$

¹ Косая черта «/» носит разделительную функцию. Ее нельзя воспринимать как символ (знак) деления чисел.

² В (4.176) и (4.177) использована наиболее употребляемая минимаксная интерпретация логических операций дизъюнкции (\cup) и конъюнкции (\cap).

Пересечением $D = A \cap B$ **нечетких множеств** называется их наименьшее нечеткое подмножество, содержащееся одновременно в A и B . Для него

$$\mu_{A \cap B}(u) = \min \{ \mu_A(u), \mu_B(u) \} \quad \forall u \in U, \quad (4.177)$$

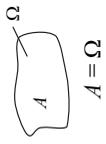
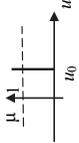
$$\text{т. е. } \mu_{A \cap B}(u) = \mu_A(u) \cap \mu_B(u).$$

Отрицанием (дополнением) множества A называется множество \bar{A} с функцией принадлежности $\mu_{\bar{A}}(u) = 1 - \mu_A(u)$. Сходство и различие между этими операциями в теории нечетких множеств и подобными им в теории случайных событий, а также различие между мерой принадлежности $\mu_A(u)$ и мерой возможности появления события A – вероятностью $P(A)$ – наглядно видно из табл. 4.6.

Как следует из изложенного и продемонстрированного в табл. 4.6, граница между двумя нечеткими множествами A и B является размытой, нечеткой (отсюда и название), а переход элементов из одного множества в другое происходит плавно, без скачков. В канторовской же теории множеств (см. колонки 1–3 в табл. 4.6) этот переход осуществляется скачкообразно, поскольку оба множества имеют четкую границу между собой.

Наличие размытых границ между нечеткими множествами обычно интерпретируется в лингвистической форме введением лингвистических переменных. **Лингвистической переменной** (linguistic variable) называется переменная, значениями которой могут быть слова или словосочетания некоторого естественного или искусственного языка. Множество всех возможных значений лингвистической переменной называется **терм-множеством** (term set), элементы которого называются **термами** (term). Обратимся в качестве примера к рис. 4.12. На рисунке отражена лингвистическая переменная «температура в каком-то контексте» (например, «температура в комнате», «температура на даче», «температура в спортзале» и т. п.). Терм-множество этой переменной на рис. 4.12 состоит из термов – лингвистических оценок «холодно», «не очень холодно», «комфортно», «более или менее комфортно», «жарко», «очень жарко», а универсальное множество $U = T$ есть множество U температур $10^\circ\text{C} \leq t \leq 30^\circ\text{C}$, т. е. $U = [10, 30^\circ\text{C}]$.

Сооставление понятий «множество случайных событий» и «нечеткое множество»

Множество случайных событий		Нечеткое множество			
Операция, свойство	Графическое представление	Свойства вероятностей $P(\omega), P(A)$	Операция, свойство	Графическое представление	Свойства функций принадлежности $\mu_A(u)$
Принадлежность элемента ω событию A		$\mu_A(\omega) = \mathbf{I}_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A; \\ 0, & \text{если } \omega \notin A; \end{cases}$ $P(\omega) \in [0, 1],$ $P(A) \in [0, 1]$	Принадлежность элемента $u \in U$ множеству A		$\mu_A(u) \in [0, 1],$ если $u \in U,$ $\mu_A(u) = 0,$ если $u \notin U$
Достоверное событие A		$P(A) = 1;$ $P(\Omega) = 1$	Унимодалное множество		$\mu_A(u) = \begin{cases} 1, & \text{если } u = u_0, \\ 0, & \text{если } u \neq u_0 \end{cases}$
Невозможное событие (пустое множество событий)	$A = \emptyset$	$P(A) = 0;$ $P(\emptyset) = 0$	Пустое нечеткое множество		$\mu_A(u) = 0 \forall u \in U$
Пересечение событий $D = A \cap B$		$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)$	Пересечение нечетких множеств $D = A \cap B$		$\mu_{A \cap B}(u) = \min\{\mu_A(u), \mu_B(u)\}$

Окончание табл. 4.6

Операция, свойство	Графическое представление	Свойства вероятностей $P(\omega), P(A)$	Операция, свойство	Графическое представление	Свойства функций принадлежности $\mu_A(u)$
Произведение событий $F = AB = A \cdot B$	A и B – независимые события	$P(AB) = P(A)P(B)$	Произведение нечетких множеств $F = AB = A \cdot B$		$\mu_{A \cap B}(u) = \mu_A(u) \mu_B(u)$
Объединение событий $C = A \cup B$		$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	Объединение нечетких множеств $C = A \cup B$		$\mu_{A \cup B}(u) = \max\{\mu_A(u), \mu_B(u)\}$
Сумма событий (объединение несовместных событий) $C = A + B = A \cup B$, если $P(A \cap B) = 0$		$P(A + B) = P(A) + P(B)$	Сумма нечетких множеств $C = A + B$		$\mu_{A+B}(u) = \mu_A(u) + \mu_B(u)$
Дополнительное к A событие \bar{A}		$P(\bar{A}) = 1 - P(A)$ $P(A \cap \bar{A}) = 0$	Отрицание (дополнение) \bar{A} множества A		$\mu_{\bar{A}}(u) = 1 - \mu_A(u)$

По аналогии со случайными множествами и многомерными случайными величинами вводятся понятия *многомерных (n-мерных) нечетких множеств A*:

$$A = \{u_1, u_2, \dots, u_n; \mu_A(u_1, u_2, \dots, u_n)\}, \quad (4.178)$$

$$u_1 \in U_1, u_2 \in U_2, u_n \in U_n; (u_1, u_2, \dots, u_n) \in U_1 \times U_2 \times \dots \times U_n.$$

Описанные нечеткие множества относятся к множествам первого рода. Их отличие в том, что $\mu_A(u)$ является четко определенной, детерминированной функцией. В последние годы разрабатывается теория *нечетких множеств* второго и в более общем виде *n-го* рода. Для них предполагается, что функция $\mu_A(u)$, т. е. ее значения для некоторого конкретного u , нечетко определена. Тогда нечеткое множество A n -го рода представляет собой пару

$$A = \{u, \mu_{(n+1)}(\mu_n(\dots(\mu_1(u))\dots))\}, \quad (4.179)$$

где $\mu_1(u) = \mu_A(u)$ – нечеткая функция принадлежности $u \in U$ множеству A ; $\mu_2(\cdot), \dots, \mu_n(\cdot)$ – нечеткие функции принадлежности U к A на уровне функции $\mu_2(\cdot), \dots, \mu_{n-1}(\cdot)$; а $\mu_{n+1}(\cdot)$ – четкая функция принадлежности $\mu_n(\cdot)$.

В последние годы активно изучаются вероятностные нечеткие множества, для которых

$$\mu_A(u) = P\{\mu_1(u) \in [0, 1]\}, \quad u \in U, \quad (4.180)$$

где $P\{\}$ – вероятность того, что для конкретного u случайная величина $\mu_1(u)$ степени принадлежности u к A принимает значение из промежутка $[0; 1]$.

Следует заметить, что в настоящее время нечеткий аппарат развит как альтернатива детерминированному и стохастическому не только на уровне описания статики объекта, нечеткой логики выводов и принятия решений. Изучаются нечеткие множества, в которых функция принадлежности задается на булевых переменных, решетках, сетях. Имеют место нечеткие постановки и модели других традиционных разделов математики: статистическая проверка гипотез, регрессионный анализ, марковские процессы, дифференциальные уравнения и др.

Нечеткие постановки, аппарат нечетких множеств, логик и выводов могут быть использованы не только на этапе модельного представления объекта, но и на других этапах исследования объекта, организации и проведения эксперимента, обработки Данных и т. д. Ведь размытость наших представлений часто имеет место на этапе постановки задачи, планирования исследования и эксперимента в широком смысле, в частности при выборе исследуемых показателей объекта, используемых моделей и характеристик сходного назначения, способа представления результатов, выборе критериев качества и т. д. Поэтому можно ожидать, что аппарат нечеткого формального описания и исследования объектов в ближайшее время найдет свое применение и развитие в более широкой сфере. Станет ли он полноценной альтернативой стохастическому аппарату при исследовании динамики процессов и временных рядов, в частности корреляционному, регрессионному, спектральному, покажет время.

4.4.4. Интервальный аппарат

Ранее мы рассматривали формальный аппарат описания объектов и исходов экспериментов в предположении, что выполняется пятое условие из указанных в подпараграфе 4.4.1, т. е. точечное представление численных результатов (параметров объекта, результатов измерений, количественных исходов экспериментов). Нашу неуверенность в корректности, адекватности, достоверности точечного представления количественных данных от опыта к опыту мы пытались описать стохастическими или нечеткими моделями. В первом, стохастическом, случае точечный результат заменялся множеством, накрывающем ожидаемое точечное значение, с определенным на нем распределением вероятностей (в теории вероятностей) или частотей (в статистике). Во втором, нечетком, случае точечный результат заменялся также множеством, но с задаваемой субъектом функцией принадлежности точечных значений интересующему нас подмножеству. Как уже указывалось в § 3.3 (см. рис. 3.4), существует интервальная форма представления данных, когда физическая величина или результат ее измерения представляется интервалом $A = [a_1, a_2]$ ее значений, где a_1 – нижняя граница (начало), а a_2 – верхняя граница (конец) интервала A , т. е. всегда $a_1 \leq a_2$. Понятно, что далее с интервалом A можно работать в рамках детерминированного, стохастического или нечеткого подходов.

В детерминированном подходе интервалы A, B, C являются обобщенной моделью вещественных чисел a, b, c , поскольку при $a_1 = a_2 = a \quad A = a$, т. е. интервал представляет собой вещественное число. Иногда и сами интервалы A, B, C называют интервальными числами, для которых вводится интервальная арифметика, аналогичная обычной. В частности, вводятся следующие операции:

- сложение $[a, b] + [c, d] = [a + c, b + d]$;
- вычитание $[a, b] - [c, d] = [a - d, b - c]$;
- умножение $[a, b][c, d] = [a, b] \times [c, d] = [\min(ac, ad, bc, bd); \max(ac, ad, bc, bd)]$;
- деление $[a, b] / [c, d] = [a, b] : [c, d] = [\min(a/c, a/d, b/c, b/d); \max(a/c, a/d, b/c, b/d)]$.

При этом оговаривается, что операция деления определена только для таких случаев, когда интервал-делитель не содержит нуля.

Для интервалов A, B, C справедливы свойства коммутативности и ассоциативности операций сложения и умножения. Однако дистрибутивное свойство имеет место только в ослабленном виде, а именно $A(B + C) \subset AB + AC$.

Для интервалов в общем случае отсутствуют обратные элементы, так как $A / A \neq 1, A - A \neq 0$. Иными словами для интервалов в общем случае вычитание не обратно сложению, а деление не обратно умножению. Для интервалов нельзя, в общем случае, приводить подобные члены, так как они «работают» только со значениями, а не с именами переменных. Здесь имеется в виду то, что одной и той же буквой A могут обозначаться разные интервалы. Если же речь идет об одном и том же интервале, то $A - A = 0, A \cdot A = A^2 = [a_1^2, a_2^2], A / A = 1$. Единственные нейтральные элементы интервального сложения и умножения – это ноль $0 = [0, 0]$ и единица $1 = [1, 1]$. Два интервала $A = [a_1, a_2]$ и $B = [b_1, b_2]$ равны тогда и только тогда, когда они заданы на одном и том же множестве элементов a, b и $a_1 = b_1$ и $a_2 = b_2$. Разность $a_2 - a_1$ называется шириной интервала A и обозначается в виде

$$r(A) = a_2 - a_1, \quad r \geq 0. \quad (4.181)$$

Расстояние $d(A, B)$ между интервалами A и B определяется метрикой Хаусдорфа

$$d(A, B) = \max \{ |a_1 - b_1|, |a_2 - b_2| \}. \quad (4.182)$$

Интервальный анализ связан с исследованием различных математических объектов, изучаемых в вещественном анализе. В частности, в исследовании функций одной или нескольких переменных, когда переменными являются вещественные интервалы, например

$$Y = (y_1, y_2) = f(\mathbf{A}) = f(A_1, A_2, \dots, A_n), \quad (4.183)$$

где

$$f(A_1, A_2, \dots, A_n) = f\left([a_{1,1}, a_{1,2}], [a_{2,1}, a_{2,2}], \dots, [a_{n,1}, a_{n,2}] \right), \quad (4.184)$$

$$y_1 = \min_{A \in \mathbf{A}} f(\mathbf{A}), \quad y_2 = \max_{A \in \mathbf{A}} f(\mathbf{A}),$$

т. е. y_1 и y_2 определяются как минимальное и максимальное значения функции $y = f(x_1, x_2, \dots, x_n)$ из всех, которые она принимает, когда $x_1 \in A_1, x_2 \in A_2, \dots, x_n \in A_n$.

В настоящее время существуют многочисленные обобщения интервального анализа на элементы теории вероятностей и нечетких множеств, в частности на регрессионный анализ, статистическую проверку гипотез и т. д. Кстати, если вернуться к теории измерений (см. § 3.5), то согласно определению неопределенности результатов измерений типа B чаще всего в качестве априорного распределения значений результата принимают равномерное на интервале $A = [a_1, a_2]$ либо нечеткое с равномерной функцией принадлежности $\mu_A(x) = 1$ для всех $x \in [a_1; a_2]$ и $\mu_A(x) = 0$ для других x . Тем самым мы фактически заменяем точечный результат измерения интервальным.

Заметим, что интервальный анализ в детерминированной, стохастической или нечеткой формах реализации его еще не нашел такого внимания и применения, как точечные методы представления, обработки и анализа Данных и исследования объектов. Сравнительные примеры его применения будут рассмотрены во второй части учебного пособия и в практикуме.

4.4.5. Аппарат теории катастроф, динамического ха́оса и самоорганизации

Если объект можно описать законовыми детерминированными моделями, то по ним можно предсказать строго достоверно поведение объекта на сколь угодно длительное время (смена фаз Луны, движение планет и т. д.). В стохастических объектах, в том числе смешанных, описываемых квазидетерминированными моделями, основанными на детерминированных моделях со случайными параметрами, также возможно предсказание, но не однозначное, а вероятностное, достоверность которого падает с увеличением лага (интервала) предсказания. Для чисто случайных моделей этот лаг зависит от длительности корреляционной, конкорреляционной или дисперсионной функции. Если она представляет собой δ -функцию, т. е. если два соседних несовпадающих отсчета $X(t_1)$ и $X(t_2)$ или $X(t_1)$ и $Y(t_2)$ при $\forall t_1 \neq t_2$ некоррелированы, то лаг предсказания нулевой. Тем более, это предсказание невозможно для абсолютно случайных стохастических процессов (типа броуновского движения), для которых $X(t_1)$ и $X(t_2)$ при любых $t_1 \neq t_2$ независимы.

Однако существует еще один класс объектов (см. рис. 3.3), которые детерминированно изменяются во времени, но имеют неустойчивые траектории изменений. Эта неустойчивость может быть обусловлена неточностью задания (описания) начального состояния объекта, окружающей среды или его параметров и быстро непредсказуемо проявляется со временем. Например, когда в фазовом пространстве $(x(t), \dot{x}(t) = dx(t)/dt)$, где $x(t)$ – некоторый характерный показатель состояния системы, все (или почти все) соседние траектории внутри некоторой области разбегаются или сбегаются; все они остаются внутри ограниченного фазового объема. Тогда это является предпосылкой к применению аппарата описываемого в данном параграфе. Довольно часто исследуемые объекты характеризуются тем, что их структура или поведение в определенных условиях, моментах развития, при плавном количественном изменении внешних условий или собственных параметров резко, непредсказуемо, скачкообразно качественно изменяются. Для исследования таких объектов используется отдельный детерминированный аппарат теории катастроф или смесь детерминированного и стохастического аппаратов теории динамического ха́оса и самоорганизации.

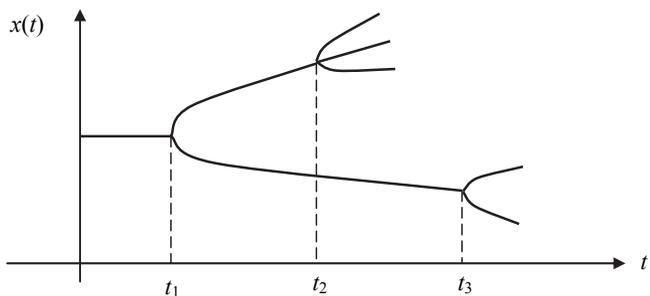
История теории катастроф начинается с работ французского математика Рене Тома (René Thom, 1958 г.). Базируясь на исследованиях Жюлья Анри Пуанкаре (Jules Henri Poincaré), Александра Александровича Аронова по теории бифуркаций (раздвоения – перестройки качественной картины движения) динамических систем при изменении их параметров и Хасслера Уитни (Hasler Whitney) по теории особенностей – обобщения исследований гладких функций на экстремумы (максимумы и минимумы), он объединил оба этих аппарата и предложил применять объединенный аппарат для описания **катастроф** – резкого качественного изменения поведения объектов, рассматриваемых как системы, при плавном количественном изменении переменных и параметров, от которых зависит их поведение. Эти идеи поддержал и начал пропагандировать английский математик Кристофер Зиман (Christopher Zeeman). В последующем этими вопросами занялся наш соотечественник Владимир Игоревич Арнольд.

В теории катастроф рассматривается поведение гладких (т. е. дифференцируемых много раз) функций в области критических экстремальных точек (**точек релетиции**), в которых не только первая производная равна нулю, но и производные более высоких порядков. Используя разложение функций в этих точках в ряды Тейлора и исследуя динамику развития таких точек при малых изменениях входных параметров, Р. Том исследовал геометрию бифуркаций на фазовой плоскости. Напомним, что **фазовое пространство** есть математическое многомерное пространство, координатами которого служат независимые параметры, определяющие состояние системы или мгновенные состояния системы по рассматриваемым показателям. **Фазовая плоскость** – двумерное фазовое пространство. Его координатами могут быть: значения внешнего воздействия на систему и реакции системы на него; значение показателя системы и его производной; предшествующее (или настоящее) и настоящее (последующее) во времени мгновенное значение показателя состояния системы; количество хищников и жертв и т. п.

Последовательность возможных состояний системы в фазовом пространстве, образующая траекторию ее эволюции, называется **фазовым портретом**, или **фазовой кривой**.

Точка на фазовом портрете отражает (задает) состояние системы. Приложенный в этой точке вектор указывает скорость изменения системы. Если вектор обращается в нуль, то это означает, что система находится в состоянии равновесия. Точки с нулевой скоростью изме-

нения системы называются *положениями равновесия*. Еще А. Пуанкаре показал, что фазовые кривые вблизи точки равновесия образуют картины, изображенные на рис. 4.13.



t_1, t_2, t_3 – точки бифуркации

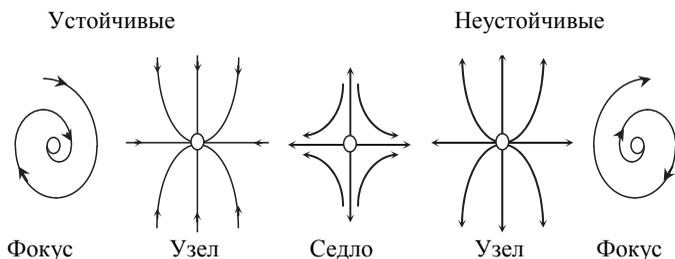


Рис. 4.13. Примеры бифуркаций

Как оказалось, даже более общие случаи поведения фазовых картин в окрестности положения равновесия на фазовой плоскости превращаются в картины, изображенные на рис. 4.13, при общем малом изменении системы.

Если система находится в устойчивом состоянии равновесия, то при небольших отклонениях от него за счет внутренних или внешних изменений ее параметров и воздействий она самостоятельно возвращается в устойчивое положение.

Другое дело, если состояние равновесия неустойчивое. Тогда система при малейшем изменении параметров и воздействий переходит в неустойчивое положение и, самоорганизуясь, может перейти в два

типа новых установившихся режимов. Первый, при *мягкой* потере устойчивости, это колебательный режим, который может стать новым состоянием равновесия. Этому состоянию соответствует замкнутая кривая на фазовой плоскости.

Второй, *жесткий*, вид потери устойчивости связан с тем, что область притяжения в точке равновесия очень мала и всегда имеющие место силы возмущения и флуктуации параметров¹ выбрасывают систему из этой области до того, как область притяжения полностью исчезнет. Тогда система уходит со своего стационарного режима скачком, разрушаясь, либо переходит на иной режим движения, состояния. Представьте себе качалку или ваньку-встаньку с перемещающимся центром тяжести либо велосипедиста с переменной скоростью движения. При достижении соответствующего положения центра тяжести качалка, ванька-встанька скачком опрокинется, упадет на бок, а при снижении скорости велосипедист не сможет удержать равновесие. Именно скачкообразный вид потери устойчивости изучает теория катастроф. Как уже упоминалось, скачкообразный переход от стационарного режима может привести к таким ситуациям, когда система перейдет в новые режимы. Они могут быть другими устойчивыми стационарными режимами, устойчивыми колебаниями или более сложными движениями.

Понимая под *катастрофой* скачкообразные изменения, возникающие в виде внезапного ответа объекта (системы) на плавное изменение окружающих условий и исследуя динамику развития объектов вблизи точек репетиции при малых изменениях активных переменных и параметров, от которых зависит устойчивость объектов, теория катастроф выделяет те области фазового пространства, в которых не просто складывается случайный узор, а формируются структурированные области стабильности. Такие точки рассматриваются как образующие центры для особых геометрических структур с низким и высоким уровнем катастрофичности в окружающих их областях фазового пространства. Р. Томом доказано, что если функция, описывающая состояние объекта, зависит от трех или менее активных переменных и пяти или менее активных параметров, то существует всего семь обобщенных структур оптимальных геометрических бифуркаций вблизи таких точек бифуркаций – репетиций. Это структуры типа *складка*, *сборка*,

¹ Флуктуации величины (параметра) – индетерминированные отклонения мгновенных значений величины (параметра) от нормального (или среднего при случайных флуктуациях) уровня.

ласточкин хвост, бабочка, гиперболическая («кошелек»), эллиптическая («пирамида») и параболическая омбилики. Пример одной из таких функций представлен на рис. 4.14. Желающие могут ознакомиться с графическими образами других структур самостоятельно с помощью Интернета.

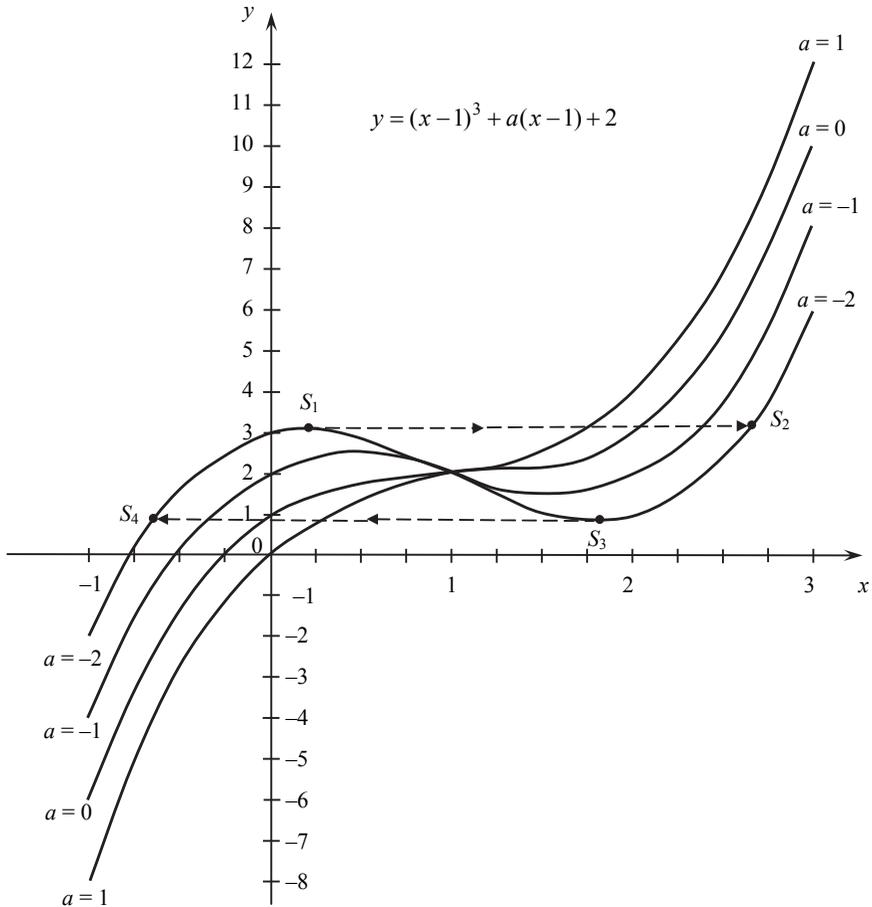


Рис. 4.14. Развитие катастрофы типа «складка» при изменении параметра a около $a = 0$:

складки начинают образовываться при переходе от $a > 0$ к $a < 0$; S – возможные скачкообразно измененные состояния

Теория катастроф является эффективным инструментом изучения и предсказания поведения сложных объектов, рассматриваемых как система, в практических ситуациях, т. е. в ситуациях, когда количественные изменения параметров объекта приводят к качественным изменениям его состояния. В частности, ее результаты позволяют предсказывать возможности появления и типы катастроф таких объектов путем исследования наличия и проявления их предвестников. Например, обнаружением и выявлением причин появления и нарастания таких новых особенностей поведения объекта, как «дребезжание», «дрожь», вибрация, автоколебания вокруг точки равновесия, которые, если ничего не предпринимать, могут привести к резкому, скачкообразному изменению состояния и даже к разрушению, разваливанию объекта. Важно обнаружить появление и проанализировать развитие следующих обычно встречающихся в совокупности **признаков (флагов) катастроф** – особенностей поведения объекта (системы), по которым принимается решение о том, что объект находится в предкатастрофическом состоянии. Вот некоторые из них (см., например, [29]).

- Переход **мульти-modalности** объекта в **бимодальность**. Мульти-modalность – это потенциальная возможность иметь несколько устойчивых состояний. В случае, если нахождение объекта в текущем состоянии становится малоустойчивым и в ближайшем будущем невозможным и все более проявляется бимодальность, начинает доминировать один вариант будущего устойчивого состояния, к тому же наименее желательный, возникает предпосылка катастрофы.

- Наличие **гистерезиса** – необратимость процесса поведения системы, невозможность вернуться к прежним условиям, состояниям, пойти по тому же пути развития, если убрать причины, послужившие отправной точкой для нахождения объекта в текущем состоянии.

- Приближение **расходимости** – продвижение объекта к такому неустойчивому состоянию, из которого он выводится слабыми возмущениями, «толчками» и способен быстро перейти в другие состояния при малых возмущениях, изменениях внешних и внутренних факторов, которые могут сопровождаться усилением шумовых флуктуаций параметров объекта вблизи точки «катастрофы» и быстро исчезать после ее прохождения.

- **Критическое замедление** ритмов функционирования, временной динамики объекта и реакций на изменение внешних условий. Критическое замедление – есть удлинение шкалы, возрастание времени релаксации объекта – постоянного возвращения его в состояние равно-

весия. Крайний случай – приближение «паралича» управления или самоорганизации объекта, когда множество усилий и разных затрат не приводит к сколько-нибудь заметному изменению ситуации или когда внешнее управление либо объект «не принимают» нужного решения, затягивают с его принятием.

Развитие теории *самоорганизованной критичности* привело к появлению дополнительных флагов, например, таких:

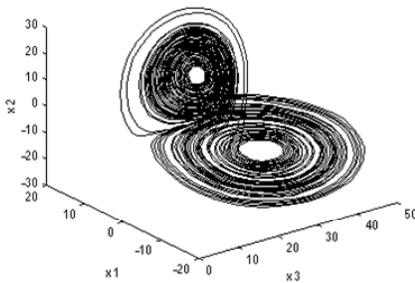
- *группировка* малых нежелательных событий перед катастрофическими (например, малых колебаний почвы перед землетрясением);
- *наведение «синергетических аварий»*, когда нестабильность одной части сложной системы многократно усиливается в другой (других), приводя к катастрофе.

Установившиеся режимы движения системы, ее состояния и структуры получили название *аттракторов* (от лат. *attrahere* – притягивать), так как они как бы «притягивают» все множество соседних режимов (переходных процессов, «траекторий» движения системы или ее структур в фазовом пространстве), определяемых различными начальными условиями. Название произошло от того, что если система попадает в область притяжения аттрактора, то она неизбежно эволюционирует к этому устойчивому режиму, состоянию, структуре. Заметим, что чаще *под аттрактором понимаются не сами режимы, состояния, структуры, а их изображение в фазовом пространстве.*

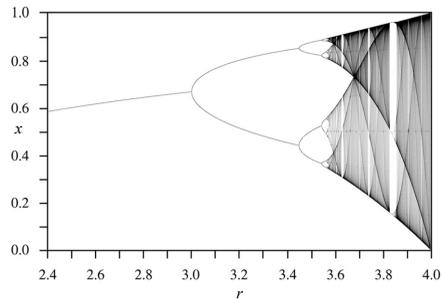
Виды аттракторов, отличных от характерных для равновесных и строго периодических колебательных режимов, состояний, структур, называют *странными* (термин ввели Д. Рюэль и Ф. Такенс, 1971 г.). Среди них особенно часто рассматривают такие, для которых, во-первых, фазовые портреты представляют собой некоторую ограниченную область в пространстве состояний системы, по которой происходят «случайные» или «хаотические блуждания»; во-вторых, характерно наличие *горизонта прогноза* – характерного времени, на которое поведение системы может быть предсказано (см. рис. 3.3). Примеры аттракторов приведены на рис. 4.13 и 4.15.

Странные аттракторы отражают установление хаотического поведения в динамических системах. Они были открыты в 1963 г. американским метеорологом Э. Лоренцем. Слово «странность» отражает две их особенности. Первая связана с их чувствительностью к малым изменениям начальных условий состояния системы. Дело в том, что если система является нелинейной детерминированной и ее аттрактор представляет собой точку (см. рис. 4.13) или цикл (для колебательных

режимов), то мы можем сделать сравнительно достоверный прогноз ее поведения $x_2(t)$ по предшествующим значениям показателя ее поведения $x_1(t)$ даже при определении производной $x'(t)$ с некоторой небольшой погрешностью, поскольку $d(t) = |x_1(t) - x_2(t)|$ быстро расти не сможет. Однако для странного аттрактора $d(t)$ растет с увеличением t быстро (для моделей кинетики как $e^{\lambda t}$, $\lambda > 0$), и поэтому через интервал времени τ_Γ ($\tau_\Gamma \approx 1/\lambda$) траектории странного аттрактора, близкие вначале, перестанут быть близкими. Это означает наличие фундаментальных ограничений на возможность прогноза в подобных нелинейных системах, т. е. наличие «горизонта прогноза», интервала, после которого прогноз невозможен из-за бифуркаций и наличия хаотических явлений (рис. 4.15). Отсюда и произошло название *детерминированный хаос*, который иногда заменяется термином *динамический хаос*).



Аттрактор Лоренца



Бифуркационная диаграмма логистического уравнения

Рис. 4.15. Примеры странного аттрактора и хаотических бифуркаций

Вторая особенность хаотических странных аттракторов связана с тем, что геометрическая структура их фазового портрета обладает *масштабной инвариантностью*. Это означает, что в мелком и крупном масштабе они выглядят примерно одинаково (ср. с самоподобным масштабированием фрактальных структур).

В связи с этим в последнее время под странными аттракторами понимают клубок траекторий с фрактально-подобной структурой в фазовом пространстве динамического хаоса диссипативной системы.

Обратим внимание на различие движений, описываемых моделями случайных функций и динамического хаоса (см. рис. 3.3). При случай-

ном движении могут быть не известны действующие силы, но известны случайные характеристики объекта, обеспечивающие такое движение, либо известен случайный характер входных, управленческих и внешних воздействий, либо и то и другое. При хаотичном же движении мы рассматриваем систему и ее движение как детерминированные, где отсутствуют случайные или иные непредсказуемые силы, возмущения, параметры, но сам характер движения неслучайным образом зависит от начальных условий. *Такая динамика имеет место только в нелинейных системах*, причем в таких из них и в таких условиях их функционирования, когда небольшие изменения в первоначальных и предшествующих состояниях вызывают большие различия в последующем поведении системы, в ее движении к новым состояниям. Из-за этого всегда имеющиеся даже незначительные погрешности в определении настоящих и предшествующих состояний системы не позволяют уточнить конкретную траекторию движения в будущем. Именно из-за этого становится невозможным предсказание поведения таких систем за пределами определенного горизонта прогноза. Иными словами, на этапе детерминированного поведения будущее однозначно диктуется прошлым, а за пределами горизонта прогноза нам доступно только в лучшем случае предсказание вероятности нахождения такой системы в разных состояниях вместо однозначного или хотя бы стохастического определения траекторий ее будущего поведения.

Появление работ по теории динамического хаоса породило одну из (вторую по хронологии) трех парадигм синергетики – *диссипативных структур, динамического хаоса и сложности*. Как оказалось, такие модели пригодны для описания турбулентности, конвективного (переносного) движения газов и жидкостей, метеопрогнозов, таких системных моделей с сильными нелинейностями, как колебания изогнутых упругих систем, динамики «колесо–рельс» и шаров бильярда, механические системы с зазором и мертвой зоной, с трением скольжения, тепловая конвекция в атмосфере, лазерные и нелинейные оптические системы, системы управления с нелинейными обратными связями. Заметим, что в 1990 году появился вал публикаций по возможности с помощью малого управления в виде обратной связи, приложенной к нелинейной хаотически колеблющейся системе, коренным образом изменять ее динамику и свойства.

4.4.6. Фракталы

В ходе наблюдений оказалось, что в природе (например, береговая линия Англии, Л.Ф. Ричардсон, 1922 г. [1]) имеет место ситуация, когда результаты измерения зависят от единицы измерения: чем меньше единица измерения, тем большее значение имеет измеряемая величина (длина береговой линии) в пересчете на одинаковые единицы. Тем самым складывалось впечатление в нарушении условия 6 из указанных в начале подпараграфа 4.4.1. Оказалось, что это связано с проблемой длины извивающихся кривых. С подобным парадоксом уже встречались математики (Д. Пеано, 1890; Х. фон Кох, 1904 и др., см. [1]). В 1919 году немецкий тополог Феликс Хаусдорф решил проблему извивающихся кривых, введя для них дробную размерность. Однако систематическое изучение подобных объектов началось только с 1975 года французским математиком польского происхождения Бенуа Б. Мандельбротом [1]. Для описания таких объектов он ввел термин **фрактал** (англ. – fractal) как производный от двух латинских слов: frangere – ломать и fractus – дробный.

Фрактал – масштабно-инвариантное рекурсивно бесконечно самоподобное множество (в частности геометрическая фигура) дробной (нецелой) размерности по Хаусдорфу.

Масштабно-инвариантное – значит, одинаковое, в каком бы масштабе его ни наблюдать. **Самоподобное множество** – это множество, представимое в виде объединения одинаковых непересекающихся подмножеств, подобных исходному множеству. **Бесконечно рекурсивная** самоподобная геометрическая фигура – это фигура, каждый фрагмент которой повторяется при уменьшении масштаба. **Дробная размерность** – это размерность, не выражаемая целым числом.

Примеры фрактальных рекурсий

Алгебраическая рекурсия $z_{n+1} = f(z_n)$, где f – некоторая детерминированная функция; $z = a + jb$ – комплексное число, $j = \sqrt{-1}$.

Например, рекурсия

$$z_{n+1} = z_n^2 + c, \quad (4.185)$$

где c – комплексное число.

Открытое множество точек комплексной плоскости, получаемых при рекурсивном итерационном преобразовании с регулярной динами-

кой (4.185), в 1905 году начал изучать Пьер Жозеф Луи Фату (математик и астроном, Франция). При фиксированном изменяемом z_0 получается замкнутое множество с хаотической динамикой Жулиа, которое в 1917–1918 годах начал изучать Гастон Морис Жулиа, французский математик из Алжира. При $c = 0$ – простое множество Жулиа. Публикации П. Фату и М. Жулиа (1917–1918), посвященные исследованию множеств (4.185), ряд исследователей относит к первым публикациям по теории хаоса. Множества, образованные рекурсией (4.185) при $z_0 = 0$ и изменяемом c , называют множеством Мандельброта. Именно с них он начал в 1975 году аналитическое исследование фракталов. Для их изучения с 1983 года началось интенсивное использование рядом ученых компьютеров. В 1986 году Б. Мандельброту впервые удалось получить четкое компьютерное изображение множества, позже и названного его именем.

Геометрическая фрактальная рекурсия

Этот вид рекурсии впервые описан в работе Б. Мандельброта «Фрактальная геометрия природы» (1977). Хотя такие рекурсии были фрагментарно известны с 1890 года (кривые Пеано, снежинка Коха и т. д., см. рис. 4.16), именно Б. Мандельброт в своей книге вводит понятия фрактала и фрактальной геометрии как новой ветви математики и отмечает всеобщий характер фрактальных форм.

Суть геометрической фрактальной рекурсии сводится к следующей рекурсивной процедуре. На каждом n -м шаге каждая часть некоторой геометрической фигуры размерности d , состоящей из N частей, полученных на $(n - 1)$ -м шаге рекурсии, уменьшается в r раз в каждой части на каждом n -м шаге, $n \geq 1$.

Пояснение сути рекурсий дано на рис. 4.16 и 4.17.

Рис. 4.16 позволяет наглядно проиллюстрировать суть дробной размерности Хаусфорда. Когда мы рассматриваем обычные геометрические объекты в евклидовом пространстве (такие как линия, квадрат, конус, шар, куб и т. п.), то их топологические размерности d_T измеряются в целых числах: 1, 2, 3 и т. п. Если фракталы рассматривать как множество точек, вложенных в евклидово пространство, то они не будут занимать все пространство соответствующей размерности, выходя своим объемом за пространство n -й размерности ($n = 1, 2, 3, \dots$) и не занимая полностью пространство $(n + 1)$ -й размерности. Например, чем более тонкой нитью мы будем измерять длину береговой линии

Англии, карта которой вписана в прямоугольник, тем она будет длиннее периметра прямоугольника, но будучи плотно плоско вложенной в него, не сможет занять весь прямоугольник, всю его площадь. Иными словами, береговая линия будет иметь фрактальную размерность больше 1, но меньше 2.

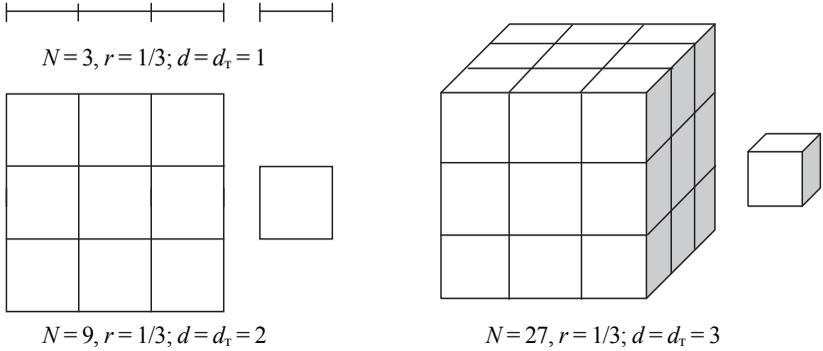


Рис. 4.16. Пояснение первого шага геометрической рекурсии

Предположим, что объемная фигура разбивается на N частей (рис. 4.16) так, что объем каждой из частей в $1/r^d = r^{-d}$ раз меньше исходного, и при этом $Nr^d = 1$. Может ли в таком случае d быть целым? Да, его можно найти как $d = \log N / \log(1/r) = \log_{(1/r)} N$. Это и есть (фрактальная) размерность объема по Хаусдорфу (см. рис. 4.17).

Ясно, что приведенное определение фрактала *математическое*, так как они *бесконечно самоподобны*. Для них характерны *дробная размерность* по Хаусдорфу, недифференцируемость и необходимость оперировать дробными производными и интегралами.

Что касается примеров, рассмотренных на рис. 4.16 и 4.17, то для них еще характерен алгоритм *детерминированного построения*. Поэтому их использование как моделей физических объектов может быть приближенным, ограниченным. Для реальных естественных и искусственных объектов можно ввести модельное понятие *физические фракталы* – геометрические объекты (линии, поверхности, тела), отличающиеся сильной изрезанностью структуры, ограниченным (по масштабу) самоподобием, детерминированным (законовым) или инде-

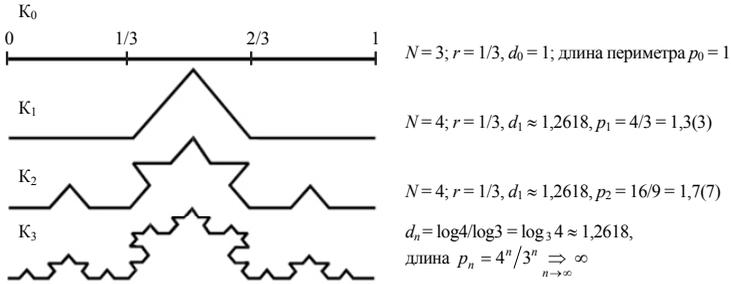
терминированным, но регулярным (закономерностным) (в частности случайным) принципом построения.

Изрезанность понятна из изложенного ранее. Именно она приводит к превышению фрактальной размерности объекта над топологической, характерной для мало изрезанных объектов. Ограниченное самоподобие означает наличие конечного числа операций, рекурсий. Регулярная (закономерностная) индетерминированность отражает тот факт, что в детерминированную процедуру создания фрактала включается внешний фактор, превращающий ее в индетерминированную, но обнаруживающую закономерностную регулярность в массовой реализации. Примерами таких регулярных индетерминированных (см. рис. 3.3) являются *случайные фракталы*, в которых на n -м такте ($n = \overline{1, N}$) каждого k -го шага происходит рандомизация – случайным образом (согласно какому-то дискретном закону распределения) выбирается направление расположения новых частей образуемого фрактала. Например, для кривой или снежинки Коха (см. рис. 4.17, *a, б*) на втором, третьем и дальнейших шагах каждая подобная другим фигурка строится на своем участке либо с наружной (наружу), либо с внутренней (внутри) стороны делимой на этом шаге линии участка согласно бинарному, биномиальному или подобному им закону распределения с конечным числом исходов, равным числу N частей фигуры.

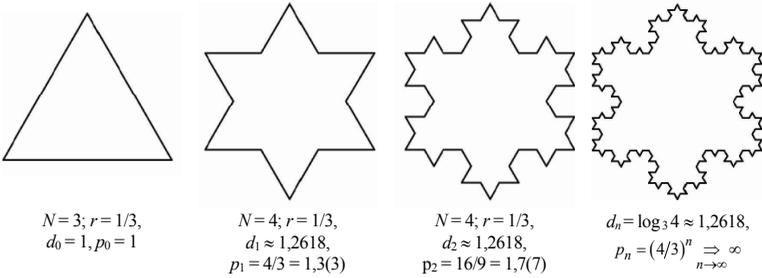
Внимание! Попробуйте сами построить кривую или снежинку Коха, используя в качестве «генератора» (датчика) последовательности выборочных значений случайной величины с бинарным законом распределения (см. § 4.3) подбрасывание монеты. Если выпадает «орел» – рисуем новую часть с наружной стороны соответствующего отрезка ломаной предыдущей рекурсии, если «решка» – внутрь (рис. 4.18).

В настоящее время фракталы широко используются для модельного представления природных объектов: модели снежинок, листьев, кустов, деревьев, лесных массивов, гор, облаков, анатомических объектов (бронхиальное или артериальное дерево), для описания и исследования различных изображений, в частности их текстур, в радиолокации, гидрологии, передаче данных и обработке сигналов, особенно с «хвостатыми» распределениями. Заметим, что «хвостатые» распределения, т.е. распределения с сильно протяженными концами (хвостами) могут не иметь моментов (см. примеры в [20]), отражают сложный, в том числе фрактальный, характер механизма образования случайности.

a – кривая Коха



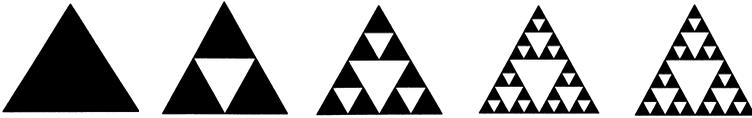
б – снежинка Коха



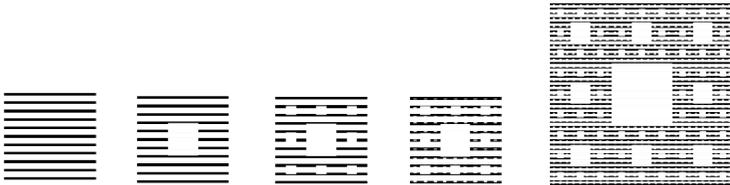
Предельная длина границы (периметр) снежинки Коха – бесконечность

в – ковры Серпинского K_1 и K_2

K_1 (треугольный), $d = \log(3)/\log(2) \approx 1,585$



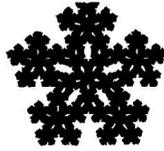
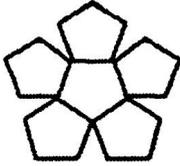
K_2 (прямоугольный), $d = \log_3 8 \approx 1,89$



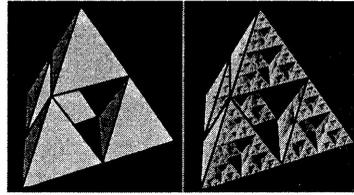
Топологическая размерность (площадь S) ковров Серпинского равна нулю

Рис. 4.17 (начало)

д – звезда Дюрера



е – пирамида Серпинского



$d = \log_2 5 \approx 2,32$; топологический объем $V = 0$

ж – губка Менгера, $d = \log_3 20 \approx 2,73$; $V = 0$

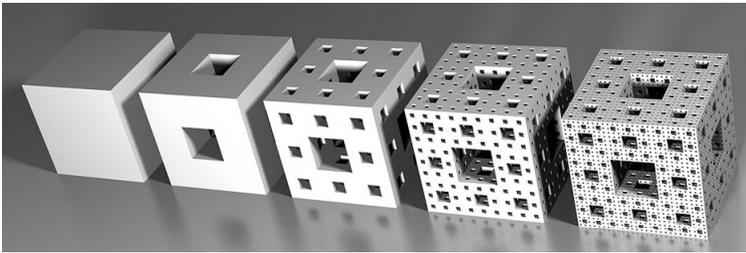
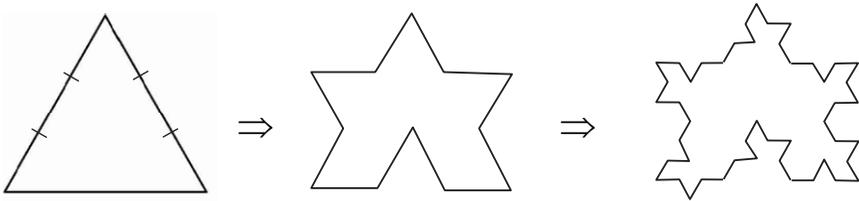


Рис. 4.17 (окончание). Примеры применения детерминированных рекурсивных алгоритмов построения геометрических фракталов¹



$$d_n = \log_3 4 \approx 1,2618; p_n = (4/3)^n \xrightarrow{n \rightarrow \infty} \infty$$

Рис. 4.18. Пример создания случайной (рандомизированной) снежинки Коха

¹ С многообразием двух- и трехмерных цифровых изображений красоты фракталов можно ознакомиться через поисковик Google: картинки фракталов.

4.4.7. Аппарат экспертного описания и исследования объектов

Как следует из подпараграфов 4.4.3, 4.4.4, предпосылкой использования нечеткого и интервального аппаратов являются ситуации, характеризующиеся следующими особенностями.

1. Субъект-исследователь не может построить объективную меру возможности появления конкретного значения исследуемой величины (параметра объекта, измеряемой величины и т. п.), принадлежности его определенному множеству (области) значений.

2. Субъект может допустить, что любое значение величины может принадлежать множеству всех возможных или рассматриваемых значений с определенной степенью предпочтения (нечеткий аппарат) или заменить точечные значения на интервальные (интервальный аппарат).

3. Исследуемые величины, переменные могут быть недостаточно определенными либо качественными, но к их поведению и представлению применим формальный аппарат, в частности описанный в подпараграфах 4.4.5 и 4.4.6.

А как быть, если вообще трудно формализовать постановку задачи и выбрать строгие объективные формальные методы решения ее. Для этого необходимо использовать разум, интеллект, интуицию, сознание, опыт, компетенции экспертов в области решаемой задачи. В таких ситуациях прибегают к аппарату экспертного исследования объектов.

Его идея сводится к привлечению для решения задачи описания и исследования объектов в условиях слабой формализации разных этапов технологического процесса решения задачи компетентных в этом специалистов (экспертов), способных искать и находить решения даже в этих «плохих» условиях. Вариант использования для этого средства искусственного интеллекта, в частности экспертных систем, будет рассмотрен позже.

Каждый эксперт – индивидуальная личность, наиболее компетентная в определенном круге вопросов и решаемых задач, зависящая от внешних и внутренних факторов, включая психологические, межличностные, мотивационные и т. п. Поэтому при организации работы группы экспертов, их числа, подбора необходимо стремиться к созданию благоприятных условий для всей группы и нейтрализации тех факторов, которые могут неблагоприятно влиять на работу экспертов. Для этого можно включать элементы подбора экспертов, выявления их согласованности, в частности в критериальной оценке исследуемых фактов, организации рандомизации итераций и обучения экспертов по

ходу исследования, формальные приемы согласования и учета их мнения и т. д. Не останавливаясь на этом подробно, отметим несколько важных моментов.

1. Существует множество вариантов экспертного решения задач (см., например, [10, 13–15, 22]) и даже рекомендации их совместного использования. Они зависят от решаемых задач, их постановки, сопутствующих условий и других причин. Сюда, в частности, относятся методы ранжирования, парного и множественного сравнения, непосредственной оценки последовательного сравнения Черчмена–Акоффа, Терстоуна, Неймана–Моргенштейна, дерева целей, мозгового штурма.

2. Типичная схема технологического процесса экспертного описания и анализа объектов, выявления проблем, выбора вариантов и принятия решений имеет вид, изображенный на рис. 4.19 (см. также [22]).

Заметим, что для каждого этапа технологического процесса существуют разные варианты его реализации.

3. Одним из важнейших этапов техпроцесса экспертного исследования объектов является формирование экспертной комиссии (группы). Существующие методы реализации этого этапа мало формализованы и не имеют строгих научных рекомендаций – как по количественному, так и по качественному составу. Имеют место лишь некоторые процедуры и подсказки. Например, следующие: подбор экспертов количественно и качественно должен осуществляться на основе анализа сложности решаемой проблемы, задачи, характеристик экспертов, затрат ресурсов, включая временные; необходимо привлекать к экспертизе специалистов различного профиля. При этом минимальное число экспертов должно определяться количеством тех трудно формализуемых аспектов, которые надлежит учесть при решении поставленной задачи в конкретных условиях. Верхняя граница численности экспертов комиссии определяется исходя из изложенного выше, относительно согласованности и разнопрофильности экспертов, ограниченности ресурсов и гипотезы: если все эксперты являются допустимо «точными измерителями», то с увеличением их числа достоверность результатов всей группы возрастает. Но можно ли всех экспертов считать «точными измерителями»? Применим ли здесь статистический подход к уменьшению статистических отклонений от среднего, не будет ли при этом проявляться систематическая погрешность за счет идеологической или какой-либо иной предвзятости всех экспертов? На эти и подобные вопросы ответы необходимо искать для каждого конкретного случая.



Рис. 4.19. Схема технологического процесса экспертного описания и анализа объектов

4. Чтобы избежать ситуаций, когда разные эксперты могут иметь диаметрально противоположные или близкие к таковым мнения по одним и тем же вопросам для каждого исследуемого объекта, а индивидуальные характеристики экспертов и разница их компетенций по каждому аспекту решаемой задачи и условий ее решения не будут способствовать повышению достоверности результатов экспертизы, необходимо выполнение этапов 3–7 (рис. 4.19). Как уже упоминалось, каждый из этих этапов может выполняться разными методами. Например, для учета индивидуальных особенностей экспертов и уровня их общей и частных компетенций в области решаемой задачи и ее разнообразных аспектов можно учитывать априорные и апостериорные данные.

В качестве априорных обычно используются итоги самооценки экспертов и взаимной оценки со стороны других экспертов. Самооценка осуществляется для предварительного отбора экспертов по анкете, содержащей перечень вопросов, подлежащих решению комиссией (группой), в виде баллов, которыми эксперт оценивает свою компетентность по каждому вопросу. Получив среднее или медианное значение баллов для всей группы по каждому вопросу, отбираются те эксперты, баллы которых предпочтительнее по решаемым вопросам или по всем вопросам.

Взаимная оценка осуществляется следующим образом. Специалисты знакомятся с открытым перечнем лиц, предлагаемых для решения поставленной проблемы (задачи). Специалист должен высказать свое бинарное мнение, т. е. «да» или «нет», о включении перечисленных лиц в экспертную группу и при желании предложить новые лица. По итогам нескольких предварительных туров опроса составляется единый полный список кандидатов в эксперты и рассчитывается коэффициент компетентности k_i каждого i -го эксперта из этого списка:

$$k_i = \frac{\sum_{j=1}^n x_{ij}}{\sum_{i=1}^n \sum_{j=1}^n x_{ij}}, \quad i = \overline{1, n}, \quad (4.186)$$

где $x_{ij} = \begin{cases} 1, & \text{если } i\text{-й специалист включил } j\text{-го кандидата;} \\ 0, & \text{если } i\text{-й специалист не включил } j\text{-го кандидата.} \end{cases}$

Понятно, что $0 \leq k_i \leq 1$, а $k_1 + k_2 + \dots + k_n = 1$.

Апостериорные данные используются, например, на седьмом этапе (рис. 4.19) для определения согласованности комиссии (группы) экспертов. Для этого проводится N тестовых решений задач, подобных рассматриваемой, и определяется один из коэффициентов согласованности экспертов (конкордации Кендалла, Устюжанинова и др.). Например, широко используемый коэффициент W конкордации (согласованности) Кендалла

$$W = \frac{12S}{n^2(N^3 - N)}, \quad (4.187)$$

где S – сумма квадратов отклонений суммы рангов в каждой тестовой задаче; n – число экспертов; N – число тестовых задач; $W \in [0, 1]$. Чем ближе W к 1, тем согласованнее группа экспертов.

Методы формализации этапа обработки мнений экспертов, в том числе с учетом их компетенций, можно найти в [10, 13–15, 22].

В заключение несколько слов о методах, способствующих повышению качества работы экспертов.

Метод Дельфи используется на этапах 3–7, 8.2, 9.2 (см. рис. 4.19). Он базируется на следующих посылах.

1. Допускается критика идей, а не персон, носителей идеи.

2. Критика мнений эксперта в его присутствии может психологически увязываться экспертом с персональной критикой.

3. Критика благотворно влияет на эксперта, если она психологически не связана с персональной конфронтацией, а проводится и обсуждается анонимно. Это способствует тому, что некоторые эксперты будут склонны не только критиковать других, но и прислушиваться к критике коллег, а возможно, даже корректировать и уточнять собственное мнение.

4. Во время коллективных обсуждений могут проявляться такие психологические факторы, как присоединение к мнению наиболее авторитетного или авторитарного специалиста, желание отказаться от публично выраженного мнения, следование за большинством и т. п. Поэтому необходим полный отказ от коллективных обсуждений.

5. Цель этапов 3–7, 8.2, 9.2 не только получение полной согласованности экспертов группы, но и выявление двух и более типов мнений, отражающих принадлежность экспертов к разным профессиям, научным школам, ведомствам, категориям, должностям и т. п.

Суть метода сводится к проведению нескольких туров индивидуальных опросов, исключая контакты между экспертами. По итогам каждого из туров экспертам сообщаются результаты обработки итогов предыдущего тура с указанием расположения оценок каждого эксперта от среднего или медианного. При этом какие-либо дискуссии между экспертами запрещены, чтобы исключить влияние психологических и эмоциональных факторов. Если оценка эксперта сильно отклоняется от среднего (медианного) значения, то его просят аргументировать свое мнение или изменить оценку. В начале следующего тура эксперты аргументируют или изменяют свою оценку с объявлением причин корректировки. Вся работа по реализации метода проводится отдельной управляющей группой, в которую обязательно входят системный аналитик и лицо, принимающее решения. При этом аноним-

ность экспертов сохраняется до конца работы обязательно, а после ее окончания – по желанию эксперта.

Обычно трех-четырёх туров достаточно, чтобы ситуация стабилизировалась. Это и служит критерием прекращения дальнейшего проведения туров опросов. Многотурность опросов требует значительных ресурсов и необходимости многократного пересмотра экспертами своих оценок, что может вызвать отрицательную реакцию на это и сказаться на результатах экспертизы. Поэтому имеется много модификаций метода (QUEST, SEER и др.).

Метод мозгового штурма (мозговой атаки, коллективной генерации идей, конференции идей) используется на тех этапах, когда в начале или в ходе исследований возникают трудные, даже тупиковые ситуации, для которых известные способы и варианты решения задачи оказываются непригодными. Тогда требуется групповое обсуждение с участием любых лиц для генерации новых идей и вариантов решения задачи, «расшивки» тупиковой ситуации. Для получения наибольшего эффекта участники коллектива групповой генерации и обсуждения новых идей и вариантов должны руководствоваться совокупностью определенных правил. Конкретная совокупность зависит от конкретных модификаций метода. В основной массе она сводится к следующим правилам.

1. Определен основной принцип отбора лиц в группу (коллектив): разнообразие профессий, квалификаций, опыта, чтобы расширить суммарный объем априорной информации о возможных вариантах решения проблемы, задачи.

2. Обеспечивается полная свобода мышления участников группы и выработки ими новых идей.

3. Приветствуются любые, даже кажущиеся «бредовыми», абсурдными, нереализуемыми, идеи, варианты.

4. Делается все для генерации как можно большего количества идей.

5. Категорически запрещается любая критика идей, но поддерживается обсуждение того полезного, что содержится в ней, развитие идеи, в том числе путем частичного улучшения чужих идей. Надо руководствоваться принципом: «Возможность критики – тормоз воображению, креативному (творческому) мышлению».

6. Технологически желательно, чтобы каждый по очереди зачитывал свою идею, а остальные члены группы записывали свои идеи, возникающие под влиянием услышанного.

7. При наличии большого количества новых идей осуществляется их «грубое» отсеивание на качественном (а не по количественным показателям) уровне, определяемом поставленной задачей. Причем отсеянные идеи не выбрасываются, из них берется то полезное, что может пригодиться в дальнейшем, при появлении возможности реализации.

8. Сбор, сортировка и анализ идей по окончании работы коллектива генераторов идей производятся, как правило, другой группой экспертов.

9. За несколько итераций можно значительно увеличить количество идей путем ассоциаций с другими и комбинацией идей.

§ 4.5. АППАРАТ ПОИСКА ОПТИМАЛЬНЫХ РЕЗУЛЬТАТОВ

4.5.1. Вводные замечания

Рассматривая в предшествующих пунктах математические модели, мы обращали внимание в основном на то, чтобы получаемый с их помощью результат был корректен и как можно адекватнее отражал истинное положение дел. Лишь изредка ставился вопрос о возможности получения наилучшего в каком-то смысле результата в тех условиях, которые сопровождают исследование объекта, экспериментирование с ним. В частности упоминали об оптимальных планах эксперимента, об оптимальных¹ (в частности эффективных) методах оценивания характеристик и параметров, робастных методах.

Однако согласно табл. 2.2 оптимальными могут быть не только разные методы, алгоритмы, планы, но и исходные либо специально построенные модели, являющиеся основой для формализации постановок.

И ранее, и сейчас совместно со словом *оптимальный* указывалось «наилучшее в каком-то заданном, принятом, требуемом смысле». Что же это означает? Во-первых, выделение того количественного или качественного показателя, признака, свойства ожидаемого результата

¹ От *лат.* *optimus* – наилучший.

или их совокупности, которые наиболее важны пользователю результата. Во-вторых, выбор предпочтений, по которым выбираются лучшие, с точки зрения поставленной цели оптимизации, решения. Иными словами, для выявления, какой результат наилучший, необходимо иметь (сформулировать, выбрать из известных) **критерий оптимальности**, ориентируясь на ту задачу, ради которой создается и используется оптимальная модель, результат оптимизации.

Желательно иметь количественную основу критерия оптимизации. Для этого необходима аналитическая модель формализованной постановки задачи оптимизации, по которой и отыскивается оптимальная модель. Чтобы ее построить, среди величин, характеризующих исследуемый объект, выделяются те величины, значения которых можно изменять, отыскивая оптимальное решение, и те, которые задаются как константы. Первые называются **параметрами оптимизации**, а вторые используются для составления математической модели. Она состоит из количественного выражения, отражающего меру достижения оптимизации, и условий в виде ограничений типа равенств или неравенств, накладываемых на возможное множество значений переменных с учетом заданных констант.

Рассмотрим три класса таких оптимальных моделей, отличающихся их аналитическими формулировками количественной меры оптимальности. Назовем их условно оптимизационные, вариационные и игровые (состязательные).

4.5.2. Детерминированные оптимизационные модели

Оптимизационные (или математического программирования) – это модели, в которых критерий оптимальности сводится к поиску экстремумов (максимумов или минимумов) функций многих переменных на их заданных числовых множествах X при наличии дополнительных условий (ограничений) на значения x_1, \dots, x_n этих переменных, представляемых в виде равенств и (или) неравенств. Общая постановка оптимизационных задач данного класса: максимизировать (прибыль, доходы, вероятность безотказной работы и т. п.) или минимизировать (потери, вероятность отказа, расходы, ...) функцию

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) \Rightarrow \underset{\mathbf{x}}{\text{extr}}, \mathbf{x} \in D \quad (4.188)$$

при условии (ограничениях)

$$g_i(x_1, x_2, \dots, x_n) \geq 0, \quad i = 1, 2, \dots, m; \quad (4.189)$$

$$h_j(x_1, x_2, \dots, x_k) \geq 0, \quad j = 1, 2, \dots, k, \quad (4.190)$$

где $\mathbf{x} = (x_1, x_2, \dots, x_n)$; D – **допустимое множество** решений задачи, любой элемент которого является **допустимым**, т. е. удовлетворяющим всем ограничениям задачи, **решением** или **допустимой точкой** решения оптимизационной задачи.

Составная часть **критерия оптимизации** (4.188) функция $f(\mathbf{x})$, отражающая зависимость критерия от переменных, называется **целевой (критериальной) функцией**. Точка $\mathbf{x}^* \in D$, в которой $f(\mathbf{x})$ достигает экстремального значения $f^* = f(\mathbf{x}^*)$, называется **оптимальным решением**, полученная при этом модель – оптимальной моделью, соответствующей искомому результату, или **оптимальным значением целевой функции** f^* . В дальнейшем будем рассматривать только конечномерные задачи, когда допустимое множество решений принадлежит n -мерному евклидову пространству R^n , т. е. $D \subset R^n$.

Математическая дисциплина, посвященная теории и методам нахождения экстремумов функций многих переменных при наличии дополнительных ограничений на эти переменные, имеющих форму равенств и неравенств, называется **математическим программированием** (МП). Поэтому оптимизационные задачи и модели такого класса называются иногда задачами и моделями математического программирования.

С подобными оптимизационными задачами мы уже сталкивались в регрессионном анализе (см., например, (4.24), (4.58), (4.67)) и в теории оценивания (см. (4.24), (4.58), (4.66), (4.128)). Рассмотрим другие задачи данного класса.

Прежде всего отметим, что если целевая функция $f(\mathbf{x})$ в (4.188) и функции g_i и h_j в ограничениях (4.189), (4.190) являются линейными, то такие задачи МП называются **задачами линейного программирования** (ЛП).

Типичная ЛП-задача (ср. с (4.188)–(4.190)):

$$\left\{ \begin{array}{l} \sum_{i=1}^n c_i x_i \Rightarrow \max(\min)_x^1; \\ \sum_{j=1}^m a_{ij} x_j \leq b_i, \quad i = \overline{1, m}; \\ x_j \geq 0, \quad j = \overline{1, n}. \end{array} \right. \quad \begin{array}{l} (4.191) \\ (4.192) \\ (4.193) \end{array}$$

В отличие от функциональных ограничений (4.189), (4.190), (4.192) ограничения вида (4.193), т. е. вида $x_j \geq d_j$, $x_j \leq d_j$, $d'_j < x_j \leq d''_j$, $j = \overline{1, k}$, называются **прямыми (тривиальными) ограничениями** на переменные задачи, причем два первых из них относятся к **односторонним**, а последнее – к **двусторонним**.

Если в (4.191) нахождение максимума (минимума) заменить на нахождение минимума (максимума), а неравенство \leq в (4.192) заменить на противоположное \geq , то такая ЛП-задача называется **двойственной** к первой (прямой) задаче (4.191)–(4.193).

Центральной в теории ЛП является **теорема двойственности**, устанавливающая равенство оптимальных значений f^* целевых функций прямой и двойственной задач.

Если в задачах МП (4.188)–(4.190) хотя бы одна из функций f , g , h нелинейная, то такие МП-задачи называются **задачами нелинейного программирования** (НП). Ясно, что среди них можно выделить отдельные подклассы МП-задач с **линейными** и **нелинейными целевыми функциями** либо **ограничениями**. При этом ограничения могут быть в виде **неравенств** (4.189) или **равенств** (4.190).

Так, если целевая функция (4.188) является квадратичной формой или суммой линейных и квадратичных форм, а g_i и h_j в (4.189), (4.190) – аффинные функции (вида (4.192)), то такой класс МП называется **квадратичным программированием** (КП). Напомним, линейная

¹ Нетрудно убедиться, что заменой $f(x)$ на $-f(x)$ можно задачу отыскания максимума заменить на задачу отыскания минимума и наоборот. Поэтому в выражениях типа (4.188), (4.191) часто пишут только \max .

функция $f(x)$ называется **аффинной**, если $f(\lambda x + \mu y) = \lambda f(x) + \mu f(y)$ для всех таких λ и μ , что $\lambda + \mu = 1$. Иными словами, для задач КП вместо ЛП в левой части (4.191) могут входить слагаемые не только вида $c_i x_i$, но и вида $c_i x_i^2$, $c_{ij} x_i x_j$, $i, j = \overline{1, n}$.

Если целевая функция $f(x)$ является выпуклой и рассматривается на выпуклом множестве, то МП-задача относится к задаче **выпуклого программирования**. Напомним, что выпуклой называется функция $f(x)$, графиком которой является выпуклая функция (см. рис. 4.20).

Другими словами, функция $f(x)$ называется выпуклой вверх (вниз) на отрезке $[a, b]$, если каждая дуга графика этой функции лежит не ниже (не выше) стягивающей ее хорды. Аналитически это означает следующее. Пусть $y = y(x) = ax + b$, уравнение прямой, отрезком которой является хорда AB , т. е. прямой, проходящей через точки $A(x_1, f(x_1))$ и $B(x_2, f(x_2))$. Функция $f(x)$ называется **выпуклой вверх (вниз)**, если $f(x) \geq y(x)$ ($f(x) \leq y(x)$) при $x_1 \leq x \leq x_2$, где x_1, x_2 – любые точки отрезка $[a, b]$ (!). Если нестрогие неравенства заменить на строгие (т. е. \geq, \leq на $>, <$), то $f(x)$ называется **строго выпуклой вверх (вниз)**. **Выпуклое множество** – это такое множество в евклидовом или другом векторном пространстве, которое вместе с любыми двумя точками содержит все точки соединяющего их отрезка.

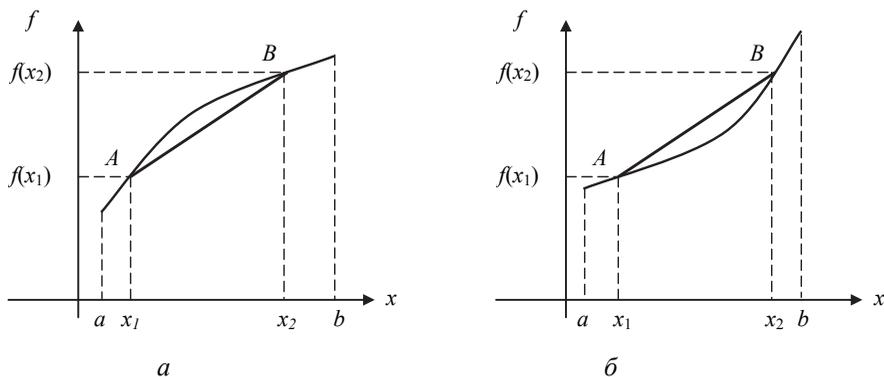


Рис. 4.20. Примеры выпуклой вверх (вогнутой вниз) (а) и выпуклой вниз (вогнутой вверх) (б) функций

Если целевая функция (4.188) является отношением двух линейных функций $f_1(\mathbf{x})$ и $f_2(\mathbf{x})$, т. е.

$$f(\mathbf{x}) = \frac{f_1(\mathbf{x}) + a}{f_2(\mathbf{x}) + b}, \quad (4.194)$$

где числа a, b заданы, а ограничения линейны (вида (4.192), (4.193)), то имеем задачу *дробно-линейного программирования* (ДЛП).

Если в (4.191) $c_i x_i$ заменить на $p_i(x_i)$, а (4.192) $a_j x_j$ заменить на $q_{ij}(x_j)$, где p и q – заданные детерминированные функции, то такой класс МП-задач называется задачами *сепарабельного программирования* (СпбП).

Если в МП-задачах (4.188)–(4.190) f и g – функции специального вида, например,

$$f(\mathbf{x}) = \sum_{i=1}^n c_i \prod_{j=1}^n x_j^{\alpha_{ij}}, \quad (4.195)$$

где α_{ij} – заданные действительные числа, то имеем дело с *задачами геометрического программирования* (ГП).

Другое деление МП-задач – это деление их по виду множества D допустимых решений. Если D представляет собой конечное множество, то такие МП-задачи называются *задачами дискретного программирования* (ДП). Среди них особый подкласс составляют *задачи целочисленного программирования* (ЦП). Они отличаются тем, что в них все переменные x_i являются целыми числами. В этом случае к условиям задач вида (4.191)–(4.193) добавляется третье условие (ограничение): x_i – целые числа, $i = \overline{1, n}$. Заметим, что в них все переменные, а не часть их, как в других задачах ДП, являются целыми числами.

Казалось бы, коль скоро множество D допустимых решений конечно, можно любые ДП-задачи решить методом полного перебора. Это означает вычисление и сравнение (при учете ограничений) всех значений целевой функции $f(\mathbf{x})$. Однако очень часто такой перебор не является тривиальным: число допустимых решений (т. е. элементов области задания ЦФ с учетом ограничений) настолько велико, что их

полный перебор трудно реализуем даже современными средствами. Поэтому для решения ДП, например ЦП-задач, чаще всего используют две группы методов.

Исторически первой группой решения ЦП-задач являются *методы отсечения*. Их идея состоит в двухэтапности решения задачи. Сначала третье условие целочисленности отбрасывается и решается обычная МП-задача. Если ее решение оказывается целочисленным, то процесс решения считается законченным. В противном случае к ограничениям типа (4.189), (4.190) или (4.192), (4.193) добавляется еще одно линейное ограничение, которому полученное нецелочисленное оптимальное решение не удовлетворяет, а все целочисленные решения удовлетворяют. Процесс продолжается до тех пор, пока полученное целочисленное решение не будет удовлетворять всей пополненной системе ограничений. Установлено, что это приводит к оптимальному целочисленному решению за конечное число шагов. Однако для этого даже при сравнительно небольших n , m , k может потребоваться весьма большое число шагов.

Вторая группа методов – *методы частичного перебора*. Среди них наиболее известны *методы ветвей и границ*. Их идея состоит в разбиении множества допустимых решений на несколько подмножеств и в отбрасывании «неперспективных», т. е. заведомо не содержащих оптимального решения. Однако объем вычислений и здесь может быть велик, иметь тенденцию к экспоненциальному (или близкому к нему) росту при увеличении числа переменных. Поэтому большое внимание уделяется другим, в том числе приближенным, методам, учитывающим специфику и содержательную постановку исследовательской задачи. Они направлены на то, чтобы ожидаемая трудоемкость была не экспоненциальной, а полиномиальной (см. далее методы искусственного интеллекта).

4.5.3. Стохастические оптимизационные методы

Этот класс оптимизационных моделей составляют модели, рассматриваемые в стохастическом программировании.

Стохастическое программирование (СП) – раздел МП, посвященный исследованию стохастических экстремальных задач. Это задача двух групп.

Первую группу составляют МП-задачи, в которых целевая функция и/или условия (ограничения) допустимости решений зависят от

случайных переменных, а максимизации или минимизации при заданных ограничениях подлежит некоторая характеристика случайной функции $\mathbf{Y} = f(\mathbf{X}) = f(X_1, X_2, \dots, X_n)$, где f – детерминированная функция, а \mathbf{X} – случайный вектор.

Чаще всего рассматривают две постановки задач СП: **М** и **Р**. Примером **М**-постановки СП-задачи является МП-задача (4.188)–(4.190), когда целевая функция $f(\mathbf{x}) = f(x_1, \dots, x_n; \Xi_1, \dots, \Xi_n)$, где Ξ_1, \dots, Ξ_n – n -мерный случайный вектор с заданным распределением, а критерий оптимальности (4.188) заменяется на подобный ему, если вместо Ξ_1, \dots, Ξ_n поставить их математические ожидания $\mathbf{M}\{\Xi_i\}$, $i = \overline{1, n}$. Например, МП-задача (4.191)–(4.193), в которой c_i заменяются случайными величинами C_i , а в целевой функции (4.191) c_i заменяются на $\mathbf{M}\{C_i\}$.

Второй, более общий, пример **М**-постановок: вместо нахождения экстремума функции $f(\mathbf{x})$ находится экстремум функции регрессии – условного математического ожидания $\mathbf{M}\{f(\mathbf{X}; \theta_1, \dots, \theta_k)\}$ при фиксированных значениях показателей $\Theta_1 = \theta_1, \dots, \Theta_k = \theta_k$ «состояния природы» случайности либо математических ожиданий $\mathbf{M}\{f(\mathbf{x}; \Theta)\}$ или $\mathbf{M}\{f(\mathbf{X}; \Theta)\}$.

Примером **Р**-постановки является СП-задача, в которой вместо целевой функции в (4.188) или (4.191) вводится вероятность

$\mathbf{P}\{f(\mathbf{x}; \Theta) \leq \alpha\}$ в (4.188) или $\mathbf{P}\left\{\sum_{i=1}^n C_i x_i \leq \alpha\right\}$ при максимизации либо

$\mathbf{P}\{f(\mathbf{x}; \Theta) \geq \alpha\}$ или $\mathbf{P}\left\{\sum_{i=1}^n C_i x_i \geq \alpha\right\}$ при минимизации ЦФ, где α – заданное число. Аналогичные замены в СП-задачах можно сделать с ограничениями.

Как видим, подобные постановки МП-задач приводят к сведению СП-задач к детерминированным МП-задачам.

Как видим, подобные постановки МП-задач приводят к сведению СП-задач к детерминированным МП-задачам.

Ко *второй группе* постановок СП-задач относятся те, в которых допустимым решением считается оптимальное распределение вероятностей на некотором множестве детерминированных решений или на

множестве функций от случайных параметров. В первом случае, например, принимаются известными значения $y = f(\mathbf{x}; \Theta)$ для отдельных $\Theta = \theta$, которые полагаются неизвестными исследователю «истинными» значениями, а $f(\mathbf{x}; \Theta)$ представляется суммой $f(\mathbf{x})$ и случайной ошибки $E_{f(\mathbf{x})}$, т. е.

$$f(\mathbf{x}; \Theta) = f(\mathbf{x}) + E_{f(\mathbf{x})}. \quad (4.196)$$

Во втором случае

$$f(\mathbf{x}; \Theta) = f(\mathbf{x}; \Theta_1, \Theta_2, \dots, \Theta_k) + E. \quad (4.197)$$

Заметим, что все СП-задачи могут решаться *обычными* и *байесовскими методами* стохастического аппарата.

4.5.4. Нечеткие и интервальные оптимизационные модели

Этот класс моделей связан с применением аппарата нечетких и интервальных моделей, рассмотренных в подпараграфах 4.4.3 и 4.4.4.

При этом нечеткость или интервальность может проявляться в разных формах и их комбинациях:

- нечеткость (интервальность) целевой функции;
- нечеткость (интервальность) выполнения ограничений;
- нечеткость (интервальность) задания переменных и/или параметров целевой функции;
- нечеткость (интервальность) задания переменных и/или параметров ограничений.

Например, ЦФ (4.188) и (4.191) или ограничения (4.189), (4.190), (4.192) и (4.193) можно смягчить, приняв возможность вместо достижения оптимального значения f^* ЦФ или строгого выполнения первых частей ограничений отклонение их в некотором диапазоне с соответствующей степенью допустимости, т. е. с соответствующей функцией принадлежности этому диапазону.

Внимание! Желающие могут ознакомиться с таким аппаратом оптимизации, а также предложить разные постановки таких задач самостоятельно.

4.5.5. Многокритериальная оптимизация

Все рассмотренные оптимизационные модели имеют общность в том, что они строились на базе нахождения экстремума одной целевой функции $f(x)$. Однако в ряде приложений необходимо находить экстремумы нескольких разных целевых функций, иногда противоречивых. Например, максимизировать скорость (ЦФ f_1 , критерий оптимальности K_1) и грузоподъемность самолета (f_2, K_2); минимизировать расход топлива (f_3, K_3) и собственный вес самолета (f_4, K_4). В этом случае речь идет о **многокритериальных (векторных) задачах оптимизации**. Пусть D – множество допустимых решений, а f_1, \dots, f_p – детерминированные вещественные целевые функции, заданные на множестве допустимых решений D . Требуется максимизировать или минимизировать функции f_1, \dots, f_p на этом множестве D . Поскольку f_1, \dots, f_p в общем случае не будут достигать требуемых критериями оптимальных значений f_1^*, \dots, f_p^* в одной и той же оптимальной точке x^* , а для каждой ЦФ $f_i, i = \overline{1, p}$, будет свое оптимальное решение x_i^* , термин *оптимальное решение* в предыдущем понимании будет не вполне корректным. Его следует пересмотреть, руководствуясь либо новыми принципами оптимальности по многим критериям, либо сводя задачу многокритериальной (векторной) оптимизации к скалярной. Последнее означает, что согласно определенным принципам многие критерии сводятся к одному, по которому проводится оптимизация в рассмотренных ранее постановках. Ознакомимся с некоторыми из таких принципов.

Поскольку зачастую ЦФ $f_i(x), i = \overline{1, p}$, имеют различную размерность и/или масштабы, их желательно привести к безразмерному виду с помощью какого-либо простого в исполнении функционального преобразования. Оно должно:

- иметь общее начало отсчета;
- иметь один порядок изменения значений ЦФ на всем множестве допустимых решений;
- быть монотонным, чтобы не менять правило предпочтений;
- допускать наличие и учитывать необходимость минимизации отклонений от f_i^* для всех ЦФ.

Это могут быть, например, преобразования вида $\varphi_i(\mathbf{x}) = [f_i(\mathbf{x}) - f_{i\min}] / [f_{i\max} - f_{i\min}]$, $\varphi_i(\mathbf{x}) = [f_{i\max} - f_i(\mathbf{x})] / [f_{i\max} - f_{i\min}]$, $\varphi_i(\mathbf{x}) = [f_{i\max} - f_i(\mathbf{x})] / f_{i\max}$ и т. п., где $f_{i\min}$ и $f_{i\max}$ – минимальное и максимальное значения $f_i(\mathbf{x})$ в области допустимых решений D . В дальнейшем по-прежнему будем писать f и только в случае необходимости указывать φ .

Принцип векторной оптимальности по Парето. Точка $\mathbf{x}^* \in D$ называется *эффективной*, или *оптимальной по Парето*, если не существует какой-либо точки $\mathbf{x}' \in D$, для которой $f_i(\mathbf{x}') \geq f_i(\mathbf{x}^*)$, $i = \overline{1, p}$, для задачи максимизации или $f_i(\mathbf{x}') \leq f_i(\mathbf{x}^*)$, $i = \overline{1, p}$, для задачи минимизации, причем хотя бы для одного i неравенство является строгим.

Поясним принцип. Если оптимальное решение \mathbf{x}^* найдено только в виде $\mathbf{x}^* = (x_1, \dots, x_i^*, \dots, x_n)$, т. е. только по i -му критерию для i -й ЦФ $f_i(\mathbf{x})$, то оно называется *субоптимальным решением*. Например, субоптимальное решение по скорости проектируемого самолета либо по грузоподъемности, либо по расходу топлива. Если для всех $i = \overline{1, p}$ и хотя бы для одного $i = k$ $f_k(\mathbf{x}') \geq f_k(\mathbf{x}'')$, то \mathbf{x}' называют *более предпочтительным решением по сравнению с \mathbf{x}''* . Иными словами, более предпочтительное решение \mathbf{x}' не хуже \mathbf{x}'' по всем рассматриваемым p критериям и лучше \mathbf{x}'' хотя бы по одному (k -му) из них. В этих понятиях эффективное решение \mathbf{x}^* – это такое, которое нельзя улучшить по какому-то из p критериев, не ухудшив при этом значения других критериев. Именно в этом и состоит суть принципа Парето. Множество эффективных решений часто называется *множеством Парето* и обозначается $P(D) \subset D \subset R^n$. Понятно, что для выбора единственного решения из $P(D)$ необходимо использовать дополнительный критерий предпочтения и селекции эффективных решений. В этом основной недостаток векторного подхода к оптимизации.

В качестве примера формирования дополнительных критериев рассмотрим два принципа их введения: компромисса и мажоритарности.

Принцип компромисса – выбирается такое (компромиссное x_0) эффективное решение, для которого взвешенные нормированные потери минимальны и равны между собой. Нормированные потери – это разность возможного наилучшего значения i -й нормированной ЦФ $\varphi_i(x_i^*)$ и значения этой функции для выбранного компромиссного решения $\varphi_i(x_0)$. Взвешенное – значит, учитываемое с весовым коэффициентом предпочтения μ_i ($0 \leq \mu_i \leq 1, \mu_1 + \dots + \mu_p = 1$).

Показано, что если x_0 – эффективное решение для данного вектора предпочтений $\mu = (\mu_1, \dots, \mu_p)$, то ему соответствует наименьшее значение δ , при котором система равенств $\mu_i \varphi_i(x_0) = \delta$ выполняется для всех $i = \overline{1, p}$. Одинаковость и минимизация значения потерь δ и составляют суть дополнительного (второго уровня) критерия. На этом уровне чаще всего используется метод ограничений, основанный на решении минимаксной задачи $\max_i \mu_i \varphi_i(x) \rightarrow \min_x$ для $x \in D$.

Принцип мажоритарности (ранжировки) критериев по значимости. Его суть: принятие оптимального решения в условиях многокритериальной оптимизации должно проводиться с учетом значимости (важности) каждого критерия для достижения главной цели исследования объекта. Исходя из этого принципа устанавливается дополнительный критерий и методы поиска наилучшего решения по нему. Например, в качестве дополнительного критерия можно выбрать невыход за допустимую зону отклонения от субоптимального значения ЦФ $f_i(x)$ (или $\varphi_i(x)$) по каждому i -му критерию первого уровня с учетом его значимости. В этом случае можно применить следующий часто используемый метод.

Метод уступок. Прежде всего все критерии (критериальные функции f_i) ранжируются по их важности и, в случае необходимости, перенумеровываются, например, по мере убывания важности: $f_{(1)}$ – самая важная, затем $f_{(2)}, \dots, f_{(p)}$. Потом находится первое субоптимальное значение $f_{(i)}^* = f_{(1)}(x_1^*)$, $i = \overline{1, p}$, где x_1^* – субопти-

мальное решение по наиболее важному первому критерию. Затем, исходя из постановки задачи исследования объекта, допускается приемлемая уступка – допустимое отклонение Δf_1 значения $f_{(1)}(\mathbf{x})$ от субоптимального $f_{(1)}^*$, когда будут отыскиваться субоптимальные решения по другим критериям. Эта уступка оформляется в виде дополнительного ограничения к постановке задачи: $f_{(1)}(\mathbf{x}) \geq f_{(1)}^* - \Delta f_1$. В новой постановке находится субоптимальное решение $f_2^* = f_{(2)}(x_2^*)$, для которого затем назначается уступка Δf_2 и т. д. Метод удобен тем, что по ходу решения сразу видно, какой уступкой по одному критерию достигаются лучшие решения по другим критериям. Однако такой метод не имеет строгой формализации выбора одного оптимального решения. Он позволяет существенно сузить множество решений, но не всегда дает возможность выявить единственное эффективное решение.

Поэтому чаще всего идут по второму пути – сведению многокритериальных задач к решению однокритериальных. Рассмотрим кратко некоторые из методов такого сведения.

Метод главного критерия. Его идея состоит в том, что выделяется главный k -й критерий из p критериев, а на остальные $p - 1$ критерии накладываются ограничения типа $f_i(\mathbf{x}) \geq a_i$ при максимизации или $f_i(\mathbf{x}) \leq b_i$ при минимизации для всех $i = \overline{1, p}$, $i \neq k$. Эти ограничения вводятся как дополнительные условия в p -критериальную задачу оптимизации. Тем самым она сводится к однокритериальной оптимизации по k -му критерию с учетом дополнительных ограничений. Его недостатки – учет лишь одного критерия и трудности назначения граничных значений – a, b .

Метод свертки соизмеримых критериев. Он основан на максимизации (минимизации) некоторой *монотонной* функции от частных критериев $f_i(\mathbf{x})$ или, лучше, $\varphi_i(\mathbf{x})$, $i = \overline{1, p}$:

$$F(\mathbf{x}) = f[f_1(\mathbf{x}), \dots, f_p(\mathbf{x})] \quad (4.198)$$

при сохранении для каждого из них ранее принятых ограничений. Чаще всего функция (4.198) является линейной, т. е. суммой типа

$$F(\mathbf{x}) = \sum_{i=1}^p \mu_i f_i(\mathbf{x}), \quad F(\mathbf{x}) = \sum_{i=1}^p \nu_i \varphi_i(\mathbf{x}), \quad (4.199)$$

$$\mu_i \geq 0, \quad \mu_1 + \dots + \mu_p = 1; \quad \nu_i \geq 0, \quad \nu_1 + \dots + \nu_p = 1. \quad (4.199a)$$

Сложность метода – в назначении весовых коэффициентов μ_i , ν_i . Именно равенство суммы весов единице с учетом возможно разной физической размерности ЦФ $f_i(\mathbf{x})$ подсказывает, что в (4.199) следует предпочесть правую сумму.

Метод идеальной точки. Идеальной называется точка $\mathbf{x}_0^* = (x_1^*, x_2^*, \dots, x_n^*)$, где x_i^* – оптимальное решение по i -му критерию. Тогда многокритериальную задачу можно свести к однокритериальной с критерием

$$\left[\sum_{i=1}^n \left[f(x_1^*, x_2^*, \dots, x_n^*) - f(x_1, x_2, \dots, x_n) \right]^2 \right]^{1/2} \Rightarrow \min_{\mathbf{x} \in D}. \quad (4.200)$$

4.5.6. О корректности постановок оптимизационных задач

Ранее уже упоминалось понятие *корректная постановка задачи* (см. подпараграф 2.1.1). Рассмотрим это понятие чуть подробнее, поскольку многие оптимизационные (экстремальные) задачи являются некорректными и требуют для своего решения определенных приемов регуляризации (например, по Тихонову), чтобы свести их к корректным.

Пусть \mathbf{x} – векторная ненаблюдаемая переменная некоторой модели, а \mathbf{y} – наблюдаемая выходная переменная. Предположим, что \mathbf{y} связано с \mathbf{x} соотношением $\mathbf{y} = \mathbf{A}\mathbf{x}$. Отображение \mathbf{A} считается заданным. Ставится задача: по имеющемуся приближенному значению $\tilde{\mathbf{y}}$ найти приближенное значение $\tilde{\mathbf{x}}$ переменной \mathbf{x} . Очень часто отображение \mathbf{A} является непрерывным, в то время как обратное к нему отображение \mathbf{A}^{-1} , т. е. $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$, не является непрерывным. Тогда могут возникнуть разные трудности при нахождении $\tilde{\mathbf{x}}$ в виде $\tilde{\mathbf{x}} = \mathbf{A}^{-1}\tilde{\mathbf{y}}$: если \mathbf{A}^{-1}

явно не задано, может быть разрывным на области своего определения, которое не всегда совпадает со всем пространством. Именно в этих ситуациях поставленная задача часто становится некорректной.

Задача нахождения вектора \mathbf{x} , принадлежащего D_X по заданному вектору $\mathbf{y} \in D_Y$ так, чтобы выполнялось равенство $A\mathbf{x} = \mathbf{y}$, называется *корректно поставленной*, если:

- 1) для любого вектора \mathbf{y} из множества Y решение существует;
- 2) решение $\tilde{\mathbf{x}}$ определено однозначно;
- 3) решение устойчиво (задача устойчива): для любого $\varepsilon > 0$ можно указать такое $\delta(\varepsilon) > 0$, при котором из неравенства $\|\tilde{A} - A\| < \delta$, $\|\tilde{\mathbf{y}} - \mathbf{y}\| < \delta$ вытекает, что решения \mathbf{x} и $\tilde{\mathbf{x}}$ соответствующих уравнений подчиняются неравенству $\|\tilde{\mathbf{x}} - \mathbf{x}\| < \varepsilon$, где $\|\cdot\|$ – евклидова норма вектора в рассматриваемом пространстве \mathbf{x}, \mathbf{y} .

Задача, не удовлетворяющая перечисленным требованиям 1–3, называется *некорректно поставленной*.

С методами решения некорректно поставленных задач вы ознакомитесь в дисциплинах, связанных с вычислительной математикой, численными методами.

4.5.7. Практические примеры оптимизационных задач

Заключая рассмотрение оптимизационных моделей, приведем схему технологического процесса решения оптимизационных задач и практические примеры, в которых используются эти модели.

Схема приведена на рис. 4.21 (ср. с рис. 3.10–3.13).

Примеры ЛП-задач: о диете (пищевом рационе), о рюкзаке туриста, транспортная, о назначении исполнителей работ, коммивояжере, раскрое материалов, об управлении выбором цели, о планировании производства продукции и простейших работ, распределении производственной программы.

Примеры НП-задач (в частности КП): о смешивании компонентов при производстве продуктов, о планировании работы в последовательных и иерархических структурах, об управлении запасами, о выборе портфеля ценных бумаг в условиях обобщенного инвестиционного риска.



Рис. 4.21. Укрупненная схема технологического процесса решения оптимизационных задач

4.5.8. Вариационные модели

Особенность *оптимизационных моделей* в том, что они базируются на определении экстремальных значений некоторых целевых *функций* $f(x)$. **Вариационные модели** основаны на нахождении экстремальных (максимальных или минимальных) значений *функционалов*¹.

¹ См. понятие функций и функционалов в § 4.2.

Задачи, связанные с исследованием функционалов на экстремумы, а также с определением функций, на которых эти экстремумы достигаются, называются *вариационными задачами* и изучаются в *вариационном исчислении*. Отцом вариационного исчисления как раздела математики считается Леонард Эйлер¹. Он первым в 1744 году начал развивать его.

Необходимым условием оптимальности решения при заданных условиях (ограничениях) в оптимизационных задачах является существование экстремумов целевых функций в этих условиях. Необходимое же условие оптимального решения в вариационных задачах – наличие (условие наличия) экстремума функционала в рамках заданных ограничений. В связи с этим методы вариационного исчисления можно условно разбить на три группы.

Первая группа (*прямые методы*) основана на непосредственном отыскании экстремума функционала. Это методы типа ломаных Эйлера, Ритца, динамического программирования (последовательного анализа вариантов), использующие идеи спуска в пространстве управлений и т. д. Вторая группа (*непрямые методы*) основана на использовании необходимых условий оптимальности, с помощью которых исходная вариационная задача сводится к краевой. Это методы, построенные на идеях принципа максимума Понтрягина² (1956). Наконец, третью группу составляют те методы, которые *не входят в первые две группы*. Это, например, методы, основанные на достаточных условиях оптимальности.

Рассмотрим некоторые из задач вариационного исчисления на уровне их постановок.

Задача о безусловном экстремуме функционала

Эта задача является одной из простейших в вариационном исчислении.

Задан функционал (!)

$$J(y) = \int_{x_1}^{x_2} f[x, y(x), y'(x)] dx, \quad (4.201)$$

¹ Л. Эйлер родился в 1707 году в Базеле (Швейцария), в 1727–1741, 1766–1783 годах работал в Петербурге (РАН), 1742–1766 – в Берлине. В последние 17 лет жизни, ослепнув, почти удвоил свое научное наследие (более 800 работ), диктуя сочинения сыну и двум ассистентам до их полного изнеможения.

² Лев Семенович Понтрягин (1908–1988) потерял зрение в 13 лет.

где f – непрерывно дифференцируемая по всем аргументам функция, зависящая от искомой непрерывно дифференцируемой функции $y(x)$ и ее производной $y'(x)$; x – произвольная вещественная переменная, в частности временная $x = t$. Требуется найти такую функцию $y(x)$, которая доставляет экстремум функционалу (4.201) и удовлетворяет следующим условиям:

$$y(x_1) = y_1, \quad y(x_2) = y_2. \quad (4.202)$$

По аналогии с оптимизационными задачами непрерывные дифференцируемые функции, удовлетворяющие краевым условиям (4.202), **называются допустимыми**.

Достижение экстремума (4.201) при ограничениях (4.202) и есть критерий оптимальности в этой вариационной задаче. Именно из-за отсутствия других условий она называется задачей о безусловном экстремуме.

Именно Л. Эйлер в 1744 году опубликовал теорему, ставшую основой вариационного исчисления: чтобы функция y обеспечивала экстремум J (4.201), она должна удовлетворять функциональному уравнению

$$\frac{df}{du} - \frac{d}{dx} \left(\frac{\partial f}{\partial u'} \right) = 0 \quad (\text{уравнение Эйлера}).$$

Другие необходимые условия вывели А. Лежандр (1776), К. Якоби (1837), К. Вейерштрасс (1879) и др. Л. Эйлер предложил искать решение этой задачи в классе непрерывных ломаных $x(t)$, состоящих из N прямолинейных отрезков с заданными абсциссами концов. Тогда функционал $J(x)$ превращается в функцию от $N+1$ ординат вершин этих ломаных, а исходная вариационная задача – в задачу нахождения экстремума функции многих переменных. Одним из широко применяемых методов решения задачи об условном экстремуме функционала, когда вводятся дополнительные к указанным выше ограничения, является метод Лагранжа. Он позволяет свести исходную задачу к простейшей задаче вариационного исчисления путем перехода к новому критерию (функционалу) оптимальности.

Из непрямых методов, используемых чаще всего для неклассических задач вариационного исчисления, рассмотрим методы, основанные на уже упомянутом **принципе максимума Понтрягина** (1956). Изначально такие методы были ориентированы на решение задач оптимального управления, когда переменная x есть время t . Именно в этой редакции рассмотрим постановки задач, обобщающих более

ранние вариационные задачи за счет расширения класса искомых оптимальных функций (решений) до кусочно-непрерывных¹ и учитывающих ограничения неклассического типа. Принцип максимума дает общие необходимые условия оптимальности управления и позволяет сводить вариационные задачи к краевым. Рассмотрим одну из формулировок принципа максимума.

Пусть требуется найти траекторию $x^*(t)$, управление $u^*(t)$ и временной отсчет t^* , обеспечивающие минимум функционалу

$$J = \int_{t_0}^{t_1} f_0(t, \mathbf{x}(t), \mathbf{u}(t)) dt \quad (4.203)$$

при условиях дифференциальной связи

$$d\mathbf{x}(t) / dt = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t)), \quad (4.204)$$

при граничных (краевых) условиях

$$\mathbf{x}(t_0) = \mathbf{x}_0, \quad \mathbf{x}(t_1) = \mathbf{x}_1, \quad (4.205)$$

а также ограничениях на координаты состояний (траекторий) $\mathbf{x}(t)$ и управление $\mathbf{u}(t)$:

$$\mathbf{x} \in X, \quad \mathbf{u} \in U, \quad (4.206)$$

где $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))$, $\mathbf{u}(t) = (u_1(t), \dots, u_m(t))$; $\mathbf{f} = (f_1, \dots, f_n)$; X и U – замкнутые множества допустимых значений n - и m -мерного пространства; t – временная переменная.

В этих условиях принцип максимума Понтрягина означает: управление $\mathbf{u}(t)$ будет оптимальным, если при каждом t оно доставит абсолютный максимум *функции Гамильтона* (!):

$$H(\tilde{\mathbf{u}}) = \max_{\mathbf{u} \in U} H(\mathbf{u}) =$$

¹ Функция $f(x)$ называется кусочно-непрерывной на отрезке $[x_1, x_2]$, если она непрерывна во всех точках $x \in [x_1, x_2]$, за исключением конечного числа точек этого отрезка, в которых она может иметь разрывы I рода, т. е. такие, когда *конечные* пределы слева и справа в этих точках не равны друг другу.

$$= \max_{\mathbf{u} \in U} \left[\sum_{i=1}^n \psi_i f_i(t, \mathbf{x}(t), \mathbf{u}(t)) - f_0(t, \mathbf{x}(t), \mathbf{u}(t)) \right], \quad (4.207)$$

в которой $\Psi = (\psi_1, \dots, \psi_i, \dots, \psi_n)$ определяется системой уравнений

$$\psi_i = -\partial H / \partial x_i, \quad i = \overline{1, n}. \quad (4.208)$$

Находя из (4.206) управление $\mathbf{u}(t, \mathbf{x}, \Psi)$ и подставляя его в (4.204) и (4.208), получаем замкнутую краевую задачу для системы из $2n$ дифференциальных уравнений (4.204) и (4.208) с граничными условиями (4.205).

Рассмотренные краевые условия (4.205) не единственные. В настоящее время исследованы и другие краевые условия:

1) когда x_0, x_1 – единственные точки, не зависящие от t_0, t_1 . В этом случае говорят, что *концы траектории закреплены*. При этом t_0 и t_1 заданы, фиксированы или не фиксированы (задано только t_0 или t_1);

2) концы траектории не закреплены. Тогда ставится задача оптимального управления со *свободным* левым или правым концом. Множество задач подобного класса определяется комбинацией концов и фиксацией или нефиксацией времени t_0, t_1 ;

3) задачи с подвижными концами траекторий. При этом один или оба из моментов t_0, t_1 могут быть не фиксированы.

Все многообразие комбинаций, когда t_0 и $t_1, x(t_0)$ и $x(t_1)$ фиксированы или подвижны (свободны), читатель может представить сам.

Заметим, что рассмотренное управление, когда $\mathbf{u}(t)$ и оптимальное $\mathbf{u}^*(t)$ зависят только от t , относится к временному *программному управлению*. Помимо него часто рассматривается *управление с обратной связью*, когда $\mathbf{u}(t)$ и, следовательно, оптимальное $\mathbf{u}^*(t)$ управления зависят не только от t , но и от координат состояний $\mathbf{x}(t)$, т. е. $\mathbf{u} = \mathbf{u}(t, \mathbf{x}(t))$.

Принцип оптимальности Беллмана. Динамическое программирование

Все предыдущие оптимизационные задачи рассматривались в предположении (по умолчанию), что исследуемый объект является

статичным (стационарным). Однако более общий и близкий к реальности случай, когда объект является *динамичным, нестационарным*, это означает, что переменные модели, описывающие такие объекты, должны зависеть от времени¹. В этих случаях типичные задачи математического программирования и вариационного исчисления заменяются задачами математического динамического программирования (ДП-задачами).

Динамическое программирование – метод оптимизации многошаговых (динамических) процессов управления и принятия решений, основанный на следующих положениях:

- 1) интуитивном принципе оптимальности Беллмана;
- 2) приеме инвариантного погружения;
- 3) методе функциональных уравнений Беллмана, построенных на основании погружения.

Обратим внимание на многошаговость решения задачи управления. Оно может либо отражать реальный многошаговый процесс принятия управленческих решений, либо вводиться искусственно, например, для сведения решения задачи в пространстве высокой размерности к задачам небольшой размерности. При этом последовательность допустимых решений на отдельных шагах называется *политикой*. Динамическое программирование направлено на поиск среди всех политик оптимальной.

Общая схема многошагового процесса принятия решений в ДП состоит в следующем. Имеется некоторый объект (управляемая система S), состояние которой в начальный момент времени t_0 есть x_0 . Разобьем процесс управления объектом на N полуинтервалов $[t_0, t_1)$, $[t_1, t_2)$, ..., $[t_{N-1}, t_N]$. **В каждый момент t_k принимается решение γ_k о том, как оптимально перевести объект (систему S) в состояние x_N .** Причем каждое решение, конечно, должно удовлетворять как ограничениям исходной постановки задачи, так и ограничениям, возникающим за счет ранее принятых в t_1, \dots, t_k решений.

Понятно, что тогда состояние объекта (системы S) в конце k -го шага, $k = \overline{1, N}$, будет зависеть только от предшествующего состояния его

¹ Понятно, что динамика объекта может рассматриваться не только во времени, но и в пространстве, по давлению, весу и другим физическим величинам. Читатель может сам рассмотреть особенности постановок задач динамического программирования для этих случаев.

x_{k-1} и управления u_k , принятого на k -м шаге. Ясно, что эта схема соответствует гипотезе *отсутствия последствия* от ранее принятых решений и может быть изображена как

$$x_0 \xrightarrow{u_1} x_1 \xrightarrow{u_2} x_2 \longrightarrow \dots \xrightarrow{u_N} x_N.$$

Еще раз укажем на два момента.

1. Согласно принципу оптимальности оптимальное управление в любой момент времени не зависит от предыстории объекта. Оно определяется только состоянием объекта в этот момент времени и *целью управления*.

2. Оптимальным является конечный участок оптимальной траектории, но не промежуточный. В качестве примера сам Р. Беллман приводит оптимальное самоуправление бегуна на дальнюю дистанцию. Его исходная цель на старте – пройти всю дистанцию за минимальное время. В ходе бега он не ставит задачу пройти каждый участок дистанции за минимальное время. В ходе бега он определяет свое состояние и возможности и распределяет силы так, чтобы с учетом состояния в настоящий момент бега *оставиющуюся* часть дистанции преодолеть за минимальное время. Тем самым он придерживается принципа оптимальности.

Принцип оптимальности Беллмана: оптимальное управление объектом (поведение объекта или оптимальная политика) – это такое, при котором, каковы бы ни были начальное состояние и принятое начальное решение (политика), последующие решения должны составлять оптимальное управление (поведение, политику) относительно состояния, возникшего в результате первоначального решения.

Принцип инвариантного погружения. Инвариантное – это независимое от числа шагов N , сведение задачи, ранее заданной для многих переменных большой размерности, в пошаговое семейство оптимизационных задач желательно как можно меньшей размерности. Идея принципа инвариантного погружения (вложения) сводится к тому, что вместо решения искомой задачи с заданным начальным состоянием x_0 объекта и числом n шагов ее решения решается семейство задач с произвольным начальным состоянием x_k и с произвольным числом шагов N .

Метод функциональных управлений – это метод, основанный на идее сведения поиска искомой оптимальной функции к системе уравнений, в которых искомые частные функции связаны с известными

функциями одной или нескольких переменных при помощи операции образования сложной функции. Например, $f(x+y) = f(x) + f(y)$; $f(x+y) = f(x)f(y)$. Наиболее часто функциональные уравнения ДП представляют собой систему рекуррентных уравнений.

Применение принципа инвариантности и метода функциональных уравнений с его рекуррентной реализацией позволяет, во-первых, свести ДП-задачу к дискретной модели динамического программирования, когда $\mathbf{x}_k = \varphi(\mathbf{x}_{k-1}, \mathbf{u}_k)$, где φ считается определенной (заданной) функцией; во-вторых, свести отыскание экстремума сложной функции многих переменных к операции последовательной экстремизации функций одной переменной. Вместо того чтобы один раз решать сложную задачу, решать N раз задачу более простую, работая всегда на перспективу. Метод поиска оптимального решения сводится к такому выбору оптимальной политики, т. е. к такому варьируемому набору управлений $(\mathbf{u}_1, \dots, \mathbf{u}_k, \dots, \mathbf{u}_N)$, для которого при заданном начальном состоянии \mathbf{x}_0 будет достигнуто требуемое экстремальное значение целевой функции перевода объекта (системы S) в финальное состояние \mathbf{x}_N . Заметим, что при этом в ряде ДП-задач надо, чтобы траектория движения от \mathbf{x}_0 к \mathbf{x}_N лежала в некоторой области. В других ДП-задачах – траектория движения $\mathbf{x}(t)$ не столь интересна. Важно, чтобы конец процесса был нужным. При непрерывном t шаги означают движение от t_0 к t_1 , затем от t_1 к t_2 , ..., от t_{N-1} к $t_N = T$, $t \in (t_0, T]$ (см. (4.203)–(4.206)) при $t_1 = T$, когда вид и/или параметры $\mathbf{x}(t)$, $\mathbf{u}(t)$ меняются на каждом интервале (t_k, t_{k+1}) , $k = \overline{0, N-1}$. При дискретном управлении число состояний $\mathbf{x}(t)$ и управленческих воздействий конечно.

Конкретные методы решения ДП-задач описаны в специальной литературе.

Помимо учета динамики изменения объекта отметим еще одну особенность ДП-задач: при постановке таких задач и схемы их решения не накладывается каких-либо специальных ограничений на используемые в них функции (на их природу и характер типа линейности, непрерывности, выпуклости и т. п.). Ограничения касаются лишь сферы применимости методов ДП.

Из изложенного ясно, что при априорном пошаговом дискретном решении ДП-задачи следует ориентироваться лишь на итоговый результат x_N , а не на узкие интересы каждого шага. Следовательно, начинать ее решение надо с последнего шага. Это единственный шаг, который просто планировать, поскольку за ним ничего не стоит, в нем не надо «оглядываться» на будущее. Но последний шаг можно сделать лишь на основе предположений об ожидаемых итогах, исходах x_{N-1} предшествующего $(N-1)$ -го шага. Для каждого исхода предшествующего шага можно выбрать условно оптимальное решение на этом шаге. Затем аналогично решаем задачу для предпоследнего $(N-2)$ -го шага и т. д., вплоть до первого шага. Прodelав такую итерацию N раз, мы каждый раз будем получать условно оптимальное решение, учитывающее разные начальные условия (состояния) x_k на каждом k -м шаге и являющееся условно оптимальным для всех $(k+1)$, $(k+2)$, N последующих, завершающих управление, шагов. После этого движемся в прямом направлении от x_0 к x_N , реализуя оптимальное управление.

Такая процедура принятия управленческих решений изображена на рис. 4.22.

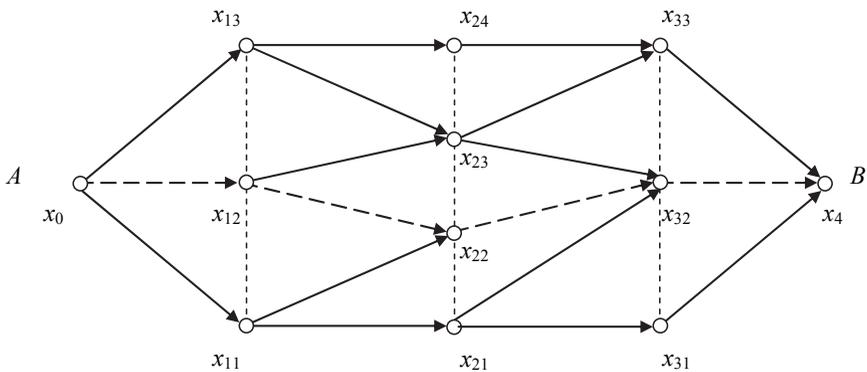


Рис. 4.22. Пример карты дорог и траекторий $x(t)$ движения от пункта А в пункт В

Попробуйте, двигаясь визуальнo от пункта В влево, определить (они очевидны) условно оптимальные пути движения из состояний (из мотелей) x_{31} , x_{32} , x_{33} в пункт В. Теперь мысленно переместимся

в состояния (мотели) x_{21} , x_{22} , x_{23} , x_{24} и выберем условно оптимальные пути движения от них к пункту B . Здесь уже необходимы сравнения. Например, какой путь $x_{21} \rightarrow x_{31} \rightarrow x_4$ или $x_{21} \rightarrow x_{32} \rightarrow x_4$ будет условно оптимальным? Ясно, тот, сумма расстояний в котором будет меньше. То же характерно для $x_{23} \rightarrow x_{32} \rightarrow x_4$ или $x_{23} \rightarrow x_{33} \rightarrow x_4$. Еще более ситуация усложнится, если мы условно переместимся в мотели $x_{1,1}$, $x_{1,2}$, $x_{1,3}$. Наконец, условно переместившись в пункт A (состояние x_0), передвигаемся в прямом направлении и находим из условно оптимальных оптимальную траекторию (путь) движения $x_0 \rightarrow x_{12} \rightarrow x_{22} \rightarrow x_{32} \rightarrow x_4$, если, конечно, графическое изображение на рис. 4.22 действительно соответствует минимальному расстоянию от A до B через все возможные дороги и мотели. Теперь начинаем реальное прямое движение от A к B . Это описание «работает» при условии, что пассажирам автомобиля удалось забронировать все ночевки в мотелях по ходу движения. А как определить оптимальный путь, когда придется ориентироваться на свободный на предстоящую ночь мотель или дожждаться освобождения места в ближайшем мотеле?

Внимание! Попробуйте решить такую задачу самостоятельно, с учетом того, в каком состоянии (мотеле) находится автомобиль и какая для этого состояния ранее была найдена условно оптимальная траектория.

4.5.9. Состязательные модели. Теория игр. Теория принятия оптимальных решений в условиях неопределенности, конфликтов

Мы уже неоднократно сталкивались с задачами выбора решений. Научный метод подготовки и выбора обоснованных рекомендаций по принятию решений называется *исследованием операций*. В более общей постановке исследование операций есть теория принятия именно оптимальных решений. Рассмотренные ранее методы оптимизации и вариационного исчисления составляют математическую основу принятия оптимальных решений, когда имеется полная определенность в едином скалярном или векторном критерии оптимальности и усилия любого исследователя направлены на достижение оптимальности именно по этому критерию. Однако существует большой класс ситуаций, когда желательно иметь математические модели и, используя их, находить количественные оптимальные решения в других условиях. А именно, когда пошаговый поиск оптимального решения осуществляют двое или более участников поиска, имеющих несовпадающие

интересы и, как следствие, скрывающих свои пошаговые действия, порождая тем самым соответствующую неопределенность для других участников. Такие ситуации принято называть **конфликтными**, само явление, связанное с участием в решении задачи нескольких участников с несовпадающими интересами, – **конфликтом**, а активных участников – **игроками**. Раздел математики, занимающийся разработкой и изучением моделей принятия оптимальных решений в условиях конфликта (неопределенностей), называют **теорией игр**, а сами модели – **моделями теории игр**, игровыми моделями, моделями состязаний или **состязательными моделями**.

Заметим, что реальный конфликт отличается от абстрактного математического представления его, по крайней мере, следующими аспектами:

- 1) наличием ограничения на ресурсы и на управление ими и конфликтом (игрой);
- 2) наличием размерных физических величин;
- 3) реальным проявлением совокупности разных видов неопределенностей, только часть из которых была рассмотрена выше.

Наконец, заметим, что игра как математическая модель отличается от реального конфликта тем, что она всегда ведется по определенным правилам. Они устанавливают права, обязанности и ответственность игроков (участников конфликта – игры), а также понимание исхода игры – выигрыш (в заданном смысле, по заданному критерию), проигрыш, ничья для каждого участника в зависимости от сложившейся обстановки.

Основная проблематика теории игр как моделей конфликтных ситуаций связана с решением трех ее внутренних задач:

- 1) выработка принципов и критериев оптимальности для разных классов игр;
- 2) исследование реализуемости выбранного принципа и достижения критерия оптимальности, т. е. доказательство реализуемости игры – существования у игры таких исходов ее, которые удовлетворяют выработанным принципам и принятым критериям оптимальности в идеальных и реальных условиях;
- 3) явное указание, как найти эти реализации, т. е. фактическое решение игры – нахождение разных реализаций игры (лучше всего их полного набора), удовлетворяющих заданным условиям.

Иными словами, теория игр занимается описанием, исследованием оптимальности игр, поиском всех реализаций решения игры и исследе-

дованием их свойств, а также условий, в которых реализуемость имеет или может иметь место. Поскольку в общем виде для всех возможных конфликтов их формальное игровое представление вряд ли возможно и оправданно, обычно ограничиваются рассмотрением определенных классов игр.

Рассмотрим их вкратце на идейном уровне без подробного описания методов решений. Для этого предварительно введем необходимые понятия.

Игра развивается во времени ходами ее игроков. *Ходом* называется выбор и осуществление игроком очередного действия, допустимого (предусмотренного) *правилами игры*. Если ход сознательно выбирается и осуществляется самим игроком, он называется *личным*. Его альтернатива – *случайный ход*, который определяется каким-либо естественным, природным или искусственным механизмом рандомизации – случайного выбора, отбора. Пример таких игр – азартные игры. Теория игр, как правило, рассматривает только такие, в которых можно оптимизировать поведение игрока в игре, в том числе при частичном наличии случайных ходов. Такие игры принято называть стратегическими.

Стратегия игрока – совокупность правил (указаний, инструкций, способов), определяющих вариант действия игрока при каждом его личном ходе или коалицию действий игроков в зависимости от сложившейся ситуации, т. е. правил, которыми руководствуются игроки, делая очередной ход. Тем самым стратегия игры определяет ее конкретную реализацию.

В зависимости от числа стратегий и вытекающих из них вариантов действий игры делятся на классы *конечных* и *бесконечных игр*. Стратегия игрока, обеспечивающая наилучший исход игры для него, т. е. максимальный в заданном смысле, по задуманному критерию, результат (выигрыш), или максимум среднего выигрыша игрока при многократном повторении игры, называется *оптимальной* (по соответствующему критерию).

Если стратегия однозначно выбирается игроком из ранее намеченных, т. е. обеспечивает полную определенность выбора ходов и, следовательно, продолжения игроком игры после каждого хода, то она называется *чистой*. Отсюда следует, что выбор игроком чистой стратегии означает выбор им конкретной реализации решения игры. Игры, в которых все игроки придерживаются чистых стратегий, называются *играми в чистых стратегиях*.

Если поведение игрока во всех очередных или в совокупности разных ходов непредсказуемо, например, определяется каким-либо случайным механизмом, то это означает, что он пользуется **смешанной стратегией**. Смешанные стратегии используются, когда игрок придерживается гибкой тактики, не знает, как поведут себя другие игроки (или один игрок-противник) в данной игре, хочет скрыть свою стратегию и т. д. Поскольку одним из лучших способов скрыть свои намерения от противника – это придать своим действиям случайный характер, когда даже сам не знаешь заранее, как поступишь в конкретной ситуации, чаще всего модельно в качестве смешанных рассматривают стратегии, основанные на их случайном (рандомизированном) выборе. Именно такие смешанные стратегии изучены в настоящее время, хотя понятно, что можно рассматривать смешанные стратегии и при других модельных представлениях неопределенности выбора действий (нечетких, экспертных, псевдослучайных, квазислучайных, интервальных и т. п.).

Согласно введенному определению смешанной стратегии, **игры со смешанными стратегиями** предполагают наличие в них комбинаций чистых и смешанных стратегий, в том числе когда отдельная смешанная стратегия может состоять из фрагментарно набранных кусков чистых стратегий. Однако чаще всего в теории игр рассматриваются простейшие игры в смешанных стратегиях, когда каждый игрок заранее случайным образом выбирает чистую стратегию из всего возможного пространства его чистых стратегий. Тогда игры в чистых стратегиях являются частным случаем игр в смешанных стратегиях, если вероятность выбора отдельных стратегий из пространства чистых принимает только значения 1 или 0. Понятно, что для таких игр в смешанных стратегиях игра приобретает случайный характер. В них значения выигрышей и проигрышей каждого игрока надо описывать случайными величинами и, следовательно, находить средние или медианные значения выигрышей и проигрышей, разбросы и доверительные интервалы их значений и т. д. Ясно, что в таких рандомизированных смешанных стратегиях случайный выбор игроком стратегии означает случайный выбор чистой стратегии и, как следствие, реализации хода (решения) игры.

Если сумма выигрышей всех игроков равна нулю, то имеет место **игра с нулевой суммой**. Ясно, что в такой игре каждый игрок может получить выигрыш только за счет соответствующего проигрыша других игроков.

Наиболее разработана теория **бескоалиционных игр**. Это такие игры, каждая из которых характеризуется следующими компонентами:

- конечным множеством своих игроков $I = (1, 2, \dots, n)$;
- последовательностью ходов, что составляет всю игру;
- множеством стратегий $\{S_i\}$ каждого i -го игрока, $i \in I$, составляющих семейство стратегий всех игроков;
- семейством выигрышей (функций выигрыша) $\{H_i\}$, $i \in I$, задаваемых на множестве всех игровых ситуаций.

Отказ от конечности числа игроков приводит к классу игр с **бесконечным множеством игроков**.

Игры, в которых их правилами некоторые игровые ситуации не допускаются, образуют класс **игр с запрещенными ситуациями**.

Перечислены лишь некоторые классы игр. В свою очередь классы игр разделяются на их частные разновидности – подклассы. Так, парные ($n = 2$) конечные бесконфликтные игры, т. е. игры двух игроков с конечным множеством стратегий, называются **антагонистическими**, или **со строгим соперничеством**. В антагонистических играх с нулевой суммой выигрыш одного (назовем его первым и обозначим буквой A) игрока с суммой a_1 означает проигрыш ему второго игрока (B) с суммой $a_2 = -a_1$. В связи с этим при исследовании таких игр можно рассматривать только выигрыш a_1 игрока A (или проигрыш $a_2 = -a_1$ игрока B). Понятно, что игрок A будет стараться максимизировать a , в то время как игрок B – минимизировать a .

Игры с нулевой суммой являются частным случаем **игр с постоянной суммой**. В них постоянным, не обязательно нулевым, является общий выигрыш $a = a_1 + a_2$ игры и a_2 не обязана равняться $-a_1$, как в антагонистических играх с нулевой суммой.

В качестве примера игр рассмотрим матричную математическую модель антагонистических конечных парных игр с нулевой суммой – **матричные игры**. Из изложенного следует, что матричная игра – это игра, в которой участвуют два игрока A и B , имеющие, во-первых, противоположные (отсюда слово антагонистические) интересы; во-вторых, конечное число чистых стратегий: $S_A = (A_1, A_2, \dots, A_n)$ для игрока A и $S_B = (B_1, B_2, \dots, B_m)$ для игрока B , где n, m – конечные натуральные числа. Обозначим a_{ij} выигрыш игрока A , выбравшего страте-

гию A_i , $i = \overline{1, n}$, когда игрок B выбрал стратегию B_j , $j = \overline{1, m}$. Если a_{ij} известны до начала игры, то игру \mathbf{I} можно свести к матричной форме $\mathbf{I}(n \times m)$, составив прямоугольную **платежную** матрицу $\mathbf{\Pi}(n \times m)$ (табл. 4.14) размера $(n \times m)$ стратегий S и выигрышей a_{ij} .

Т а б л и ц а 4.14

Платежная матрица

	B_1	B_2	...	B_j	...	B_m
A_1	a_{11}	a_{12}		a_{1j}		a_{1m}
A_2	a_{21}	a_{22}		a_{2j}		a_{2m}
\vdots						
A_i	a_{i1}	a_{i2}		a_{ij}		a_{im}
\vdots						
A_n	a_{n1}	a_{n2}		a_{nj}		a_{nm}

Клетка (точка) матрицы, т. е. пара (i_0, j_0) номера строки и столбца, для которых при любых $i = \overline{1, n}$ и $j = \overline{1, m}$ верны неравенства

$$a_{ij_0} \leq a_{i_0j_0} \leq a_{i_0j}, \tag{4.209}$$

называется **седловой точкой**¹ платежной матрицы $\mathbf{\Pi}(n \times m)$.

На примере матричных игр в чистых стратегиях рассмотрим, как решают упомянутые ранее три внутренние задачи теории игр.

Задача 1. Выработка принципов и критериев оптимальности решения игры.

¹ В геометрии седловой точкой гладкой поверхности называется такая, в которой одновременно достигается минимум z по одной координате (x или y) и максимум по другой (y или x) (вспомни лошадиное седло). Иными словами, это точка, вблизи которой поверхности лежат по разные стороны от своей касательной плоскости.

Для данного класса задач наиболее часто в качестве принципа оптимальности используется два из них: **принцип минимакса** (максимального результата) и **принцип максимина** (гарантированного результата). Принцип минимакса предписывает игрокам выбирать такие рискованные стратегии, на которых достигается максимальный гарантированный выигрыш. Поэтому иногда его называют принципом наибольшего гарантированного результата (выигрыша). Иными словами, принцип минимакса означает: поступай каждый раз так, чтобы при наихудшем для тебя поведении противника получить максимальный выигрыш или, другими словами, выбирай такую стратегию, при которой твой минимальный выигрыш будет максимальным.

В отличие от принципа минимакса принцип максимина ориентирует игрока выбирать такие стратегии, на которых достигается наибольший выигрыш из тех минимально возможных, которые могут быть в наихудших для тебя условиях. Такая осторожная стратегия обеспечит тебе гарантированный выигрыш. Эти принципы основаны на соотношениях максимина и минимакса. Согласно ему для любой функции (!) $f(x, y)$, $x \in X$, $y \in Y$, справедливо неравенство

$$\max_{x \in X} \min_{y \in Y} f(x, y) \leq \min_{y \in Y} \max_{x \in X} f(x, y). \quad (4.210)$$

Предположим, что противники вездесущи, могут угадывать ходы другого, ориентируются каждый на оптимальную стратегию.

Введем следующие обозначения. Прежде всего заменим a_{ij} на $a_{i,j}$. Если на каждую лучшую чистую стратегию A_i , $i = \overline{1, n}$, игрока A игрок B отвечает такой своей чистой стратегией $j(i)$, $j = \overline{1, m}$, что выигрыш (плата, цена) $a_{i,j(i)} \leq a_{i,j}$ для любой стратегии B_j , $j = \overline{1, m}$, то такую его стратегию назовем наилучшей. Пусть $\alpha_i = a_{i,j(i)} = \min_{1 \leq j \leq m} a_{i,j}$, $i = \overline{1, n}$. Тогда наилучшая *максиминная стратегия* первого игрока A A_{i_0} (или просто i_0) будет такая, которая максимизирует α_i для $\forall i \in [1, n]$, т. е. обеспечивает

$$\max_{1 \leq i \leq n} \min_{1 \leq j \leq m} a_{i,j} = a_{ij} = \alpha_{i_0} = \alpha. \quad (4.211)$$

Величина α называется *нижней ценой игры в чистых стратегиях* и определяет тот гарантированный выигрыш (ту минимальную цену, плату), ниже которого (которой) игрок A , принимающий максиминную стратегию (в нашем примере игрок A), получит при любых действиях противника – второго игрока.

Если игрок, например B , выбирает *минимаксную стратегию* j_0 , то он должен минимизировать величину $\beta_j = a_{i(j),j} = \max_{1 \leq i \leq n} a_{i,j}$, т. е. обеспечить величину β

$$\min_{1 \leq j \leq m} \max_{1 \leq i \leq n} a_{i,j} = \beta_{j_0} = \beta, \quad (4.212)$$

которая называется *верхней ценой игры в чистых стратегиях*. Отсюда следует, что α есть гарантированный выигрыш, который игрок, ведя себя осторожно, получит в игре, т. е. меньше которого выигрыш быть не может. В то же время β есть то значение выигрыша, больше которого разумный противник заведомо не отдаст противнику. Из принципов оптимальности и соотношений (4.211), (4.212) следуют максиминный и минимаксный критерии оптимальности.

Задача 2. Исследование реализуемости принципов оптимальности.

В основе доказательства реализуемости рассмотренных принципов оптимальности лежат два обстоятельства.

Первое связано с тем, что для парных конечных игр в чистых стратегиях всегда $\alpha \leq \beta$, т. е. выигрыш, который может обеспечить себе игрок A , не может превысить проигрыш, который будет иметь второй игрок. Необходимым и достаточным условием равенства нижней цены α и верхней цены β матричной игры в чистых стратегиях является существование седловой точки (i_0, j_0) платежной матрицы игры (см. (4.209)). Именно седловая точка (i_0, j_0) обеспечивает в таких играх *положение равновесия*, когда ни одному из игроков невыгодно (из-за неустойчивости) отклоняться от его оптимальной стратегии, если противник придерживается своей оптимальной стратегии. Если же $\alpha \neq \beta$, то в играх рассматриваемого класса положения равновесия не существует и, следовательно, оптимального решения игры найти (построить) не удастся. Тогда необходимо либо руководствоваться принципом минимакса, либо переходить к смешанным стратегиям.

Второе обстоятельство связано с понятием *игры с полной информацией*. Ими называются игры, в которых каждый игрок при каждом

личном ходе знает всю предысторию ее развития, все результаты предыдущих ходов (личных и случайных). Пример – шахматы, шашки, «крестики-нолики» и т. д. В теории игр доказано, что любая игра с полной информацией имеет седловую точку. Это означает, что она может быть решена в чистых стратегиях, т. е. для нее существует пара оптимальных стратегий S_A и S_B игроков A и B , дающих устойчивый выигрыш, равный цене игры γ , где $\gamma = \alpha = \beta$.

Задача 3. Поиск решений.

Из изложенного при рассмотрении задачи 2 ясно указание по выбору оптимальной стратегии – это стратегии седловой точки (A_{j_0}, B_{j_0}) .

Если же игра такова, что $\alpha < \beta$, т. е. седловой точки нет, то нужно выбрать принцип минимакса или перейти к смешанным стратегиям в частной их форме рандомизированных чистых стратегий. Отличие игр в рандомизированных (случайным образом выбираемых чистых стратегиях) смешанных стратегий от рассмотренных игр в чистых стратегиях в следующем. В играх с чистой стратегией игроком A однозначно, наверняка выбирается какая-то одна чистая стратегия A_i из n стратегий A_1, A_2, \dots, A_n , а игроком B – также однозначно чистая стратегия B_j из B_1, B_2, \dots, B_m . В рандомизированных же смешанных стратегиях игроком A выбирается любая из чистых стратегий $A_1, \dots, A_i, \dots, A_n$ с вероятностью $p_1, \dots, p_i, \dots, p_n$, $p_i \geq 0$, $p_1 + p_2 + \dots + p_n = 1$, а игроком B – какая-то чистая стратегия $B_1, \dots, B_j, \dots, B_m$ с соответствующей вероятностью $q_1, \dots, q_j, \dots, q_m$, $q_j \geq 0$, $q_1 + q_2 + \dots + q_m = 1$. Распределения вероятностей p и q полностью определяют характер игры и смешанные стратегии $S_A(p)$ и $S_B(q)$ игроков A и B , где

$$S_A(p) = \begin{bmatrix} A_1 & A_2 & \dots & A_i & \dots & A_n \\ p_1 & p_2 & \dots & p_i & \dots & p_n \end{bmatrix}; \quad (4.213)$$

$$S_B(q) = \begin{bmatrix} B_1 & B_2 & \dots & B_j & \dots & B_m \\ q_1 & q_2 & \dots & q_j & \dots & q_m \end{bmatrix} \quad (4.213a)$$

есть ряды распределений.

Как уже упоминалось, в смешанных стратегиях величины выигрыша (или проигрыша) становятся случайными. Поэтому для их описания необходимо применять характеристики случайных величин вида математического ожидания (среднего выигрыша), медианы, средне-квадратического отклонения выигрыша от среднего или медианы и т. п. Чаще всего для описания выбирают математическое ожидание. Тогда средний выигрыш игрока A при применении соответствующих стратегий $S_A(p)$ и $S_B(q)$ будет равен

$$\mathbf{M}(p, q) = \sum_{i=1}^n \sum_{j=1}^m p_i q_j a_{ij}. \quad (4.214)$$

Аналогично играем в чистых стратегиях. Игрок A должен выбрать свою стратегию, максимизируя по p функцию $\alpha(p) = \alpha(p_1, p_2, \dots, p_n) = \min_q \mathbf{M}(p, q)$, а игрок B – минимизируя по q функцию $\beta(q) = \beta(q_1, q_2, \dots, q_m) = \max_p \mathbf{M}(p, q)$.

В этом случае максимин α , т. е. нижняя цена игры в смешанных стратегиях, будет равна

$$\alpha = \max_p \min_q \mathbf{M}(p, q), \quad (4.215)$$

а минимакс β – верхняя цена игры –

$$\beta = \min_q \max_p \mathbf{M}(p, q). \quad (4.216)$$

Согласно теореме фон Неймана, в рассматриваемой матричной игре существует такая пара (p^*, q^*) вероятностей p_i , $i = \overline{1, n}$, и q_j , $j = \overline{1, m}$, и оптимальных смешанных стратегий $S_A(p^*)$ игрока A и $S_B(q^*)$ игрока B , для которых, во-первых (ср. с седловой точкой (4.209)),

$$\mathbf{M}(p, q^*) \leq \mathbf{M}(p^*, q^*) \leq \mathbf{M}(p^*, q); \quad (4.217)$$

во-вторых, значения нижней (максимина α) и верхней (минимакса β) цен игры совпадут и будут равны *цене игры* γ , т. е. $\alpha = \beta = \gamma = \mathbf{M}(p^*, q^*)$.

Пару вероятностей (p^*, q^*) , обеспечивающих достижение цены игры γ , называют *оптимальной*, а стратегии $S_A(p^*)$ и $S_B(q^*)$ – оптимальными стратегиями игры игроков A и B соответственно в рассматриваемых рандомизированных смешанных стратегиях.

§ 4.6. АППАРАТ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ (ОЧЕРЕДЕЙ)

Завершим рассмотрение различных видов математического аппарата, использующего куомодные модели, элементарными основами *теории массового обслуживания* (ТМО), иногда называемой *теорией очереди*.

ТМО – основанный на теории вероятностей прикладной раздел математики, связанный с построением математических моделей потоков требований (заявок) на обслуживание, поступающих в системах массового обслуживания (СМО), самих СМО, условий их работы и связи их с показателями эффективности СМО. Что касается показателей эффективности СМО, то их разработка и исследование являются одной из задач ТМО. Понятно, что показатели должны учитывать поставленную задачу исследований, назначение СМО, суть происходящих явлений в СМО, с необходимой точностью отражать количественную связь между числом единиц приборов СМО, осуществляющих обслуживание, их тип, характеристики потока требований, очередь и качество обслуживания и т. п. Под качеством обслуживания при этом понимается количественная характеристика того, насколько хорошо и своевременно организовано и проведено (а не насколько добротнo осуществляется) обслуживание поступающих в систему требований. Добротность обслуживания характеризуют другие критерии. Набор таких характеристик и критериев качества может быть разным для разных СМО. Примерами показателей (характеристик) СМО являются длина очереди в различные моменты времени; общая продолжительность нахождения требований (в очереди и на самом обслуживании); доля времени, в течение которого обслуживающие приборы были не заняты. Чаще всего в ТМО рассматриваются задачи, связанные с поиском оптимальных

потоков требований для конкретных СМО. Для этого применяются различные методы оптимизации, вариационного исчисления, теории игр. Общий для всех СМО – случайный характер исследуемых явлений в СМО. Например, количество требований на обслуживание, временные интервалы между их поступлениями, длительность их обслуживания конкретной единицей (прибором) и т. п. Именно поэтому ТМО базируется на вероятностно-статистическом аппарате.

Обобщенная структурная схема СМО представлена на рис. 4.23. Изображенная СМО имеет n потоков (**многопоточная**) требований (заявок на обслуживание), единую очередь, s каналов (совокупности приборов, единиц) обслуживания, т. е. является **многоканальной**, каждый l -й из этих каналов имеет q_l , $l = \overline{1, s}$, фаз обслуживания (**многофазная** СМО), с **обратной связью**. Примеры **входящих потоков требований** (запросов на удовлетворение потребности в обслуживании): потоки данных на обслуживание в ЦВМ, в сети; клиентов в банке, в парикмахерской; пациентов в больнице; самолетов при заходе на посадку на каком-то аэродроме; телефонных вызовов; жителей в ремонтных мастерских, билетных кассах, магазинах.

В общем случае приборы и число фаз в каждом канале могут быть разными (**разнотипная СМО**), например, по производительности обслуживания. Однако чаще всего приборы рассматриваются однотипными (**однотипная СМО**). Рассмотрим некоторые частные разновидности СМО. Самая простая СМО содержит всего один поток требований ($n=1$), одну обслуживающую единицу (прибор, канал, т. е. $s=1$) и одну фазу обслуживания ($q=1$), без обратной связи, когда все требования ожидают обслуживания в единой очереди. Пример такой **одноканальной** системы: парикмахерская с одним мастером, одна бензозаправочная колонка. Иногда, наоборот, каналов так много, что СМО математически удобнее рассматривать как систему с **бесконечным числом** каналов ($s \rightarrow \infty$).

По времени пребывания требований в системе на начало обслуживания СМО делятся на системы с отказами, с неограниченным временем ожидания и смешанного типа.

В системах с **отказами** вновь поступившее требование покидает систему, если все ее приборы заняты. Пример таких систем – автоматические телефонные станции. Не следует путать такие системы с **неполнодоступными**, в которых не обеспечивается возможность каждого требования поступить на обслуживание (а не быть необслуженным!) любого прибора.

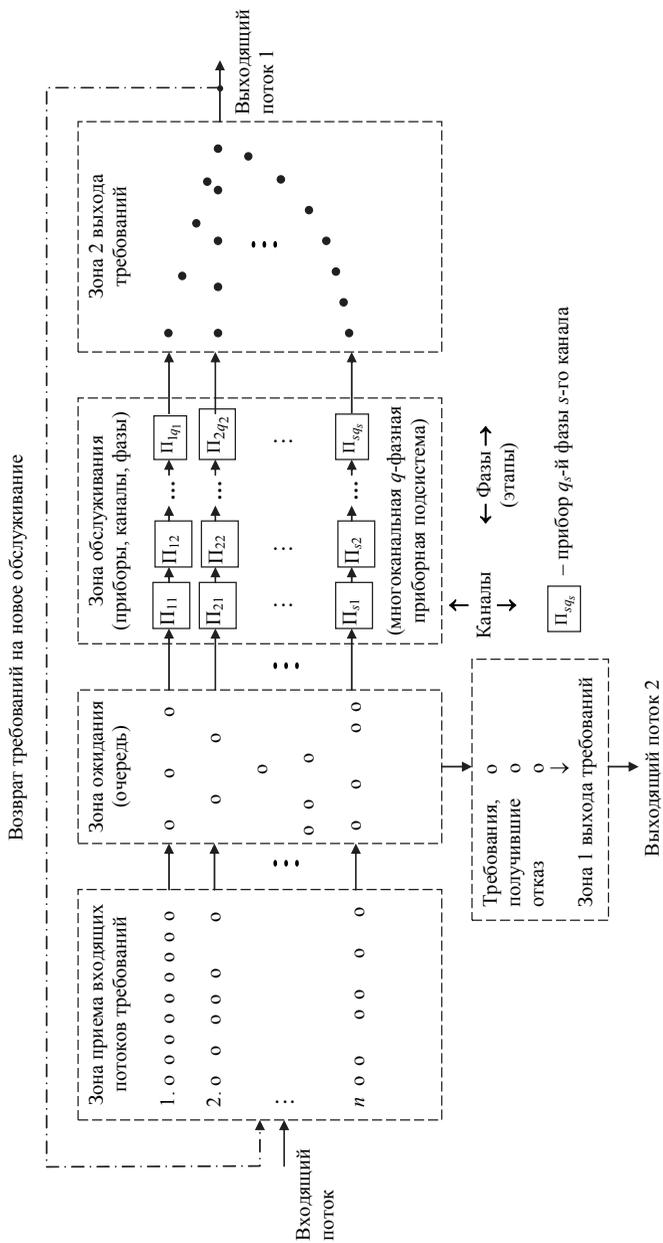


Рис. 4.23. Обобщенная структурная схема системы массового обслуживания (СМО)

В отличие от систем с отказами в системах с **неограниченным временем ожидания** все поступившие требования ожидают свою очередь до тех пор, пока какой-либо прибор не освободится. Это наиболее многочисленный и изученный класс СМО. В системах **смешанного типа** либо каждое требование находится в очереди свое ограниченное время, после которого, если оно не обслужено, покидает систему, либо ограничены длина очереди или время пребывания требования в очереди (см. рис. 4.23, зона 1 выхода).

Внимание! Попробуйте привести примеры таких СМО самостоятельно.

Существует дальнейшая детализация рассмотренных систем.

По порядку занятия свободных приборов СМО делятся на системы со строгим порядком обслуживания (системы с **упорядоченным обслуживанием**, когда, например, СМО состоит из разнотипных приборов); с обслуживанием по типу: вновь поступившее требование обслуживается в порядке освобождения приборов; со случайным занятием приборов. Могут быть и другие дисциплины обслуживания: строго в порядке очередности, по приоритетам на обслуживание, со случайным приемом требований на обслуживание и т. д.

Наконец, **выходящий поток требований** – это поток, покидающий систему. Требования выходящего потока могут быть полностью обслужены приборами системы (выход 2 на рис. 4.23), не обслужены (выход 1) либо обслужены частично (обратная связь на рис. 4.23) и поступают опять на вход системы для обслуживания на других приборах (в частности **замкнутые системы с ограниченным потоком требований**).

Кроме рассмотренных исследуются СМО с накопителем требований, с межфазными очередями и без них, а также стационарными и нестационарными условиями функционирования, с абсолютно надежными и ненадежными приборами, в частности использующими резервирование, с переменной структурой и др.

Для получения представления об аппарате ТМО рассмотрим несколько простейших СМО. Но вначале опишем наиболее часто рассматриваемый **первый важный элемент СМО** – входящий поток требований (заявок).

Прежде всего заметим, что входящий в СМО поток однородных или неоднородных требований является двойко случайным: по числу требований, поступающих за промежуток времени τ , т. е. от t до $t + \tau$; по промежуткам времени (интервалам) между моментами поступления соседних требований. При этом законы распределения временных

интервалов между появлениями требований в потоке могут быть различными. Однако чаще всего в ТМО рассматривается пуассоновский поток, в котором вероятность $P_k(\tau)$ поступления в промежуток времени τ равно k требований, т. е. вероятность того, что случайное число K поступивших за отрезок τ требований, будет равно k , подчиняется дискретному закону распределения Пуассона [20, 28, 30, 31]:

$$\mathbf{P}\{(K = k) | \tau\} = P_k(\tau) = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}, \quad k = 0, 1, 2, \dots, \quad (4.218)$$

где $\lambda > 0$ – параметр потока, характеризующий его интенсивность. Это объясняется, по крайней мере, тремя причинами [30]. Во-первых, для других видов потоков показатели качества СМО получить гораздо сложнее. Во-вторых, для приборов (единиц, средств обслуживания) такой поток является «тяжелым» в обслуживании. В-третьих, такой поток в ТМО играет роль, аналогичную той, которую в теории вероятностей и математической статистике играет нормальный (Гаусса) [20] закон распределения. В частности, аналогично предельному закону теории вероятностей (см. прил. 3), показано, что при сложении нескольких случайных потоков требований образуемый поток по своим свойствам приближается к пуассоновскому, называемому простейшим.

Простейший поток обладает тремя основными свойствами: стационарностью, ординарностью и отсутствием последствия.

Случайный поток называется *стационарным*, если вероятность $P_k(\tau)$ поступления определенного количества требований (k) в течение определенного времени зависит только от величины τ и не зависит от начала отсчета t_0 этого τ на оси времени t , т. е. если

$$P_k(t_0, t_0 + \tau) = P_k(\tau). \quad (4.219)$$

Это означает, что для простейшего (стационарного) потока достаточно задать только его параметр (интенсивность потока) λ .

Отсутствие последствия в потоке означает, что вероятность $P_K(\tau)$ поступления k требований за отрезок $(t_0, t_0 + \tau)$ не зависит от предыстории потока, т. е. от того, сколько требований уже поступило в систему до этого t_0 . Иными словами, отсутствие последствия отражает взаимную независимость протекания процессов в потоке требований в неперекрывающиеся соседние промежутки времени.

Ординарность потока – это практическая невозможность появления двух или более требований в один и тот же момент времени, т. е. существование и равенство нулю предела

$$\lim_{\Delta t \rightarrow 0} \frac{P_{>1}(\Delta t)}{\Delta t} = 0, \quad (4.220)$$

что означает вероятность появления более одного требования в малом интервале Δt как значения функции от Δt , является бесконечно малой величиной (при $\Delta t \rightarrow 0$) большего порядка, чем Δt . Итак, простейший поток – это пуассоновский стационарный ординарный поток без последствия.

Интенсивность простейшего потока – математическое ожидание случайного числа K требований в единицу времени ($M\{K\}$ при $\tau = 1$) равно λ , поскольку для распределения Пуассона (4.218) $M\{K\} = \lambda\tau$. Отсюда видно, что интенсивность простейшего потока $\lambda(t)$ как функция t является константой. Это следствие стационарности потока. В противном случае поток является нестационарным.

Как следует из распределения Пуассона (4.218) при $k=1$, закон распределения временных промежутков в потоке является показательным (экспоненциальным [20]). Верно и обратное утверждение: если закон распределения временных интервалов между появлениями требований является показательным, то такой поток простейший.

Вторым элементом после входного потока требований считается подсистема обслуживания. Именно от входного потока и подсистемы обслуживания зависят прежде всего основные показатели работы СМО. Для описания элементов (приборов, каналов, фаз (этапов)) обслуживания также в основном используется вероятностный аппарат. При этом допустимо, чтобы законы распределения времени обслуживания были разными из семейства правосторонних, т. е. используемых для описания неотрицательных случайных величин. Наиболее часто в теоретических задачах и практических приложениях закон распределения времени обслуживания полагается показательным (сокращенно пишется и говорится: время обслуживания показательное). Это означает, что функция распределения времени на обслуживание требования отдельным прибором (как случайная величина T) имеет функцию распределения (ср. с (4.218) при $k=1$):

$$F_T(t) = P\{T \leq t\} = 1 - e^{-\lambda t} \quad (4.221)$$

или плотность распределения

$$W_T(t) = \begin{cases} 0, & t < 0; \\ \mu e^{-\mu t}, & t > 0, \mu \geq 0, \end{cases} \quad (4.221a)$$

где μ – интенсивность потока обслуживания, равная $\mu = 1/\bar{T}$, $\bar{T} = \bar{t}_{\text{обс}}$ – среднее время (математическое ожидание $\mathbf{M}\{T\}$ времени) обслуживания. **Поток обслуживания** – это поток требований, обслуживаемых соответствующей непрерывно занятой единицей зоны обслуживания (прибором, каналом, фазой (этапом), всей СМО).

Отсюда, если рассмотреть одноканальную ($s=1$ на рис. 4.23) q -фазную (q -этапную) СМО с одинаковой интенсивностью показательного потока обслуживания каждым прибором, равной μ , то время обслуживания всеми приборами будет равно $T = T_1 + \dots + T_q$, а распределение случайной величины T будет подчиняться закону Эрланга с плотностью [20]:

$$W_T(t) = \frac{\mu^q t^{q-1}}{(q-1)!} e^{-\mu t}, \quad (4.222)$$

где «!» – знак факториала (ср. с (4.218)).

Допустим, что СМО является однофазной s -канальной, в которой все s приборов начинают обслуживать каждое требование одновременно и прекращают обслуживание сразу же, как только один из приборов первым выполнил заявку (обслужил требование)¹. Предположим, что приборы разнотипные, а время обслуживания требования i -м прибором $i = \overline{1, s}$ равно μ_i . Тогда время обслуживания требования всеми приборами СМО будет также показательным с распределением (4.221), в котором $\mu = \mu_1 + \mu_2 + \dots + \mu_s$. Этот факт отражает важное для характеристик эффективности СМО свойство показательного времени обслуживания. Действительно, если все приборы имеют одинаковую производительность, т. е. $\mu_1 = \mu_2 = \dots = \mu_s$, то $\mu = s\mu_1$. Поскольку

¹ Несколько официантов обслуживают посетителя ресторана, как в Китае. Обстрел или бомбардировка цели несколькими орудиями, стрелками, бомбардировщиками одновременно.

$\mu = 1/\bar{T} = 1/\bar{t}_{\text{обс}}$, это означает, что среднее время обслуживания такой s -канальной СМО уменьшается в s раз по сравнению со временем обслуживания требований только одним прибором. При этом среднеквадратическое отклонение времени обслуживания T также уменьшается в s раз.

Третий важный элемент СМО – это очередь. От того, какая она и как организована, зависят многие показатели работы (эффективности) СМО. Как уже упоминалось, очередь может быть ограничена и не ограничена. Ограничения на очередь касаются ее длины, времени ожидания, дисциплины обслуживания и т. д.

Теперь кратко рассмотрим некоторые показатели работы СМО.

• Прежде всего связь среднего времени пребывания требования в системе со средним числом требований в системе и интенсивностью входного потока требований. Для любой СМО введем следующие обозначения: $X(t)$ – число заявок, прибывших в СМО до момента t ; $Y(t)$ – число заявок, покинувших СМО до момента t ; $Z(t) = X(t) - Y(t)$ – число заявок, находящихся в СМО. Понятно, что в рамках принятого описания $X(t)$, $Y(t)$ и $Z(t)$ – случайные функции, траектории которых являются ступенчатыми, длины ступенек определяются законом распределения временных промежутков между соседними требованиями в потоке, а высоты – количеством требований k , возникающих одновременно. Ясно, что при простейшем потоке $k = 1$, а закон распределения экспоненциальный (показательный).

Рассмотрим большой промежуток времени L . Тогда среднее число $\mathbf{M}\{Z\}$ стационарного процесса $Z(t)$ можно оценить через траекторное эмпирическое *среднее число требований, находящихся в системе*, \bar{z} , которое будет равно

$$\bar{z} = \frac{1}{L} \int_0^L z(t) dt, \quad (4.223)$$

где $z(t)$ – конкретная траекторная (реализация) $Z(t)$.

Поскольку λ – интенсивность потока, т. е. среднее число требований в потоке, то λL есть *среднее число требований, поступивших за время L* . Отсюда легко получается знаменитая формула Литтла [31], связывающая среднее время $\bar{\theta}$ пребывания требования в системе $\bar{\theta}$:

$$\bar{\theta}_{\text{сист}} = \bar{z}_{\text{сист}} / \lambda. \quad (4.224)$$

Формула Литтла (4.224) означает, что для любой СМО, т. е. при любом типе потока требований, распределении времени, дисциплине обслуживания, канальности СМО *среднее время пребывания требования в системе равно среднему числу заявок в системе, деленному на интенсивность потока требований.*

Оказывается, подобная формула (вторая формула Литтла) имеет место и для очередей: среднее время пребывания требования в очереди [31]

$$\bar{\theta}_{\text{оч}} = \bar{z}_{\text{оч}} / \lambda . \quad (4.225)$$

Из других показателей рассмотрим следующие [30, 31].

- Вероятность потери требования. Применительно к СМО с потерями она равна вероятности занятости обслуживанием требований всех n приборов системы и обозначается p_n или $p_{\text{отг}}$.

- Вероятность p_k того, что обслуживанием требований в системе занято k , $k = \overline{0, n}$, приборов, частным случаем которой являются p_n и p_0 – все приборы свободны.

- Среднее число занятых приборов (каналов, фаз)

$$N_3 = \sum_{k=1}^n k p_k , \quad (4.226)$$

характеризующее степень загрузки СМО.

- Среднее число свободных (от обслуживания) приборов

$$N_0 = \sum_{k=0}^{n-1} k p_k = N - N_3 . \quad (4.227)$$

- Коэффициент простоя приборов (каналов, ...)

$$K_{\text{п}} = N_0 / n \quad (4.228)$$

и коэффициент загрузки (занятости)

$$K_3 = N_3 / n = 1 - K_{\text{п}} . \quad (4.229)$$

Для систем с ожиданием это такие характеристики как:

- вероятность $P_k(t_{\text{ож}} > \tau)$ того, что время ожидания $t_{\text{ож}} > \tau$ при условии, что в текущий момент t поступления требования в систему в ней уже обслуживалось k требований;
- вероятность того, что время ожидания поступившим требованием обслуживания $t_{\text{ож}} > t$

$$P(t_{\text{ож}} > t) = \sum_{k=0}^n p_k P_k(t_{\text{ож}} > t); \quad (4.230)$$

- среднее время ожидания требований в очереди

$$T_{\text{ож}} = - \int_0^{\infty} t dP(t_{\text{ож}} > t); \quad (4.231)$$

- средняя длина очереди

$$M_{\text{ож}} = \sum_{k=n}^{\infty} (k-n) p_k \quad \text{при } k \geq n \quad (4.232)$$

и среднее число требований, находящихся в системе,

$$\bar{z} = M_{\text{ож}} + N_3. \quad (4.233)$$

Понятно, что для каждой разновидности СМО (см. рис. 4.23) конкретное значение перечисленных показателей будет по-своему связано с плотностью потока заявок λ , параметром обслуживания $\mu = 1/\bar{t}_{\text{обс}}$, где $\bar{t}_{\text{обс}}$ – среднее время обслуживания требований в системе, числом n приборов в ней, ее конфигурацией и т. д.

Так, в СМО с потерями при условии поступления простейшего потока заявок и показательном времени обслуживания имеем [30]

$$p_k = p_0 \alpha^k / (k!), \quad \alpha = \lambda / \mu; \quad (4.234)$$

$$p_0 = \left[\sum_{m=0}^{\infty} \alpha^m / m! \right]^{-1}; \quad (4.235)$$

$$p_n = p_0 \alpha^n / n!; \quad (4.236)$$

$$N_3 = p_0 \sum_{k=1}^n \alpha^k / (k-1)! \quad (4.237)$$

В системах с накоплением заявок [30]

$$p_k = \alpha^k / (\alpha + 1)^{k+1} \text{ при } k < n; \quad (4.238)$$

$$p_n = \alpha^n / (\alpha + 1)^n \quad (4.239)$$

и т. д.

Помимо перечисленных функциональных показателей, отражающих способность системы обслуживать требования, могут быть введены также показатели, отражающие качество самого обслуживания, стоимостные и иные показатели.

Так, например, среди стоимостных часто рассматриваются следующие показатели [30]:

- $q_{об}$ – стоимость обслуживания каждого требования в системе;
- $q_{ож}$ – стоимость потерь, связанных с простаиванием требований в очереди в единицу времени;
- q_y – стоимость убытков, связанных с уходом требований из системы;
- q_k – стоимость эксплуатации каждого прибора СМО в единицу времени;
- $q_{пк}$ – стоимость единицы времени простоя прибора и т. д.

ЗАКЛЮЧЕНИЕ

В четвертой главе рассмотрена первая часть формального аппарата, используемого при моделировании объектов.

Прежде всего рассмотрен детерминированный аппарат. Большое внимание уделено очень часто применяемому стохастическому аппарату, а именно моделям и методам, рассматриваемым в теории вероятностей, математической статистике и статистической имитации.

Из другого формального аппарата рассмотрены модели, которые могут быть использованы в ситуациях, когда определенные детерминированные и стохастические модели не адекватны моделируемым объектам, условиям их функционирования, имеющимся данным, целям и задачам моделирования. Среди них модели и методы, использующие стохастический аппарат (в частности байесовский), нечеткий, интервальный, динамического хаоса, фрактальный, экспертный, а также аппарат поиска оптимальных решений, теории игр и теории массового обслуживания.

Особое внимание уделено следующим аспектам применения аналитического модельного представления объектов:

- выявлению условий, при которых тот или иной аппарат применим;
- возможности и допустимости описания одних и тех же объектов разными моделями с учетом цели и условий моделирования;
- желательности использования разных моделей различных классов или подклассов при решении прикладных задач и, как следствие, автоматизации процедуры выбора моделей под решаемую прикладную задачу в конкретных условиях.

Завершая четвертую главу, обращаем еще раз внимание на замечания, изложенные в § 2.7. Особенно это касается оптимальных моделей.

ВОПРОСЫ ДЛЯ САМОПОДГОТОВКИ

1. Что понимается под детерминированными, законовыми моделями? Почему они так называются? При каких условиях могут использоваться на практике?

2. Приведите примеры детерминированного аппарата.

3. Приведите примеры разновидностей функциональных моделей и укажите условия, при которых они рассматриваются и применяются.

4. Что такое стохастические модели? Каковы предпосылки их практической применимости?

5. Перечислите разновидности стохастических моделей. В чем их сходство и различие?

6. Какова структура вероятностного и статистического математических аппаратов? Что такое случайный элемент?

7. Приведите постановки вероятностных, статистических отражательных и статистических имитационных задач.

8. Можете ли вы построить таблицы соответствия между понятиями теории физического эксперимента и разных аппаратов описания его результатов: точечных и интервальных, детерминированного, вероятностного, статистического, нечеткого, экспертного?

9. Перечислите основные вероятностные и статистические характеристики случайных величин, векторов, функций. Укажите их свойства и связь с характеристиками сигналов.

10. Что представляют собой следующие виды вероятностно-статистического анализа: дескриптивный, распределений значений, регрессионный, корреляционный, спектральный и т. п.?

11. В чем идея и особенности конкорреляционных и дисперсионных характеристик и методов анализа?

12. Приведите примеры характеристик случайных элементов, инвариантных к разным условиям: отдельным видам преобразования случайных элементов, типам измерительной шкалы и т. п.

13. В чем сходство и отличие между вероятностными и статистическими характеристиками, анализами?

14. Что такое генеральная совокупность и выборка из нее? Статистика? Всякая ли статистика может быть оценкой какой-либо характеристики, параметра?

15. Что такое оценка параметра или характеристики случайного элемента?

16. Перечислите требования к генеральной совокупности и оценке характеристики.

17. Укажите и охарактеризуйте методы непараметрического оценивания.

18. Укажите и охарактеризуйте методы параметрического оценивания и оценивания параметров характеристик случайных элементов.

19. Что такое смешанные методы оценивания характеристик случайных элементов? Приведите примеры смешанных методов.

20. Приведите постановки задач статистической проверки гипотез и проверки статистических гипотез.

21. Приведите основные понятия и методы статистической проверки гипотез.

22. Приведите обобщенную схему статистической проверки гипотез. От чего зависят критерии проверки гипотез?

23. В чем особенность последовательных методов проверки гипотез?

24. Что понимается под статистической имитацией объектов и случайных элементов?

25. Что такое метод статистических испытаний? Что такое метод Монте-Карло?

26. Приведите примеры методов имитации случайных элементов – последовательностей их выборочных значений.

27. В чем отличие между случайными, псевдослучайными и квазислучайными последовательностями чисел? С каким типом математического аппарата (детерминированным, вероятностным, статистическим) они связаны?

28. Что представляет собой байесовский стохастический аппарат? Когда он применяется?

29. Что представляет собой аппарат нечетких множеств? Приведите его основные понятия.

30. Что такое интервальный анализ и когда он применяется?

31. Что такое теория катастроф? Приведите и поясните примеры флагов (признаков) катастроф.

32. Что понимается под динамическим ха́осом? Приведите основные понятия и модели его теории. Каковы условия его применимости?

33. Что такое фрактал? Приведите примеры фрактальных моделей. Что такое фрактальная размерность?

34. Что такое экспертное описание и исследование объектов, каковы условия их применимости? Приведите их часто используемые методы и технологии.

35. Что такое оптимизационные модели? Каковы их разновидности и основные понятия?

36. Приведите примеры постановки оптимизационных задач линейного и нелинейного программирования.

37. Что такое многокритериальная оптимизация? Приведите примеры методов решения ее задач.

38. Что означает корректность постановки задачи в общем виде и оптимизационных задач в частности?

39. В чем суть и различие вариационных постановок задач и моделей объектов?

40. Что понимается под динамическим программированием? Каковы особенности постановки и решения задач динамического программирования?

41. В чем суть и различие состязательных моделей, теории игр, принятия решений в условиях конфликтов? Укажите разновидности теории игр.

42. Приведите основные понятия и методы теории игр.

43. Что такое теория и системы массового обслуживания (СМО)? Каковы разновидности СМО?

44. Приведите основные понятия теории и характеристики СМО.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Губарев В.В.* Информатика: прошлое, настоящее, будущее: учеб. пособие / В.В. Губарев. – М.: Техносфера, 2011. – 432 с.
2. *Перегудов Ф.П.* Основы системного анализа: учебник / Ф.И. Перегудов, Ф.П. Тарасенко. – Томск: Изд-во НТЛ, 1997. – 396 с.
3. *Губарев В.В.* Концептуальные основы информатики: учеб. пособие / В.В. Губарев. – Новосибирск: Изд-во НГТУ, 2001. – Ч. 1. Сущностные основы информатики. – 149 с.
4. *Губарев В.В.* Информатика в рисунках и таблицах: учеб. пособие / В.В. Губарев. – Новосибирск: Изд-во НГТУ, 2003. – 198 с.
5. *Губарев В.В.* Алгоритмы спектрального анализа случайных сигналов: монография / В.В. Губарев. – Новосибирск: Изд-во НГТУ, 2005. – 660 с.
6. *Советов Б.Я.* Моделирование систем / Б.Я. Советов, С.А. Яковлев. – М.: Высшая школа, 2007. – 346 с.
7. *Альсова О.К.* Прогнозирование притока реки Обь в створе Новосибирской ГЭС на основе вариативного моделирования: дис. ... канд. техн. наук / О.К. Альсова; науч. руков. В.В. Губарев; Новосиб. гос. техн. ун-т, Новосибирск, 2002. – 162 с.
8. *Губарев В.В.* Системный анализ в экспериментальных исследованиях / В.В. Губарев. – Новосибирск: Изд-во НГТУ, 2000. – Ч. 1. – 99 с.
9. *Флейшман Б.С.* Основы системологии / Б.С. Флейшман. – М.: Радио и связь, 1982. – 368 с.
10. *Теория систем и системный анализ в управлении организациями: Справочник* / Под ред. В.Н. Волковой и А.А. Емельянова. – М.: Финансы и статистика, 2006. – 848 с.
11. *Костюк В.Н.* Изменяющиеся системы / В.Н. Костюк. – М.: Наука, 1993. – 352 с.
12. *Лившиц В.Р.* Математические методы обработки результатов наблюдений: учеб. пособие / В.Р. Лившиц. – Новосибирск: НГУ, 2007. – Ч. 1, 2. Анализ данных. – 292 с.
13. *Анфилатов В.С.* Системный анализ в управлении: учеб. пособие / В.С. Анфилатов, А.А. Емельянов, А.А. Кукушкин. – М.: Финансы и статистика, 2002. – 368 с.
14. *Тарасенко Ф.П.* Прикладной системный анализ (наука и искусство решения проблем): учебник / Ф.П. Тарасенко. – Томск: Изд-во ТГУ, 2004. – 186 с.

15. *Смолин Д.В.* Введение в искусственный интеллект: конспект лекций / Д.В. Смолин. – М.: Физматлит, 2007. – 264 с.
16. *Загоруйко Н.Г.* Когнитивный анализ данных / Н.Г. Загоруйко. – Новосибирск: Акад. изд-во «Гео», 2013. – 186 с.
17. *Руководство по выражению неопределенности измерения* / под ред. В.А. Слава. – СПб.: ВНИИМ, 1999.
18. *Сергиенко А.Б.* Цифровая обработка сигналов / А.Б. Сергиенко. – СПб.: Питер, 2003. – 604 с.
19. *Яковлев А.Н.* Основы теории сигналов в примерах, упражнениях и задачах: учеб. пособие / А.Н. Яковлев. – Новосибирск: Изд-во НГТУ, 2012. – 472 с.
20. *Губарев В.В.* Вероятностные модели: справочник / В.В. Губарев. – Новосибирск: НЭТИ, 1992. – 421 с.
21. *Губарев В.В.* Практикум по теории вероятностей: учеб. пособие / В.В. Губарев, В.М. Зыбарев. – Новосибирск: Изд-во НГТУ, 1995. – 157 с.
22. *Ехлаков Ю.П.* Теоретические основы компьютерных систем обработки информации и управления / Ю.П. Ехлаков, В.В. Яворский. – Караганда: Изд-во КарГТУ, 2005. – 394 с.
23. *Губарев В.В.* Алгоритмы статистических измерений / В.В. Губарев. – М.: Энергоатомиздат, 1985. – 272 с.
24. *Лемешко Б.Ю.* Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. – Новосибирск: Изд-во НГТУ, 2011. – 888 с.
25. *Фрумкин В.Д.* Теория вероятностей и статистика в метрологии и измерительной технике / В.Д. Фрумкин, Н.А. Рубичев. – М.: Машиностроение, 1987. – 168 с.
26. *Кельтон В.* Имитационное моделирование. Классика CS / В.Д. Кельтон, А.М. Лоу. – СПб.: Питер; Киев: Изд. группа ВН, 2004. – 847 с.
27. *Соболь И.М.* Точки, равномерно заполняющие многомерный куб / И.М. Соболь. – М.: Знание, 1985. – 32 с.
28. *Коршунов Ю.М.* Математические основы кибернетики / Ю.М. Коршунов. – М.: Энергоатомиздат, 1987. – 496 с.
29. *Гилмор Р.* Прикладная теория катастроф: В 2 кн. / Р. Гилмор. – М.: Мир, 1984. – Кн. 1 – 350 с., кн. 2 – 285 с.
30. *Новиков О.А.* Прикладные вопросы теории массового обслуживания / О.А. Новиков, С.И. Петухов. – М.: Сов. радио, 1969. – 400 с.
- Вентцель Е.С.* Исследование операций. Задачи, принципы, методология / Е.С. Вентцель. – М.: Высшая школа, 2001. – 208 с.

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ 1

ОПЕРАТОРЫ УСРЕДНЕНИЯ

П1.1. **Оператор усреднения по вероятностной мере** (оператор математического ожидания) $\mathbf{M}\{f(\aleph)\}$ случайного элемента \aleph (!)

$$\mathbf{M}\{f(X_1, \dots, X_n)\} = \begin{cases} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dF_{\mathbf{X}}(x_1, \dots, x_n) \\ \text{для всех } \mathbf{X} = (X_1, \dots, X_n); \\ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) W_{\mathbf{X}}(x_1, \dots, x_n) dx_1, \dots, dx_n \\ \text{для абсолютно непрерывных } \mathbf{X}; \\ \sum_{i_1=-\infty}^{\infty} \dots \sum_{i_n=-\infty}^{\infty} f(X_{1,i_1}, \dots, X_{n,i_n}) \mathbf{P}\left\{\bigcap_{s=1}^n (x_s = x_{s,i_s})\right\} \\ \text{для дискретных } \mathbf{X}, \end{cases}$$

когда хотя бы одна из двух (положительная или отрицательная) частей n -мерных интегралов конечна, где $F(\cdot)$ – функция, $W(\cdot)$ – плотность распределения вероятностей, $\mathbf{P}(\cdot)$ – вероятность. Положительной части интеграла соответствует область значений $\mathbf{x} = (x_1, x_2, \dots, x_n)$, при которой $f(\mathbf{x}) \geq 0$, отрицательной – $f(\mathbf{x}) < 0$.

П1.2. **Оператор ансамблевого усреднения** $\mu_N\{f(\chi)\} = \frac{1}{N} \sum_{i=1}^N f(\varkappa_i)$ по выборке $\chi = (\varkappa_i, i = \overline{1, N})$ случайного элемента \aleph .

Для случайных процессов $X(t)$, $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_n(t))$ элемент выборки \varkappa_i – это $\varkappa_i = x_i(t)$, либо $\varkappa_i = (x_i(t_1), x_i(t_2), \dots, x_i(t_n))$ для $X(t)$ или $\varkappa_i = (x_{1,i}(t_1), x_{2,i}(t_2), \dots, x_{n,i}(t_n))$ для $\mathbf{X}(t)$.

П1.3. Оператор траекторного усреднения

$\mu_{\Gamma} \{f[\mathbf{x}(t)]\}$ – непрерывного ($t = t$ – непрерывно, $\Gamma = T$) или дискретного (t – дискретно; $t = i$ или, как правило, $t = i\Delta t$, $\Delta t > 0$; $\Gamma = N$) усреднения по траектории $f[\mathbf{x}(t)]$:

$$\mu_{\Gamma} \{f[\mathbf{x}(t)]\} = \begin{cases} \mu_T \{f[\mathbf{x}(t)]\} = \frac{1}{T} \int_0^T f[\mathbf{x}(t)] dt, & t - \text{непрерывно;} \\ \mu_N \{f[\mathbf{x}(i\Delta t)]\} = \frac{1}{N} \sum_{i=0}^{N-1} f[\mathbf{x}(i\Delta t)], & t - \text{дискретно.} \end{cases}$$

П1.4. Оператор усреднения по условной вероятностной мере

$W_{X_i}(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ случайного вектора $\mathbf{X} = (X_1, \dots, X_n)$:

$$\mathbf{M}_{X_i} \left\{ f \left(X_i \mid \bigcap_{\substack{s=1, \\ s \neq i}}^n (X_s = x_s) \right) \right\} = \begin{cases} \int_{-\infty}^{\infty} f(x_i) W_{X_i}(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) dx_i, & \text{если } \mathbf{X} - \text{абсолютно непрерывный вектор;} \\ \sum_{k=-\infty}^{\infty} f(x_{i,k}) \mathbf{P} \left\{ (X_i = x_{i,k}) \mid \bigcap_{\substack{s=1, \\ s \neq i}}^n (X_s = x_s) \right\}, & \text{если } \mathbf{X} - \text{дискретный вектор.} \end{cases}$$

СВОЙСТВА ВЕРОЯТНОСТНЫХ ХАРАКТЕРИСТИК

Т а б л и ц а П 2.1

Свойства функций регрессии случайных величин X и Y

№ п/п	Условие, наименование свойства	$m_Y(x)$	$m_X(y)$	Примечание
1	Существование	$m_Y(x)$ и $m_X(y)$ существуют, если существуют моменты случайных величин X и Y первых двух порядков	$m_X(y)$	Если разрешима проблема моментов
2	X и Y независимые	$m_Y(x) = m_Y$	$m_X(y) = m_X$	Обратные утверждения в общем случае неверны
3	Свойство асимметрии	$m_Y(x) \neq m_X(y)$		В общем случае
4.1	$Y = f(x)$, $f(\cdot)$ – однозначная функция	$m_Y(x) = f(x)$	$m_X(y)$ зависит от вида обратной функции $x = \varphi(y)$ и распределения Y	Если зависимость у от x взаимно однозначная, то $m_Y(x) = f(x)$, $m_X(y) = \varphi(y)$
4.2	$X = \varphi(Y)$	$m_Y(x)$ зависит от вида функции $y = f(x)$ и распределения X	$m_X(y) = \varphi(y)$	Функция регрессии аддитивна
5	$Y = aX + b$, $X = cY + d$	$m_Y(x) = ax + b$	$m_X(y) = cy + d$	
6	Пусть $Y = f(x)$ или $X = \varphi(y)$, f, φ – подбираемые функции	$\min_f \Delta_Y^2 = \min_f \mathbf{M} \left\{ [Y - f(X)]^2 \right\} = \min_f \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - f(x)]^2 W(x, y) dx dy = \Delta_{X \min}^2 = \mathbf{M} \left\{ \sigma_X^2(X) \right\}$, когда $f(x) = m_Y(x)$	$\min_{\varphi} \Delta_X^2 = \min_{\varphi} \mathbf{M} \left\{ [X - \varphi(Y)]^2 \right\} = \min_{\varphi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x - \varphi(y)]^2 W(x, y) dx dy = \Delta_{Y \min}^2 = \mathbf{M} \left\{ \sigma_Y^2(Y) \right\}$, когда $\varphi(y) = m_X(y)$	Лучшее среднеквадратическое (СК) приближение функциональной зависимости Y от X или X от Y дает функция регрессии
6.1	Следствие	$\min_a \mathbf{M} \left\{ [Y - a]^2 \right\} = \sigma_Y^2$, когда $a = m_Y$	$\min_b \mathbf{M} \left\{ [X - b]^2 \right\} = \sigma_X^2$, когда $b = m_X$	Математическое ожидание – лучшее СК числовое приближение значений случайной величины
7	Функции регрессии линейные	$m_Y(x) = m_Y + \frac{\sigma_{YX}}{\sigma_X} \rho_{XY} (x - m_X)$	$m_X(y) = m_X + \frac{\sigma_{XY}}{\sigma_Y} \rho_{XY} (y - m_Y)$	ρ_{XY} – коэффициент корреляции X и Y

Свойства скадастических функций случайных величин X и Y

№ п/п	Условие, наименования свойства	$\sigma_Y(x)$	$\sigma_X(y)$	Примечание
1	Существование	$\sigma_Y(x)$ и $\sigma_X(y)$ существуют, если распределение имеет первые и вторые моменты	$\sigma_X(y)$	Должны выполняться условия разрешения проблемы моментов
2	X и Y независимы	$\sigma_Y(x) = \sigma_Y$	$\sigma_X(y) = \sigma_X$	Обратное утверждение в общем случае неверно
3	Свойство асимметрии	$\sigma_Y(x) \neq \sigma_Y(y)$		В общем случае
4.1	$Y = f(x)$, $f(\cdot)$ – однозначная функция	$\sigma_Y(x) = 0$	$\sigma_X(y)$ зависит от вида $x = \varphi(y)$ и распределения Y	При взаимно однозначной зависимости Y и X имеем $\sigma_Y(x) = \sigma_X(y) = 0$
4.2	$X = \varphi(Y)$, $\varphi(\cdot)$ – однозначная функция	$\sigma_Y(x)$ зависит от вида функции $f(x)$ и распределения X	$\sigma_X(y) = 0$	
5	$Y = aX + b$, $X = cY + d$	$\sigma_Y(x) = 0$	$\sigma_X(y) = 0$	Следствие свойства 4
6.1	Разложение дисперсии	$\mathbf{D}\{Y\} = \sigma_Y^2 = \mathbf{D}\{m_Y(X)\} + \mathbf{M}\{D_Y(X)\} = \mathbf{M}\{[m_Y(X) - m_Y]^2\} + \mathbf{M}\{\sigma_Y^2(X)\}$	$\mathbf{D}\{X\} = \sigma_X^2 = \mathbf{D}\{m_X(Y)\} + \mathbf{M}\{D_X(Y)\} = \mathbf{M}\{[m_X(Y) - m_X]^2\} + \mathbf{M}\{\sigma_X^2(Y)\}$	$\sigma_Y^2(x)$ – условная дисперсия Y при конкретном значении $X = x$
6.2	Линия регрессии – прямая	Если $m_Y(x) = ax + b$, то $\mathbf{D}\{m_Y(X)\} = \sigma_Y^2 \rho_{XY}^2$; $\Delta_{Y \min}^2 = \Delta_{Y \max}^2 = \sigma_Y^2 (1 - \rho_{XY}^2)$	Если $m_X(y) = cy + d$, то $\mathbf{D}\{m_X(Y)\} = \sigma_X^2 \rho_{XY}^2$; $\Delta_{X \min}^2 = \Delta_{X \max}^2 = \sigma_X^2 (1 - \rho_{XY}^2)$	ρ_{XY} – коэффициент корреляции X и Y
7	Связь с другими характеристиками	$\mathbf{D}\{m_Y(X)\} = \sigma_Y^2 \eta_{Y X}^2$; $\mathbf{M}\{\sigma_Y^2(X)\} = \sigma_Y^2 [1 - \eta_{Y X}^2]$	$\mathbf{D}\{m_X(Y)\} = \sigma_X^2 \eta_{X Y}^2$; $\mathbf{M}\{\sigma_X^2(Y)\} = \sigma_X^2 [1 - \eta_{X Y}^2]$	$\eta_{Y X}^2$ – корреляционное отношение Y по X
8	Связь с Δ_{\min}^2	$\Delta_{Y \min}^2 = \sigma_Y^2 [1 - \eta_{Y X}^2] = \sigma_Y^2 c_{Y X}^2$	$\Delta_{X \min}^2 = \sigma_X^2 [1 - \eta_{X Y}^2]$	Минимальное СКО Y от $f(X)$ по X
9	Линия регрессии прямая	$\Delta_{Y \min}^2 = \Delta_{Y \max}^2 = \mathbf{M}\{\sigma_Y^2(X)\} = \sigma_Y^2 (1 - \rho_{XY}^2)$	$\Delta_{X \min}^2 = \Delta_{X \max}^2 = \mathbf{M}\{\sigma_X^2(Y)\} = \sigma_X^2 (1 - \rho_{XY}^2)$	Минимум СКО Y при линейной аппроксимации

Свойства корреляционных характеристик связи случайных величин X и Y

№ п/п	Условие, наименование свойства	Свойства			Примечание
		ρ_{XY}	$\eta_{Y X}$	$\eta_{X Y}$	
1	Существование	Существуют только, если существуют первые и вторые моменты распределения			Существуют всегда
2	X и Y независимые	$\rho_{XY} = 0$	$\eta_{Y X} = 0$	$\eta_{X Y} = 0$	$\chi_{XY} = 0$
3	Свойство симметрии (асимметрии)	$\rho_{XY} = \rho_{YX}$	$\eta_{Y X} \neq \eta_{X Y}$		$\chi_{XY} = \chi_{YX}$
4	Экстремальные значения	$ \rho_{XY} \leq 1$, $ \rho_{XY} = 1$ тогда и только тогда, когда X и Y линейно функционально связаны, т. е. $Y = aX + b$, $ a > 0$	$0 \leq \eta_{Y X} \leq 1$, $\eta_{Y X} = 1$ тогда и только тогда, когда $Y = f(X)$ и $f(\cdot)$ – однозначная функция!	$0 \leq \eta_{X Y} \leq 1$, $\eta_{X Y} = 1$ тогда и только тогда, когда $X = \varphi(Y)$ и $\varphi(\cdot)$ – однозначная функция!	$ \chi_{XY} = 1$ тогда и только тогда, когда X и Y связаны монотонной взаимно однозначной функцией
5	$Y = aX + b$, $ a > 0$	$\rho_{XY} \text{ sign}(a)$	$\eta_{Y X} = 1$	$\eta_{X Y} = 1$	$\chi_{XY} \text{ sign}(a)$
6	$Y = f(X)$; $X = \varphi(Y)$	$ \rho_{XY} = 1$, если $f(\cdot)$ и $\varphi(\cdot)$ – линейные. Иначе значение ρ_{XY} зависит от вида $f(\cdot)$, $\varphi(\cdot)$ и одномерных распределений X, Y	$\eta_{Y X} = 1$, если $f(\cdot)$ – однозначная функция $\eta_{Y X}$ может иметь любое значение в зависимости от $f(\cdot)$ и $W_Y(Y)$	$\eta_{X Y}$ может иметь любое значение в зависимости от $\varphi(\cdot)$ и $W_X(X)$	$\chi_{XY} = 1$, если $f(\cdot)$ и $\varphi(\cdot)$ – взаимно однозначные монотонные, иначе любое в зависимости от $f(\cdot)$ и $W_X(X)$, $W_Y(Y)$

Окончание табл. П2.3

№ п/п	Условие, наименование свойства	Свойства			Примечание
		ρ_{XY}	$\eta_{Y X}$	χ_{ZY}	
7	$Z = aX + b$; $V = cY + d$	$\rho_{ZY} = \rho_{XY} \text{sign}(ac)$	$\eta_{ZY V} = \eta_{XY}$	$\chi_{ZY} = \chi_{XY} \text{sign}(ac)$	
8	$Z = f(X)$, $V = g(Y)$	$\rho_{ZY} = \rho_{XY} \text{sign}(\varepsilon_f \varepsilon_g)$, если $f(\cdot)$ и $g(\cdot)$ – линейные функции; ρ_{ZY} может быть любым, если $f(\cdot)$ и $g(\cdot)$ – нелинейные функции	$\eta_{ZY V} = \eta_{XY}$, если $f(\cdot)$ – однозначная функция; $\eta_{ZY V} = \eta_{XY}$, если $f(\cdot)$ и $g(\cdot)$ – взаимно однозначные функции	$\chi_{ZY} = \varepsilon_f \varepsilon_g \chi_{XY}$, если $f(\cdot)$ и $g(\cdot)$ – любые монотонные функции; если $f(\cdot)$ и $g(\cdot)$ – взаимно однозначные функции	$\varepsilon_f, \varepsilon_g > +1$, если f, g монотонно возрастают; $\varepsilon_f, \varepsilon_g \leq -1$, если f, g монотонно убывают в области значений X, Y
9	Связь между собой	$\delta_{Y X}^2 = \frac{1}{\sigma_Y^2} \mathbf{M} \left\{ [m_Y(X) - \bar{m}_Y(X)]^2 \right\}$; $\delta_{X Y}^2 = \frac{1}{\sigma_X^2} \mathbf{M} \left\{ [m_X(Y) - \bar{m}_X(Y)]^2 \right\}$	$\eta_{Y X}^2 = \rho_{XY}^2 + \delta_{Y X}^2$; $v_{Y X}^2 = \frac{\eta_{XY}^2 - \rho_{XY}^2}{\eta_{Y X}^2}$	$\eta_{X Y}^2 = \rho_{XY}^2 + \delta_{X Y}^2$; $v_{X Y}^2 = \frac{\eta_{XY}^2 - \rho_{XY}^2}{\eta_{X Y}^2}$	$\eta \neq 0$
10	Связь с коэффициентом регрессии	$\rho_{XY}^2 = r_{Y X} \cdot r_{X Y}$ (см. (4.38) и (1.24))	$\eta_{Y X}^2 = r_{Y X} \cdot r_{X Y} + \delta_{Y X}^2$	$\eta_{X Y}^2 = r_{Y X} \cdot r_{X Y} + \delta_{X Y}^2$	

При этом другое КО может принимать любое значение из диапазона [0, 1] (см. рис. 4.5) в зависимости от того, какой вид имеет функция, обратная к $f(\cdot)$ для $\eta_{X|Y} = 1$, или к $\varphi(\cdot)$ при $\eta_{X|Y} = 1$ для $\eta_{Y|X}$ и какой закон распределения X для $\eta_{X|Y}$ или Y для $\eta_{Y|X}$. Если же зависимость X и Y взаимно однозначная, то $\eta_{Y|X} = \eta_{X|Y} = 1$. Верно и обратное утверждение.

Матрица соответствия задач и характеристик связи

Содержание задачи	Характеристики связи									
	ρ_{XY}	χ_{XY}	$\eta_{Y X}$	$\eta_{X Y}$	ρ, χ	$(\rho, \eta_{Y X})$	$(\rho, \eta_{X Y})$	$(\rho, \eta_{Y X}, \eta_{X Y})$	$(\chi, \eta_{Y X}, \eta_{X Y})$	$(\rho, \chi, \eta_{Y X}, \eta_{X Y})$
1. Определение направления (тенденции) связи	+	+	-	-	+	+	+	+	++	++
2. Обнаружение наличия функциональной связи:	*				*	*	*	*	*	*
а) линейной	+	+	+	+	+	+	+	+	+	+
б) нелинейной взаимно однозначной	-	*	+	+	-	-	+	+	+	+
в) нелинейной однозначной $Y=f(X)$	-	-	+	+	-	+	+	+	+	+
г) нелинейной однозначной $X=f(Y)$	-	-	-	+	-	-	+	+	+	+
3. Выявление вида функциональной связи, т. е. а, б, в, г и др. рис. 4.4, 4.5 или более сложной	-	-	-	-	-	-	-	-	-	+
4. Определение тесноты (степени статистичности) связи:										
а) линейной	+	+	+	+	+	+	+	+	+++	+++
б) нелинейной взаимно однозначной	-	+	+	+	+	+	+	+	+	+
в) нелинейной однозначной по оси y	-	-	+	-	-	+	-	+	+	++
г) нелинейной однозначной по оси x	-	-	-	+	-	-	+	+	+	++
5. Определение степени нелинейности соответствующей функции регрессии:										
а) $m_X(x)$	-	-	-	-	-	+	-	+	-	+
б) $m_Y(y)$	-	-	-	-	-	-	+	+	-	+
в) обеих	-	-	-	-	-	-	+	+	-	+
6. Пригодность для любых случайных величин X, Y	-	+	-	-	-	-	-	-	-	-

* Определяется взаимно однозначно.

Таблица П2.5

Основные соотношения и свойства корреляционных функций

№ п/п	Наименование соотношения, свойства	Определяющие условия	Соотношения и свойства				Основа (идея) доказательства
			$R_{X(t_1, t_2)}$	$\rho_{XX}(t_1, t_2)$	$R_{X_1 X_2}(t_1, t_2)$	$\rho_{X_1 X_2}(t_1, t_2)$	
1	Связь с вероятностными характеристистиками	<p>3. $X(t), \mathbf{X}(t)$ – дискретные процессы</p> <p>$\xi_j = x_j - \mathbf{M}\{X(t_j)\}$ для $X(t)$;</p> <p>$\xi_k = x_k - \mathbf{M}\{X(t_k)\}$ для $\mathbf{X}(t)$;</p> <p>$P_{\theta n}(t_1, t_2) = P\left\{\sum_{k=1, k \neq n}^n (X_k(t_1) - x_{1,k}) \cap \sum_{k=1, k \neq n}^n (X_k(t_2) - x_{2,n})\right\}$,</p> <p>$P_{k, n}(t) = P\{X_k(t) = x_{1,k}\}$</p>	<p>4. $R_{X(t_1, t_2)}$</p> <p>1. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_X(x_1, x_2; t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_{X, X}(x_1, x_2; t_1, t_2)$</p> <p>2. $\frac{\partial^2 \Theta}{\partial x_1 \partial x_2}(t_1, t_2; t_1, t_2) \Big _{x_1 = x_2 = 0} = \frac{\partial^2 H_x}{\partial x_1 \partial x_2}(t_1, t_2; t_1, t_2) \Big _{x_1 = x_2 = 0}$</p> <p>3. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_X(x_1, x_2; t_1, t_2) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_X(x_1, x_2; t_1, t_2) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_X(x_1, x_2; t_1, t_2) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_X(x_1, x_2; t_1, t_2)$</p>	<p>5. $\rho_{XX}(t_1, t_2)$</p> <p>1. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_X(x_1, x_2; t_1, t_2) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1)^2 dF_X(x_1; t_1) \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_2)^2 dF_X(x_2; t_2)$</p> <p>2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 W'_X(x_1, x_2; t_1, t_2) dx \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1)^2 W'_X(x_1; t_1) dx \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_2)^2 W'_X(x_2; t_2) dx$</p> <p>3. $\sum_{k=1, k \neq n}^n \sum_{l=1, l \neq n}^n P_{k, n}(t_1) \times \sum_{l=1, l \neq n}^n \sum_{m=1, m \neq n}^n P_{k, n}(t_2)$</p>	<p>6. $R_{X_1 X_2}(t_1, t_2)$</p> <p>$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \times dF_{X_1 X_2}(x_1, x_2; t_1, t_2)$</p>	<p>7. $\rho_{X_1 X_2}(t_1, t_2)$</p> <p>$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 dF_{X_1 X_2}(x_1, x_2; t) \times \left[\int_{-\infty}^{\infty} (x_1)^2 dF_{X_1}(x_1; t_1) \times \int_{-\infty}^{\infty} (x_2)^2 dF_{X_2}(x_2; t_2) \right]^{\frac{1}{2}}$</p> <p>Аналогично $\rho_{XX}(t_1, t_2)$ с заменой $F_X(t)$ на $F_{X'}(t)$, $W'_X(t)$ на $W_{X'}(t)$, $P_{k, n}(t)$ на $P_{k, n}(t)$</p>	8

2	Стационарность	$X(t)$ – стационарный процесс, $X_1(t), X_2(t)$ – стационарно-связанные процессы	$R_{XX}(t_2 - t_1) = R_{XX}(0) \rho_{XX}(\tau)$, $\tau = t_2 - t_1$	$\rho_{XX}(t_2 - t_1) = \frac{R_{XX}(t_2 - t_1)}{R_{XX}(0)} = \frac{R_{XX}(\tau)}{\sigma_X^2}$	$R_{X_1 X_2}(t_2 - t_1) = R_{X_1 X_2}(0) \rho_{X_1 X_2}(\tau)$ $\rho_{X_1 X_2}(t_2 - t_1) = \frac{R_{X_1 X_2}(t_2 - t_1)}{R_{X_1 X_2}(0)} = \frac{R_{X_1 X_2}(\tau)}{\sigma_{X_1} \sigma_{X_2}}$	Определение стационарности
3	Связь с другими моментами и семинвариантами	1. $X(t), X(t)$ – произвольные процессы 2. $X(t), X(t)$ – стационарные, $t \in (-\infty, \infty)$	1. $\rho_{XX}(t_1, t_2) = [B_{XX}(t_1, t_2) - m_X(t_1)m_X(t_2)] \times [m_{2,X}(t_1) - m_X^2(t_1)] \times [m_{2,X}(t_2) - m_X^2(t_2)]^{-1/2}$ 2. $\rho_X(t) = \frac{B_X(t) - m_X^2}{m_{2,X} - m_X^2}$	$R_{X_1 X_2}(t_1, t_2) = [B_{X_1 X_2}(t_1, t_2) - m_{X_1}(t_1) \times m_{X_2}(t_2)] \times [m_{2,X}(t_1) - m_{X_1}^2(t_1)] \times [m_{2,X}(t_2) - m_{X_2}^2(t_2)]^{-1/2}$	Связь начальных и центральных моментов	
4	Независимость	$X(t_1), X(t_2)$ или $X_1(t_1), X_2(t_2)$ независимы	0	0	0	
5	Периодичность	$X(t), X_1(t), X_2(t)$ – периодические СПс с периодом T , $t \in (-\infty, \infty)$ $s, s_1, s_2 = 0, 1, 2, \dots$	$R_{XX}(t_1, t_2) = R_{XX}(t_1 \pm s_1 T, t_2 \pm s_2 T)$, $R_{XX}(t) = R_{XX}(t \pm sT)$	$\rho_{XX}(t_1, t_2) = \rho_{XX}(t_1 \pm s_1 T, t_2 \pm s_2 T)$, $\rho_{XX}(t) = \rho_{XX}(t \pm sT)$	$R_{X_1 X_2}(t_1, t_2) = R_{X_1 X_2}(t_1 \pm s_1 T, t_2 \pm s_2 T)$, $R_{XX}(t) = R_{XX}(t \pm sT)$	Учесть, что $X(t) = X(t \pm sT)$
6	Свойство симметрии (четности)	$X(t), X(t)$ – произвольные процессы	$R_{XX}(t_1, t_2) = R_{XX}(t_2, t_1)$, $R_{XX}(t) = R_{XX}(-t)$	$\rho_{XX}(t_1, t_2) = \rho_{XX}(t_2, t_1)$, $\rho_{XX}(t) = \rho_{XX}(-t)$	$\rho_{X_1 X_2}(t_1, t_2) = \rho_{X_2 X_1}(t_2, t_1)$, $\rho_{X_1 X_2}(t) = \rho_{X_2 X_1}(-t)$	Использовать неравенство Коши-Буняковского-Шварца
7	Экстремальные значения		$ R_{XX}(t_1, t_2) \leq \sigma_{X_1}(t_1) \sigma_{X_1}(t_2)$, $ R_{XX}(t) \leq \sigma_X^2 = R_{XX}(0)$	$ R_{XX}(t_1, t_2) \leq \sigma_{X_1} \sigma_{X_2}$	$ R_{X_1 X_2}(t_1, t_2) \leq \sigma_{X_1} \sigma_{X_2}$	

Окончание табл. П2.5

1	2	3	4	5	6	7	8
8	Инвариантность к задержке	$X(t)$, $X(t)$ – произвольные процессы, $t \in (-\infty, \infty)$, $Y(t) = X(t + \theta)$, $Y_1(t) = X_1(t + \theta_1)$, $Y_2(t) = X_2(t + \theta_2)$	$R_{YY}(t_1, t_2) = R_{XX}(t_1 + \theta_1, t_2 + \theta)$, $R_{YY}(\tau) = R_{XX}(\tau)$	$\rho_{YY}(t_1, t_2) = \rho_{XX}(t_1 + \theta, t_2 + \theta)$, $\rho_{YY}(\tau) = \rho_{XX}(\tau)$	$R_{YY_2}(t_1, t_2) = R_{X_1 X_2}(t_1 + \theta_1, t_2 + \theta_2)$, $R_{YY_2}(\tau) = R_{X_1 X_2}(\tau + \theta_2 - \theta_1)$	$\rho_{YY_2}(t_1, t_2) = \rho_{X_1 X_2}(t_1 + \theta_1, t_2 + \theta_2)$, $\rho_{YY_2}(\tau) = \rho_{X_1 X_2}(\tau + \theta_2 - \theta_1)$	Определения
9	Неотрицательная определенность	$X(t)$ – произвольный СП; $A > 0$ – произвольное вещественное число; $\varphi(t)$ – любая комплексная неслучайная функция; m – натуральное число; «*» – знак комплексной сопряженности	$A \int_0^A \int_0^A R_{XX}(t_1, t_2) \varphi(t_1) \times \varphi^*(t_2) dt_1 dt_2 \geq 0$; $\sum_{k=1}^m R_{XX}(t, k) \varphi(k) \varphi^*(k) \geq 0$	$A \int_0^A \int_0^A R_{XX}(t_1, t_2) \varphi(t_1) \times \varphi^*(t_2) dt_1 dt_2 \geq 0$; $\sum_{k=1}^m R_{XX}(t, k) \varphi(k) \varphi^*(k) \geq 0$	–	–	Учсть, что $\mathbf{M} \left\{ \int_0^A X(t) \times \int_0^A \varphi(t) dt \right\}^2 \geq 0$
10	Поведение при изменении масштаба аргумента /	$Y(t) = X(\alpha t)$; $Y_1(t) = X_1(\alpha t)$; $Y_2(t) = X_2(\beta t)$; α, β – положительные действительные числа	$R_{YY}(t_1, t_2) = R_{XX}(\alpha t_1, \alpha t_2)$; $R_{YY}(\tau) = R_{XX}(\alpha \tau)$	$\rho_{YY}(t_1, t_2) = \rho_{XX}(\alpha t_1, \alpha t_2)$; $\rho_{YY}(\tau) = \rho_{XX}(\alpha \tau)$	$R_{YY_2}(t_1, t_2) = R_{X_1 X_2}(\alpha t_1, \beta t_2)$; $R_{YY_2}(t, t + \tau) = R_{X_1 X_2}(t, t + \tau)$; $R_{YY_2}(\tau) = R_{X_1 X_2}(\beta - \alpha) \tau + \beta \tau$	$\rho_{YY_2}(t_1, t_2) = \rho_{X_1 X_2}(\alpha t_1, \beta t_2)$; $\rho_{YY_2}(t, t + \tau) = \rho_{X_1 X_2}(t, t + \tau)$; $\rho_{YY_2}(\tau) = \rho_{X_1 X_2}(\beta - \alpha) \tau + \beta \tau$	Учсть, что $\mathbf{M} \left\{ \int_0^A X_1(t_1) \times \int_0^A X_2(t_2) \right\} = \mathbf{M} \left\{ \int_0^A X_1(\alpha t) \times \int_0^A X_2(\beta t) \right\}$
11	Связь с аналогичными характеристиками при дифференцировании в среднем квадратическом СП	$Y(t) = \frac{d^k X(t)}{dt^k} = X^{(k)}(t)$; $Y_1(t) = X_1^{(k)}(t)$; $Y_2(t) = X_2^{(k)}(t)$	$R_{YY}(t_1, t_2) = \frac{\partial^{2k} R_{XX}(t_1, t_2)}{\partial t_1^k \partial t_2^k}$; $R_{YY}(\tau) = (-1)^k \frac{d^{2k} R_{XX}(\tau)}{d\tau^{2k}}$	–	$R_{YY_2}(t_1, t_2) = \frac{\partial^{k+n} R_{XX}(t_1, t_2)}{\partial t_1^k \partial t_2^n}$; $R_{YY_2}(\tau) = (-1)^k \frac{d^{k+n} R_{XX}(\tau)}{d\tau^{k+n}}$	–	Учсть определение для $X^{(k)}(t)$

Связь с аналогичными характеристиками	$Y(t) = \int_0^t X(t-u) \times h(u) du;$ $Y_1(t) = \int_0^t X_1(t-u) \times h_1(u) du;$ $Y_2(t) = \int_0^t X_2(t-u) \times h_2(u) du$	$R_{YY}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} R_{XX}(t_1 - \tau_1, t_2 - \tau_2) \times h(\tau_1) h(\tau_2) d\tau_1 d\tau_2,$ $R_{YY}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} h(u) \times R_{XX}(t_1 - t_2 - u) du$		$R_{YY}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} R_{X_1 X_2}(t_1 - \tau_1; t_2 - \tau_2) h(\tau_1) h(\tau_2) d\tau_1 d\tau_2$	Учесть, что $\mathbf{M} \left\{ \begin{pmatrix} t \\ \tau \end{pmatrix} \right\} = \begin{pmatrix} t \\ 0 \end{pmatrix} = \mathbf{M} \{ \tau \}$
12 Изменения при других линейных преобразованиях СП	$Y(t) = a(t) + \lambda(t) X(t);$ $Y_1(t) = a_1(t) + \lambda_1(t) X_1(t);$ $Y_2(t) = a_2(t) + \lambda_2(t) X_2(t)$	$R_{YY}(t_1, t_2) = \lambda(t_1) \lambda(t_2) R_{XX}(t_1, t_2)$	$\rho_{YY}(t_1, t_2) = \rho_{XX}(t_1, t_2) \times \text{sign}[\lambda(t_1) \lambda(t_2)]$	$R_{Y_1 Y_2}(t_1, t_2) = \rho_{Y_1 Y_2}(t_1, t_2) \times \text{sign}[\lambda_1(t_1) \lambda_2(t_2)]$	См. свойства 2 и 3
14 Значения для суммы СП	$Y(t) = \sum_{i=1}^n \lambda_i(t) X_i(t);$ $Z(t) = \sum_{i=1}^n \lambda_i(t) X_i(t);$ $V(t) = \sum_{i=1}^n v_i(t) Y_i(t)$	$R_{YY}(t_1, t_2) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i(t_1) \lambda_j(t_2) \times R_{X_i X_j}(t_1, t_2)$		$R_{ZV}(t_1, t_2) = \sum_{i=1}^n \sum_{j=1}^k \lambda_i(t_1) v_j(t_2) \times R_{X_i Y_j}(t_1, t_2)$	Учесть, что $\mathbf{M} \left\{ \sum \tau \right\} = \sum \mathbf{M} \{ \tau \}$

Предостережение. Обратим внимание на то, что, как и для оценок коэффициентов корреляции, для оценок НКФ $\rho_{XX}(\tau)$, $\rho_{XY}(\tau)$ имеют место смещение (см. [5]) и слабая сходимость распределения оценок к асимптотическому. Следовательно, требуют аккуратного применения при малых объемах выборки как значения $\hat{\rho}(\tau)$, так и критерии проверки гипотез о равенстве ρ заданному (например, нулевому) значению даже в условиях справедливости гипотез о виде двумерного распределения $X(t)$, $Y(t)$.

Некоторые специфические свойства КФ, ККФ, ДФ стационарных СФ

Свойство	КФ $R_{XY}(\tau)$	ККФ $K_{XY}(\tau)$	ДФ $Z_{XY}(\tau), Z_{Y X}(\tau)$
Связь с другими характеристиками	$\mathbf{M}\{X(t)Y(t+\tau)\} = m_X m_Y;$ $\rho_{XY}(\tau) = R_{XY}(\tau) / \sigma_X \sigma_Y,$ где $\sigma_X = \sqrt{R_{XX}(0)}$	$K_{XY}(\tau) / \sqrt{K_{XX}(0)K_{YY}(0)} = \chi_{XY}(\tau)$ Для случайных процессов с абсолютно непрерывными распределениями $\mathbf{M}\{F_X[X(t)]F_Y[Y(t+\tau)]\} = \frac{1}{4}$; $\chi_{XY}(\tau) = 12K_{XY}(\tau)$	$Z_{XY}(\tau) =$ $= \mathbf{M}\left\{m_X^2 [Y; t, t+\tau]\right\} - m_X^2;$ $\eta_{XY}(\tau) = \sqrt{Z_{XY}(\tau) / \sigma_X^2};$ $\zeta_{XY}(\tau) = 1 - Z_{XY} / \sigma_X^2$
Симметричность	$R_{XY}(-\tau) = R_{YX}(\tau); R_{XX}(\tau) = R_{XX}(-\tau)$	$K_{XY}(-\tau) = K_{YX}(\tau)$	$Z_{XY}(\tau) \neq Z_{Y X}(\tau)$
Экстремальные значения	$ R_{XY}(\tau) \leq \sqrt{R_{XX}(0)R_{YY}(0)};$ $ R_{XY}(\tau) \leq \frac{1}{2}(\sigma_X^2 + \sigma_Y^2);$ $ \rho_{XY}(\tau) \leq 1; \rho_{XX}(0) = 1$	$ K_{XY}(\tau) \leq \sqrt{K_{XX}(0)K_{YY}(0)};$ $ \chi_{XY}(\tau) \leq 1; \chi_{XX}(0) = 1$	$0 \leq Z_{XY}(\tau) \leq \sigma_X^2;$ $Z_{Y X}(\tau) Z_{XY}(-\tau) \leq \sigma_Y^2 \sigma_X^2;$ $0 \leq \eta_{XY}(\tau) \leq 1, 0 \leq \zeta_{XY}(\tau) \leq 1;$ $\eta_{Y X}(0) = 1$
Характерные соотношения	$R_{XY}(\tau)$ существенно зависит от параметров формы распределения	$K_{XY}(\tau)$ не зависит от параметров формы семейства распределений	В общем случае $\eta_{Y X}(\tau)$ и $\eta_{X Y}(\tau) \geq \rho_{XY}(\pm\tau) $ Для случайных процессов с линейной регрессией $\eta_{Y X}(\tau) = \eta_{X Y}(-\tau) = \rho_{XY}(-\tau) $
Инвариантность к функциональным преобразованиям $U(t) = f[X(t)];$ $V(t) = g[Y(t)]$	$f(\cdot), g(\cdot)$ – линейные функции, т.е. $f(x) = ax + b, g(y) = cy + d$; $\rho_{UV}(\tau) = \rho_{XY}(\tau) \text{sign}(ac)$	$f(\cdot), g(\cdot)$ – любые монотонные взаимно однозначные функции $K_{UV}(\tau) = \varepsilon_f \varepsilon_g K_{XY}(\tau);$ $\varepsilon_f = \text{sign}\{df(x)/dx\}$	$f(\cdot), g(\cdot)$ – любые однозначные функции $\eta_{UV}(\tau) = \eta_{XY}(\tau)$

<p>Учет детерминированных связей $Y(t) = f[X(t)]$</p>	<p>$f(\cdot)$ – линейная функция, т. е. $f(x) = ax + b$, $\rho_{XY}(\tau) = \rho_{XX}(\tau) \text{sign}(a)$</p>	<p>$f(\cdot)$ – любая монотонная взаимно однозначная функция $\chi_{XY}(\tau) = \varepsilon_f \chi_{XX}(\tau)$</p>	<p>$f(\cdot)$ – любая однозначная функция $\eta_{YX}(\tau) = \eta_{XY}(\tau)$; $f(\cdot)$ – линейная функция $\eta_{YX}(\tau) = \eta_{XY}(\tau) = \rho_{XY}(\tau)$</p>
<p>Периодичность</p>	<p>$X(t)$ и $Y(t)$ – периодические функции с периодом T_0, т. е. $X(t) = X(t \pm kT_0)$, $Y(t) = Y(t \pm nT_0)$, $k, n = 0, 1, 2, \dots$; $\rho_{XY}(\tau)$, $\chi_{XY}(\tau)$, $\eta_{YX}(\tau)$ и $\eta_{XY}(\tau)$ – периодические функции с периодом T_0</p>		
<p>Связь между собой для разных $X(t)$, $Y(t)$: $X(t)$, $Y(t + \tau)$ – гауссовы СФ с параметром связи $\psi(\tau)$</p>	<p>$\rho_{XY}(\tau) = \psi(\tau)$</p>	<p>$\chi_{XY}(\tau) = \frac{6}{\pi} \arcsin \left[\frac{\psi(\tau)}{2} \right] \approx \rho_{XY}(\tau)$</p>	<p>$\eta_{Y X}(\tau) = \eta_{XY}(-\tau) =$ $= \rho_{XY}(-\tau) = \psi(-\tau)$</p>
<p>$X(t)$, $Y(t + \tau)$ имеют двумерное распределение Вейбулла с параметром связи $\psi(\tau)$ [20]</p>	<p>$\rho_{XY}(\tau)$ связана с $\psi(\tau)$ через вырожденную гипергеометрическую функцию</p>	<p>$\chi_{XY}(\tau) = 3\psi^2(\tau) [4 - \psi^2(\tau)]^{-1}$</p>	<p>Сложная связь с $\psi(\tau)$</p>
<p>$X(t) = a \sin(\lambda_0 t + \Phi)$, Φ – равномерно распределенная на $(0, 2\pi)$ случайная величина</p>	<p>$\rho_{XX}(\tau) = \cos(\lambda_0 \tau)$</p>	<p>$\chi_{XX}(\tau) = \frac{96}{\pi^4} \sum_{k=1}^{\infty} \frac{\cos[(2k-1)\lambda_0 \tau]}{(2k-1)^4} \approx$ $\approx \rho_{XX}(\tau)$</p>	<p>$\eta_{Y X}(\tau) = \eta_{XY}(\tau) = \rho_{YX}(\tau)$</p>

Основные соотношения и свойства спектральных плотностей стационарных СФ (!)

№ п/п	Наименование соотношения, свойства	Определяющие условия	Соотношения и свойства				Примечания
			$S_{XX}(\omega)$	$s_{XX}(\omega)$	$S_{XY}(\omega)$	$s_{XY}(\omega)$	
1	Связь с другими характеристиками	Для процессов, у которых существует $R(\tau)$, $\rho(\tau)$	$S_{XX}(\omega) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-j\omega\tau} d\tau;$ $R_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XX}(\omega) \times \exp\{j\omega\tau\} d\omega$	$s_{XX}(\omega) = \int_{-\infty}^{\infty} \rho_{XX}(\tau) e^{-j\omega\tau} d\tau;$ $\rho_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} s_{XX}(\omega) \times \exp\{j\omega\tau\} d\omega$	$S_{XY}(\omega) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j\omega\tau} d\tau;$ $R_{XY}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}(\omega) \times \exp\{j\omega\tau\} d\omega$	$s_{XY}(\omega) = \int_{-\infty}^{\infty} \rho_{XY}(\tau) e^{-j\omega\tau} d\tau;$ $\rho_{XY}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} s_{XY}(\omega) \times \exp\{j\omega\tau\} d\omega$	Между СПМ и КФ взаимно однозначное соответствие
2	Независимость, δ -коррелированность, δ -дельта-функция Дирака	$X(t)$ – δ -коррелированные $R_{XX}(\tau) = \sigma_X^2 \delta(\tau)$, $X(t)$ и $Y(t)$ – независимы	$S_{XX}(\omega) = \sigma_X^2$	$s_{XX}(\omega) = 1$	$S_{XY}(\omega) = 0$	$s_{XY}(\omega) = 0$	Если, конечно, $S(\omega)$ и $s(\omega)$ существуют
3	Экстремальные значения		$S_{XX}(\omega) \geq 0$ (*)	$s_{XX}(\omega) \geq 0$	$ S_{XY}(\omega) ^2 = S_{XY}(\omega) S_{YX}(\omega) \leq S_{XX}(\omega) S_{YY}(\omega)$	$ s_{XY}(\omega) \leq \sqrt{s_{XX}(\omega) s_{YY}(\omega)}$	(*) Следствие неотрицательной неотрицательности $R_{XX}(\tau)$
4	Комплексной сопряженности (симметрии, четности)	Для всех процессов, для которых $S(\omega)$, $s(\omega)$ существуют	$S_{XX}(\omega) = S_{XX}(-\omega)$	$s_{XX}(\omega) = s_{XX}(-\omega)$	$S_{XY}(\omega) = P(\omega) - jQ(\omega) = \tilde{S}_{YX}(\omega);$ $P_{XY}(\omega) = P_{YX}(-\omega);$ $Q_{XY}(\omega) = -Q_{YX}(-\omega);$ $S_{XY}(\omega) = S_{XY}(\omega) \times \exp\{-j\theta_{XY}(\omega)\};$ $\theta_{XY}(\omega) = \arctg \left[\frac{Q_{XY}(\omega)}{P_{XY}(\omega)} \right]$	$s_{XY}(\omega) = P_{XY}(\omega) - jQ_{XY}(\omega) = \tilde{s}_{YX}(\omega);$ $P_{XY}(\omega) = P_{YX}(-\omega);$ $Q_{XY}(\omega) = -Q_{YX}(-\omega)$	\sim – знак комплексной сопряженности $(a + jb) = a - jb$, $P(\omega)$ – коспектральная плотность; $Q(\omega)$ – квадратурная спектральная плотность

5	Периодичность, T – период	$X(t) = X(t \pm KT)$; $Y(t) = Y(t \pm KT)$	$S_{XX}(\omega) = \sum_{k=-\infty}^{\infty} A_k \delta\left(\omega + \frac{2k\pi}{T}\right)$	$S_{XX}(\omega) = \sum_{k=-\infty}^{\infty} a_k \delta\left(\omega + \frac{2k\pi}{T}\right)$	$S_{XY}(\omega) = \sum_{k=-\infty}^{\infty} Z_k \delta\left(\omega + \frac{2k\pi}{T}\right)$	$s_{XY}(\omega) = \sum_{k=-\infty}^{\infty} z_k \delta\left(\omega + \frac{2k\pi}{T}\right)$	Спектры – решетчатые функции
6	Связь с мощностью процессов	Для всех стационарных СФ, у которых $S(\omega)$ существуют	$\int_{-\infty}^{\infty} S_{XX}(\omega) d\omega = 2 \int_0^{\infty} S_{XX}(\omega) d\omega = \sigma_X^2$ есть мощность (дисперсия) процесса	$\int_{-\infty}^{\infty} S_{XX}(\omega) d\omega = 2 \int_0^{\infty} S_{XX}(\omega) d\omega = 1$	$\int_{-\infty}^{\infty} S_{XY}(\omega) d\omega = R_{XY}(0) = \sigma_X \sigma_Y \rho_{XY}(0)$	$\int_{-\infty}^{\infty} s_{XY}(\omega) d\omega = \rho_{XY}(0)$	$S(\omega)$ – отношение мощностей в полосе $\left(\omega - \frac{\Delta\omega}{2}, \omega + \frac{\Delta\omega}{2}\right)$ к $\Delta\omega$ при $\Delta\omega \rightarrow 0$
7	Определение мощности в полосе частот	$P_{XX}(\omega_1, \omega_2)$ – мощность в полосе частот (ω_1, ω_2) , $\omega_2 \geq \omega_1$	$P_{XX}(\omega_1, \omega_2) = 2 \int_{\omega_1}^{\omega_2} S_{XX}(\omega) d\omega$	$P(\omega_1, \omega_2) = P(-\infty, \infty) = 2 \int_{\omega_1}^{\omega_2} s_{XX}(\omega) d\omega$	$P_{XY}(\omega_1, \omega_2) = \int_{\omega_1}^{\omega_2} S_{XY}(\omega) d\omega + \int_{-\omega_1}^{-\omega_2} S_{XY}(\omega) d\omega$	$P_{XY}(\omega_1, \omega_2) = P_{XY}(-\infty, \infty) = \int_{\omega_1}^{\omega_2} s_{XY}(\omega) d\omega + \int_{-\omega_1}^{-\omega_2} s_{XY}(\omega) d\omega$	Для $X(t)$, $Y(t)$ – взаимная мощность $P_{XY}(\omega_1, \omega_2)$
8	Реакция на задержку процесса во времени	$Z(t) = X(t + \theta)$, $Y(t) = Y(t + \theta)$, $ \theta < \infty$, $ \theta < \infty$	$S_{ZZ}(\omega) = S_{XX}(\omega)$; $S_{YY}(\omega) = S_{YY}(\omega)$	$s_{YZ}(\omega) = s_{YX}(\omega) \times \exp(-j\omega\theta)$	$S_{ZY}(\omega) = S_{YX}(\omega) \times \exp(-j\omega(\theta - 0))$	$s_{ZY}(\omega) = s_{YX}(\omega) \times \exp(-j\omega(\theta - 0))$	
9	Реакция на изменение масштаба времени	$Z(t) = X(\alpha t)$, $F(t) = Y(\beta t)$, $\alpha > 0$, $\beta > 0$	$S_{ZZ}(\omega) = \frac{1}{\alpha} S_{XX}\left(\frac{\omega}{\alpha}\right)$	$s_{YZ}(\omega) = \frac{1}{\beta} s_{YX}\left(\frac{\omega}{\beta}\right)$	$S_{ZY}(\omega)$, $S(\omega)$ – не определена из-за нестационарности при $\alpha \neq \beta$	$s_{ZY}(\omega) = \frac{1}{\alpha} s_{YX}\left(\frac{\omega}{\alpha}\right)$ при $\alpha = \beta$	Свойство можно использовать для проверки взаимной нестационарности $Z(t)$ и $F(t)$
10	Линейность	$Z(t) = a + \lambda X(t)$, $F(t) = b + \nu Y(t)$, $ a , b < \infty$, $ \lambda , \nu < \infty$	$S_{ZZ}(\omega) = \lambda^2 S_{XX}(\omega)$; $S_{XZ}(\omega) = \lambda S_{XX}(\omega)$	$s_{YZ}(\omega) = s_{YX}(\omega)$; $s_{YY}(\omega) = s_{YY}(\omega)$	$S_{ZY}(\omega) = \lambda \nu S_{XY}(\omega)$	$s_{ZY}(\omega) = s_{XY}(\omega)$	
11	Значения для сумм	$Z(t) = \sum_{j=1}^n \lambda_j X_j(t)$, $F(t) = \sum_{j=1}^m \nu_j Y_j(t)$	$S_{ZZ}(\omega) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j S_{X_i X_j}(\omega)$		$S_{ZY}(\omega) = \sum_{i=1}^n \sum_{j=1}^m \lambda_i \nu_j S_{X_i Y_j}(\omega)$		

Окончание табл. П2.7

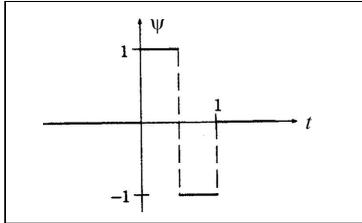
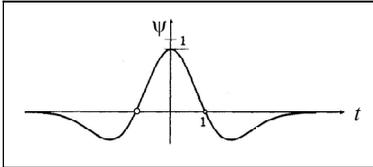
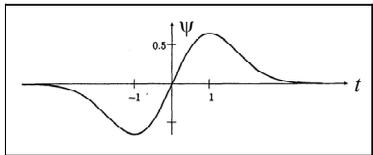
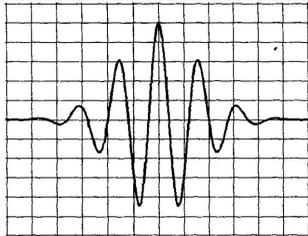
№ п/п	Наименование соотношения, свойства	Определяющие условия	Соотношения и свойства			Примечания	
			$S_{XX}(\omega)$	$S_{YY}(\omega)$	$S_{XY}(\omega)$		$S_{XY}(\omega)$
12	Изменения при дифференцировании в среднем квадратичном	$Z(t) = \frac{d^k X(t)}{dt^k};$ $V(t) = \frac{d^n Y(t)}{dt^n}$	$S_{ZZ}(\omega) = \omega^{2k} S_{XX}(\omega)$	$S_{ZZ}(\omega) =$ $= (j\omega)^k S_{XX}(\omega)$	$S_{ZY}(\omega) = (-1)^k \times$ $\times (j\omega)^{k+n} S_{XY}(\omega)$	$S_{XY}(\omega) =$ $= (j\omega)^n S_{XY}(\omega)$	
13	Изменения при прохождении линейных систем $H(j\omega) =$ $= \int_{-\infty}^{\infty} h(t) e^{j\omega t} dt,$ $h(t) = 0, t < 0$	$Z(t) =$ $= \int_0^t X(t-u) h(u) du;$ $V(t) =$ $= \int_0^t Y(t-u) g(u) du$	$S_{ZZ}(\omega) =$ $= H(j\omega) ^2 S_{XX}(\omega);$ $S_{YZ}(\omega) =$ $= H(j\omega) S_{XY}(\omega)$	$S_{ZY}(\omega) =$ $= G(j\omega) ^2 S_{YY}(\omega);$ $S_{YZ}(\omega) =$ $= \tilde{H}(j\omega) S_{XY}(\omega)$	$S_{ZY}(\omega) = H(j\omega) \times$ $\times G(j\omega) S_{XY}(\omega);$ $S_{YZ}(\omega) = \tilde{H}(j\omega) \times$ $\times G(j\omega) S_{XY}(\omega)$		
14	Значение СПМ в нуле	$\omega = 0$	$S_{XX}(0) =$ $= 2 \int_0^{\infty} R_{XX}(\tau) d\tau$	$S_{XX}(0) =$ $= 2 \int_0^{\infty} R_{XX}(\tau) d\tau$	$S_{XY}(0) =$ $= \int_{-\infty}^{\infty} R_{XY}(\tau) d\tau$	$S_{XY}(0) =$ $= \int_{-\infty}^{\infty} R_{XY}(\tau) d\tau$	Условие нормировки

Специфические свойства конкорреляционных и дисперсионных спектральных плотностей «мощности» (!)

№ п/п	Наименование соотношения, свойства	Определяющие условия	Соотношения и свойства			Примечания
			КПМ $C_{XY}(\omega)$	ДПМ $D_{Y X}(\omega)$	ДПМ $D_{X Y}(\omega)$	
1	Связь с другими характеристиками	Для всех стационарных СФ, где характеристики существуют	$C_{XY}(\omega) = \int_{-\infty}^{\infty} K_{XY}(\tau) \exp\{-j\omega\tau\} d\tau;$ $K_{XY}(\tau) = \frac{1}{2\pi} \times \int_{-\infty}^{\infty} C_{XY}(\omega) \exp\{j\omega\tau\} d\omega$	$D_{Y X}(\omega) = \int_{-\infty}^{\infty} Z_{Y X}(\tau) \times \exp\{-j\omega\tau\} d\tau;$ $Z_{Y X}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} D_{YX}(\omega) \times \exp\{j\omega\tau\} d\omega$	$D_{X Y}(\omega) = \int_{-\infty}^{\infty} Z_{X Y}(\tau) \times \exp\{-j\omega\tau\} d\tau;$ $Z_{X Y}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} D_{YX}(\omega) \times \exp\{j\omega\tau\} d\omega$	Для нормированных характеристик аналогично
2	Область существования	Виды случайных функций	Любые СФ	СФ, имеющие моменты не ниже 2-го порядка	СФ, имеющие моменты не ниже 2-го порядка	
3	Связь с «мощностью»	$Z_{Y X}(\tau) = \mathbf{M} \left\{ m^2 \begin{matrix} X(t); t+\tau \\ Y \end{matrix} \right\},$ $Z_{X Y}(\tau) = \mathbf{M} \left\{ m^2 \begin{matrix} Y(t+\tau); t \\ X \end{matrix} \right\},$ $m_Y(x; \tau) = \mathbf{M} \{ Y(t+\tau) X(t) = x \} -$ условное математическое ожидание $Y(t)$, функция регрессии	$C_{XX}(0) = \int_{-\infty}^{\infty} K_{XX}(\tau) d\tau =$ $= \mathbf{D} \{ F_X [X(t)] \} = P_{F_X}(\omega)$	$D_{Y X}(\omega) = \int_{-\infty}^{\infty} Z_{Y X}(\tau) d\tau =$ $= \int_{-\infty}^{\infty} \mathbf{D} \left\{ m_Y [X(t); \tau] \right\} d\tau$	$D_{YX}(\omega) = \int_{-\infty}^{\infty} Z_{Y X}(\tau) d\tau =$ $= \int_{-\infty}^{\infty} \mathbf{D} \left\{ m_Y \begin{matrix} Y(t+\tau); \tau \\ X \end{matrix} \right\} d\tau$	
4	Изменение при нелинейных преобразованиях	$Z(t) = f[X(t)];$ $f'(t) = g[Y(t)];$ $f'(t), g(t) - \text{взаимно однозначные монотонные функции}$	$C_{Z Y}(\omega) = \varepsilon_f \varepsilon_g C_{XY}(\omega),$ $\varepsilon_f = \text{sign} \left[\frac{df(x)}{dx} \right],$	$f'(t) - \text{однозначная функция}$ $D_{Y X}(\omega) = \frac{\sigma_f^2}{\sigma_X^2} D_{YX}(\omega)$	$g(t) - \text{однозначная функция}$ $D_{Y X}(\omega) = \frac{\sigma_f^2}{\sigma_Y^2} D_{YX}(\omega)$	

№ п/п	Наименование соотношения, свойства	Определяющие условия	Соотношения и свойства			Примечания
			КПМ $C_{XY}(\omega)$	ДПМ $D_{YX}(\omega)$	ДПМ $D_{XY}(\omega)$	
5	Реакция на временную задержку	$Z(t) = f[X(t+\theta)],$ $Y(t) = g[Y(t+\theta)],$ $f(\cdot), g(\cdot)$	$C_{ZY}(\omega) = \varepsilon_f \varepsilon_g C_{XY}(\omega) \times$ $\times \exp\{-j\omega(\theta - \theta)\}$	ДПМ $D_{YX}(\omega)$ При $Z(t) = X(t+\theta),$ $Y(t) = Y(t+\theta)$ $D_{ZY}(\omega) = D_{YX}(\omega) \times$ $\times \exp\{j\omega\theta\}$	ДПМ $D_{XY}(\omega)$ При $Z(t) = X(t+\theta),$ $D_{ZY}(\omega) = D_{YX}(\omega) \times$ $\times \exp\{j\omega\theta\}$	
6	Прохождение через линейные инерционные системы	$Y(t) = \int_0^t X(t-u)h(u)du,$ $K_{XY}(j\omega) = \int_0^\infty k(u)e^{-j\omega u} du =$ конкордатная характеристика системы, $k(u)$ – конкорватная функция	$C_{XY}(\omega) = K_{XY}(j\omega)C_{XX}(\omega),$ $C_{YX}(\omega) = K_{XY}(j\omega) ^2 C_{XX}(\omega)$ (по определению $K(j\omega)$)			
7	Прохождение через линейные инерционные системы Гаммерштейна-Винера	$Z(t) = f[X(t)],$ $Y(t) = \int_0^t Z(t-u)h(u)du,$ $Y(t) = g[Y(t)],$ $f(\cdot), g(\cdot)$ – взаимно однозначные монотонные функции	$C_{XX}(\omega) K_{ZY}(j\omega) ^2,$ $C_{XY}(\omega) = \varepsilon_f \varepsilon_g C_{XX}(\omega) \times$ $\times K_{ZY}(j\omega)$			

**Примеры вейвлетных функций («основных вейвлетных
(волновых) пакетов») $\psi(t)$**

№ п/п	Наименование	Аналитическое описание $\psi(t)$	Графическое представление
1	Хаара	$\begin{cases} 1, & 0 \leq t < 0,5; \\ -1, & 0,5 \leq t < 1; \\ 0 & \text{иначе} \end{cases}$	
2	Мексиканская шляпа	$\frac{2(1-t^2)}{\sqrt{3}\sqrt{\pi}} e^{-\frac{t^2}{2}}$	
4	Производная гауссовой кривой	$\sqrt{\frac{2}{\pi}} t e^{-\frac{t^2}{2}}$	
5	Морле	$\exp\left\{j\pi t - \frac{t^2}{2\lambda^2}\right\};$ $\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(t-\tau)^2}{2\lambda^2} + j2\pi \frac{(t-\tau)}{\lambda}\right\}$	

Внимание! Запишите самостоятельно функции 1–4 в обобщенном виде по типу функции 5.

НЕКОТОРЫЕ НЕРАВЕНСТВА И ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ
ТЕОРИИ ВЕРОЯТНОСТЕЙ¹

ПЗ.1. Проблема моментов

Прямая проблема. Для всякого ли распределения вероятностей случайных величин существуют моменты (начальные m_k , центральные μ_k) либо родственные им характеристики (например, кумулянты χ_k)? Ответ отрицательный. Для того чтобы для конечного k существовал момент m_k (или μ_k , χ_k), достаточно, чтобы выполнялось условие $\mathbf{M}\{|\mathbf{X}|^k\} < \infty$. Примеры распределений, для которых не существуют моменты даже при малых $k = 1, 2, 3, \dots$, можно найти в [20].

Обратная проблема. Если распределение имеет соответствующие моменты, то они однозначно определяются им (по нему). А верно ли обратное утверждение: можно ли однозначно определить вид закона распределения вероятностей и охарактеризовать его свойства по ее заданной счетной² совокупности моментов? В общем виде ответ отрицательный – не каждое распределение однозначно определяется своими моментами. Основные вопросы, решаемые в рамках этой проблемы: разрешимость – существует ли хотя бы одна функция $W(x)$, имеющая заданные моменты m_k , μ_k (или χ_k); определенность и неопределенность – при каких ограничениях на последовательности моментов существует лишь одна функция $W(x)$, имеющая заданные моменты (определенность), и в каких случаях таких функций может быть несколько (неопределенность); проблема класса – при каких условиях последовательность m_k (или μ_k , χ_k) является последовательностью моментов для функций из того или иного класса.

Чтобы распределение однозначно определялось всеми своими моментами, достаточно выполнения условия Карлемана

$$\sum_{k=0}^{\infty} (m_{2k})^{-\frac{1}{2k}} = \infty. \quad (\text{ПЗ.1})$$

¹ Для примера рассмотрим только случайные величины X, Y, Z .

² Понятно, что одну и ту же конечную совокупность моментов без каких-либо дополнительных условий может иметь много распределений.

Например, неоднозначно определяется заданной счетной совокупностью моментов распределение $(X - m_X)^k$ при $k \geq 3$, если X – нормально распределенная случайная величина (другие примеры см. в [20]).

П3.2. Неравенства для моментных характеристик

Неравенство Ляпунова

$$\left[\mathbf{M}\{|X|^\alpha\} \right]^{\frac{1}{\alpha}} \leq \left[\mathbf{M}\{|X|^\beta\} \right]^{\frac{1}{\beta}}, \quad 0 < \alpha \leq \beta. \quad (\text{П3.2})$$

Неравенство Гельдера–Минковского

$$\mathbf{M}\{XY\} \leq \left[\mathbf{M}\{|X|^p\} \right]^{\frac{1}{p}} \left[\mathbf{M}\{|Y|^q\} \right]^{\frac{1}{q}}, \quad (\text{П3.3})$$

если

$$\frac{1}{p} + \frac{1}{q} = 1, \quad p > 1, \quad q > 1, \quad \mathbf{M}\{|X|^p\} < \infty, \quad \mathbf{M}\{|Y|^q\} < \infty.$$

Неравенство Коши–Буняковского–Шварца

$$\mathbf{M}\{XY\} \leq \left[\mathbf{M}\{|X|^2\} \mathbf{M}\{|Y|^2\} \right]^{\frac{1}{2}}. \quad (\text{П3.4})$$

Неравенство Иенсена

$$f(\mathbf{M}\{X\}) \leq \mathbf{M}\{f(X)\}, \quad (\text{П3.5})$$

если $f(\cdot)$ – выпуклая функция, т. е. если $f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y)$.

П3.3. Неравенства для вероятностей

Неравенство Чебышева 1-го рода

$$\left. \begin{aligned} \mathbf{P}\{X \geq \varepsilon\} &\leq \frac{\mathbf{M}\{X\}}{\varepsilon}, \quad X \geq 0, \quad \varepsilon > 0; \\ \mathbf{P}\{|X| \geq \varepsilon\} &\leq \frac{\mathbf{M}\{|X\}}{\varepsilon}, \quad \varepsilon > 0 \end{aligned} \right\}. \quad (\text{П3.6})$$

Неравенство Чебышева 2-го рода

$$\mathbf{P}\{|X - \mathbf{M}\{X\}| \geq \varepsilon\} \leq \frac{\mathbf{D}\{X\}}{\varepsilon^2}. \quad (\text{П3.7})$$

Неравенство Маркова

$$\mathbf{P}\{|X| \geq \varepsilon\} \leq \frac{\mathbf{M}\{|X|^k\}}{\varepsilon^k}. \quad (\text{П3.8})$$

Неравенство Пирсона

$$\mathbf{P}\{|X - \mathbf{M}\{X\}| \geq \varepsilon\} \leq \frac{\mathbf{M}\{|X - m_X|^k\}}{\varepsilon^k}. \quad (\text{П3.9})$$

Двумерный аналог неравенства Чебышева 2-го рода. Пусть X и Y – случайные величины с коэффициентом корреляции ρ_{XY} . Для них

$$\mathbf{P}\{(|X - m_X| \geq \varepsilon\sigma_X) \text{ или } (|Y - m_Y| \geq \varepsilon\sigma_Y)\} \leq \frac{1 + \sqrt{1 - \rho_{XY}^2}}{\varepsilon^2}. \quad (\text{П3.10})$$

П3.4. Основные виды вероятностной сходимости

Последовательность случайных величин X_1, X_2, \dots, X_N называется сходящейся при $N \rightarrow \infty$ к случайной величине X (в частности к константе $X = a$):

а) *по вероятности* (краткое обозначение $X_N \xrightarrow{\mathbf{P}} X$ при $N \rightarrow \infty$, или $X_N \xrightarrow{\text{ПВ}} X$ при $N \rightarrow \infty$), если для любого $\varepsilon > 0$:

$$\lim_{N \rightarrow \infty} \mathbf{P}\{|X_N - X| \geq \varepsilon\} = 0 \quad (\text{П3.11})$$

или

$$\lim_{N \rightarrow \infty} \mathbf{P}\{|X_N - X| < \varepsilon\} = 1; \quad (\text{П3.11a})$$

б) *почти наверное или с вероятностью единица* ($X_N \xrightarrow{1} X$ при $N \rightarrow \infty$, или $X_N \xrightarrow{\text{ПН}} X$ при $N \rightarrow \infty$), если

$$\lim_{N \rightarrow \infty} \mathbf{P}\{|X_N - X| > \varepsilon\} = 0 \quad (\text{П3.12})$$

хотя бы при одном $m > N$ и любом $\varepsilon > 0$, или

$$\mathbf{P}\left\{\lim_{N \rightarrow \infty} X_N = X\right\} = 1; \quad (\text{ПЗ.12a})$$

в) в среднем порядка k ($X_N \xrightarrow{\text{СП } k} X$ при $N \rightarrow \infty$), если

$$\lim_{N \rightarrow \infty} \mathbf{M}\left\{|X_N - X|^k\right\} = 0; \quad (\text{ПЗ.13})$$

сходимость в среднем порядка $k = 2$ называется сходимостью в *среднем квадратичном* (или по дисперсии), кратко записывается в виде $X_N \xrightarrow{D} X$ или $X_N \xrightarrow{\text{СК}} X$ при $N \rightarrow \infty$ либо в виде $\text{l.i.m. } X_N = \text{l.i.m. } X_N = X$;

г) по распределению ($X_N \xrightarrow{\text{ПР}} X$ или $X_N \xrightarrow{F} X$ при $N \rightarrow \infty$), если

$$\lim_{N \rightarrow \infty} F_{X_N}(x) = F_X(x), \quad (\text{ПЗ.14})$$

где $F_{X_N}(x), F_X(x)$ – функции распределения X_N и X .

Используя соответствующие неравенства, нетрудно убедиться, что сходимость по вероятности следует из сходимости почти наверное и из сходимости в среднем порядка k при $k \geq 1$. Самой слабой сходимостью обладает сходимость по распределению.

ПЗ.5. Закон больших чисел

Рассмотрим последовательность случайных величин $X_1, X_2, \dots, X_N, \dots$ с математическими ожиданиями $\mathbf{M}\{X_i\} = a_i$ и дисперсиями $\mathbf{D}\{X_i\} = D_i = \sigma_i^2$.

Образует новую величину Y_i и последовательность

$$Y_N = \frac{1}{N} \sum_{i=1}^N X_i, \quad N = 1, 2, 3, \dots; \quad (\text{ПЗ.15})$$

$$\mathbf{M}\{Y_N\} = \frac{1}{N} \sum_{i=1}^N \mathbf{M}\{X_i\}, \quad (\text{ПЗ.16})$$

полагая, что $\mathbf{M}\{Y_N\}$ принимает конечное значение при $N \rightarrow \infty$.

Под законом больших чисел (ЗБЧ) принято понимать совокупность теорем, устанавливающих необходимые (для того, чтобы..., необходимо...)

и/или достаточные (если выполняются такие-то условия, то имеет место...) условия сходимости Y_N к $\mathbf{M}\{Y_N\}$ при $N \rightarrow \infty$ по вероятности. Если вместо сходимости по вероятности рассматривается *сходимость почти наверное*, то вместо ЗБЧ имеет место *усиленный закон больших чисел* (УЗБЧ) (термин введен А.Я. Хинчиным, 1927, 1928 гг.).

Закон больших чисел устанавливает сходимость (в каком-либо смысле) случайной величины Y_N к неслучайной величине (ее математическому ожиданию – среднему) $\mathbf{M}\{Y_N\}$ при неограниченном увеличении числа слагаемых, или, иными словами, сходимость среднего арифметического случайных величин (ПЗ.15) к среднему арифметическому их математических ожиданий (ПЗ.16).

Сам закон может быть сформулирован так: средний результат большого числа случайных явлений (величин) практически перестает быть случайным и (следовательно) может быть предсказан в опытах с большой степенью определенности.

Разные формы ЗБЧ и УЗБЧ называются по именам авторов, установивших разные исторически первые, достаточные, необходимые или и те, и другие условия сходимости Y_N к $\mathbf{M}\{Y_N\}$ по вероятности тех же или разных видов случайных величин X_i , $i = \overline{1, N}$. Это формы Чебышева, Хинчина, Бернулли, Пуассона, Маркова, Колмогорова, Слуцкого, Феллера и др.

Приведем еще одну форму ЗБЧ в формулировке А.А. Маркова: достаточным условием существования ЗБЧ является сходимость к нулю дисперсии $\mathbf{D}\{Y_N\}$ при $N \rightarrow \infty$. При этом возможна любая зависимость величин X_1, \dots, X_n , при которой $\lim \mathbf{D}\{Y_N\} = 0$ при $N \rightarrow \infty$.

Аналогом ЗБЧ для случайных процессов $X(t)$ являются эргодические теоремы, устанавливающие сходимость по вероятности или по дисперсии траекторных средних к ансамблевым. Их примером являются сходимость по дисперсии и по вероятности траекторных характеристик к ансамблевым.

ПЗ.6. Предельные теоремы

Рассмотрим последовательность случайных величин X_1, \dots, X_N , а также образуемые из них последовательности Y_N (ПЗ.15) и

$$Z_N = \frac{Y_N - \mathbf{M}\{Y_N\}}{\sqrt{\mathbf{D}\{Y_N\}}} = \frac{\sum_{i=1}^N (X_i - \mathbf{M}\{X_i\})}{\sqrt{\mathbf{D}\left\{\sum_{i=1}^N X_i\right\}}} . \quad (\text{ПЗ.17})$$

Предельные теоремы устанавливают необходимые и (или – чаще!) достаточные условия сходимости по распределению Y_N или Z_N к некоторой величине Y или Z , имеющей вполне определенный (предельный) закон распределения. Первыми предельными теоремами были теоремы Я. Бернулли (1713), А. Муавра (1730), П. Лапласа (1812), С. Пуассона (1837), П.Л. Чебышева (1887), А.А. Маркова (1898), А.М. Ляпунова (1901), С.Н. Бернштейна (1926). Дальнейшие исследования в этом направлении связаны с работами А.Н. Колмогорова, Б.В. Гнеденко, П. Леви, П. Линдберга, У. Феллера и других исследователей.

Чаще всего рассматривается сходимость Y_N , Z_N к величинам, имеющим нормальное (Гаусса) распределение. Такие теоремы принято называть *центральными предельными* (ЦПТ).

Из них рассмотрим лишь *теорему Ляпунова*, устанавливающую достаточные условия применимости ЦПТ, т. е. сходимости Z_N к стандартной нормальной Z с нулевым средним и единичной дисперсией:

- а) X_1, \dots, X_N – независимые;
- б) выполняется условие

$$\lim_{N \rightarrow \infty} \frac{1}{B_N^{2+\delta}} \sum_{i=1}^N \mathbf{M} \left\{ |X_i - \mathbf{M}\{X_i\}|^{2+\delta} \right\} = 0, \quad (\text{ПЗ.18})$$

где $B_N = D_1 + \dots + D_N$, $\delta > 0$ (сам А.М. Ляпунов доказал теорему для $\delta = 1$).

Сделаем несколько замечаний по поводу предельных теорем.

1. Условия ЦПТ являются достаточными или необходимыми лишь при условии *пренебрежимо* равномерной малости вклада слагаемых.

2. ЦПТ находит широкое применение в самых разных областях благодаря ее универсальности, устойчивости и высокой точности. Первые два качества ЦПТ означают, что теорема применима к любым величинам, если их распределения имеют дисперсии и они пренебрежимо малы; незначительные отклонения от условий теоремы, в том числе наличие умеренной зависимости между X_i , не меняют нормальности предельного распределения. Третье условие означает, что нормальная (Гаусса) аппроксимация применима даже при умеренных значениях N .

3. Если не выполняются какие-либо условия ЦПТ (в том числе вместо Y_N и Z_N рассматриваются другие величины), то предельное распределение Z_N (или других величин) не обязательно будет нормальным. Оказывается, что предельные распределения величин Y_N , Z_N образуют целый класс так называемых *предельных распределений*, который совпадает с классом *устойчивых распределений* (сохраняющих свою форму при суммировании одинаковых независимых случайных величин с такими распределениями, отличающимися только параметрами положения a и масштаба λ).

ПРИЛОЖЕНИЕ 4

ПРИМЕРЫ ВЕРОЯТНОСТНЫХ ХАРАКТЕРИСТИК

Таблица П4.1

Примеры характеристик вида (4.115)–(4.117) стационарных случайных процессов $X(t)$, $Y(t)$

Обобщенная характеристика $Q(\mathfrak{Q})$	Конкретная характеристика			
	Наименование	Обозначение, определение	Вид аргумента \mathfrak{Q}	Конкретный вид обобщенного преобразования $g_Q(\cdot)$ случайного элемента \mathfrak{N} или оператора $A\{\cdot\}$
$Q_1(\mathfrak{Q}) = \mathbf{M}\{g_Q(\mathfrak{N}; \mathfrak{Q})\}$	Математическое ожидание	$m_X = \mathbf{M}\{X(t)\}$	Нет, $\mathfrak{Q} = \emptyset$	$g_Q(\mathfrak{N}; \mathfrak{Q}) = X(t)$
	Начальный момент k -го порядка	$m_{kX} = \mathbf{M}\{X^k(t)\}$	Нет, $\mathfrak{Q} = \emptyset$	$g_Q(\mathfrak{N}; \mathfrak{Q}) = X^k(t)$
	Ковариационная функция	$B_{XY}(\tau) = \mathbf{M}\{X(t)Y(t+\tau)\}$	$\mathfrak{Q} = \tau$	$g_Q(\mathfrak{N}; \mathfrak{Q}) = X(t)Y(t+\tau)$
	Функция распределения	$F_X(x) = \mathbf{P}\{X(t) \leq x\} = \mathbf{M}\{1\{X(t) \leq x\}\}$	$\mathfrak{Q} = x$	$g_Q(\mathfrak{N}; \mathfrak{Q}) = 1\{X(t) \leq x\}$
	Совместная функция распределения	$F_{XY}(x, y; \tau) = \mathbf{P}\{[X(t) \leq x] \cap [Y(t+\tau) \leq y]\} = \mathbf{M}\{1\{[X(t) \leq x] \cap [Y(t+\tau) \leq y]\}\}$	$\mathfrak{Q} = (x, y; \tau)$	$g_Q(\mathfrak{N}; \mathfrak{Q}) = 1\{[X(t) \leq x] \cap [Y(t+\tau) \leq y]\}$

Продолжение табл. 4.1

Обобщенная характеристика $Q(\mathfrak{Q})$	Конкретная характеристика		
	Наименование	Обозначение, определение	Вид аргумента \mathfrak{Q}
	Взаимная срединная корреляционная функция	$I_{XY}(\tau) = \mathbf{M} \left\{ \left[X(t) - x_{0,5} \right] \times \left[Y(t+\tau) - y_{0,5} \right] \right\}$	$\mathfrak{Q} = \tau$
	Взаимная полукорреляционная функция	$H_{XY} = \mathbf{M} \left\{ \left[X(t) - m_X \right] \times \left[F_Y [Y(t+\tau)] - \mathbf{M} \{ F_Y [Y(t)] \} \right] \right\}$	$\mathfrak{Q} = \tau$
$Q_{II}(\mathfrak{Q}) =$ $\mathbf{M} \{ g_Q(\mathfrak{N};$ $Q_1, \dots, Q_s; \mathfrak{Q}) \}$	Взаимная дисперсионная функция II рода	$\zeta_{Y, X}(\tau) = \mathbf{M} \left\{ \left[Y(t+\tau) - m_Y(t+\tau; X) \right]^2 \left[X(t) - x \right] \sigma_Y^2 \right\}$	$\mathfrak{Q} = \tau$
	Центральный момент порядка $k_1 + k_2 + \dots + k_n$	$\mu_{k_1, \dots, k_n}(\tau_1, \dots, \tau_{n-1}) =$ $= \mathbf{M} \left\{ \left[X_1(t) - m_{X_1} \right]^{k_1} \times \left[X_2(t+\tau_1) - m_{X_2} \right]^{k_2} \times \dots \times \left[X_n(t+\tau_{n-1}) - m_{X_n} \right]^{k_n} \right\}$	$\mathfrak{Q} = (\tau_1,$ $\tau_2, \dots, \tau_{n-1})$
			<p>Конкретный вид обобщенного преобразования $g_Q(\cdot)$ случайного элемента \mathfrak{N} или оператора $A\{\}$</p> $g_Q(\mathfrak{N}; Q_1, \dots, Q_s; \mathfrak{Q}) = \left[X(t) - x_{0,5} \right] \times \left[Y(t+\tau) - y_{0,5} \right]; Q_1 = x_{0,5}; Q_2 = y_{0,5}; s=2$ $g_Q(\mathfrak{N}; Q_1, \dots, Q_s; \mathfrak{Q}) = \left[X(t) - m_X \right] \times \left[F_Y [Y(t+\tau)] - \mathbf{M} \{ F_Y [Y(t)] \} \right]; Q_1 = m_X, Q_2 = F_Y(y); Q_3 = \mathbf{M} \{ F_Y [Y(t)] \}; s=3$ $g_Q(\mathfrak{N}; Q_1, \dots, Q_s; \mathfrak{Q}) = \left[Y(t+\tau) - m_Y(t+\tau; X(t)) \right]^2 / \sigma_Y^2;$ $Q_1 = \mathbf{M} \left[Y(t+\tau) X(t) = x \right] = m_Y(t+\tau; x); Q_2 = \sigma_Y^2; s=2$ $g_Q(\mathfrak{N}; Q_1, \dots, Q_s; \mathfrak{Q}) =$ $= \prod_{i=1}^n \left[X_i(t+\tau_{i-1}) - m_{X_i} \right]^{-k_i},$ $\tau_0 = 0; Q_1 = m_{X_1}, \dots; Q_n = m_{X_n}; s = n$

					$A\{Q_1, \dots, Q_s; \mathfrak{P}\} = \sqrt{Q_s} =$ $= \sqrt{D_X}, Q_1 = D_X, s=1$
Среднее квадратическое отклонение				Нет, $\mathfrak{P} = 0$	
Взаимная нормированная корреляционная функция				$\mathfrak{P} = \tau$	$A\{Q_1, \dots, Q_s; \mathfrak{P}\} = Q_1(\mathfrak{P}) / \sqrt{Q_2 Q_3},$ $Q_1(\mathfrak{P}) = R_{XY}(\tau); Q_2 = R_{XX}(0) = D_X = \sigma_X^2;$ $Q_3 = R_{YY}(0) = D_Y = \sigma_Y^2, s=3$
Плотность распределения вероятностей				$\mathfrak{P} = x$	$A\{Q_1, \dots, Q_s; \mathfrak{P}\} = dQ_1(\mathfrak{P})/d\mathfrak{P},$ $Q_1(\mathfrak{P}) = F_X(x); s=1$
Спектральная плотность мощности				$\mathfrak{P} = \omega$	$A\{Q_1, \dots, Q_s; \mathfrak{P}\} =$ $= \int_{-\infty}^{\infty} Q_1(\tau) \exp\{-j\omega\tau\} d\tau,$ $Q_1 = R_{XY}(\tau); s=1$
Кумулянтная функция				$\mathfrak{P} =$ $= (u, v; \tau)$	$A\{Q_1, \dots, Q_s; \mathfrak{P}\} = \ln Q_1(\mathfrak{P}),$ $Q_1(\mathfrak{P}) = \Theta_{XY}(u, v; \tau); s=1$
Мода абсолютно непрерывного случайного процесса $X(t)$				Нет, $\mathfrak{P} = 0$	$A\{Q_1, \dots, Q_s; \mathfrak{P}\} = \arg \max_{\mathfrak{P}} Q_1(\mathfrak{P}),$ <p>т. е. неявный оператор, $Q_1(\mathfrak{P}) = W_X(x; \tau); s=1$</p>
Квантиль порядка p				Нет, $\mathfrak{P} = 0$	<p>Неявный оператор нахождения корня уравнения (1), т. е.</p> $A\{Q_1, \dots, Q_s; \mathfrak{P}\} = \arg [Q_1(\mathfrak{P}) = p],$ $Q_1(\mathfrak{P}) = F_X(x); s=1$

$$Q_{III}(\mathfrak{P}) =$$

$$= A\{Q_1, \dots, Q_s; \mathfrak{P}\}$$

**ЧАСТО ИСПОЛЬЗУЕМЫЕ ОБОЗНАЧЕНИЯ
И СОКРАЩЕНИЯ**

П5.1. Принятые обозначения

D – область практической деятельности (§ 2.3.1).

M – методы (§ 1.3).

Π_{jk} – плата за решение γ_k , когда объект (система) находится в состоянии S_j (§ 4.3.7).

t – знак транспонирования векторов и матриц.

A – случайное событие.

$A = [a_1, a_2]$ – интервал, интервальное число A (§ 4.4.4).

$a(t), \lambda(t)$ – аддитивная и мультипликативная трендовые составляющие моделей нестационарного случайного процесса (§ 4.3.6.1).

a_k, b_k, A_k – коэффициенты разложения $x(t)$ в ряд Фурье (§ 3.6).

(a, λ, α) – параметры: сдвига (a), масштаба (λ), формы (α) (§ 4.2, § 4.3).

$C(\lambda)$ – конспектральная плотность мощности.

$\text{cor}\{X, Y\} = R_{XY}$ – корреляционный момент (корреляция).

D – допустимое множество решений (§ 4.5).

d – фрактальная размерность (§ 4.4.6).

d_τ – топологическая размерность (§ 4.4.6).

$\mathbf{D}\{\cdot\}$ – оператор нахождения дисперсии (§ 4.3.3).

D_X – дисперсия случайной величины X .

E – энергия сигнала (§ 3.6).

E – шумовая составляющая (§ 3.7, § 4.3).

E_X – срединное отклонение (§ 4.3.3).

$E_{W_f}(\lambda)$ – скалограмма (§ 4.3.6.4, см. (4.110)).

F – статистика Фишера (§ 4.3.4).

$F(x) = F_X(x)$ – функция распределения вероятностей случайной величины X .

$F(x)$ – свертка соизмеримых критериев (§ 4.5.5).

$F_f(\omega, \tau)$ – динамическое представление Фурье (§ 4.3.6.4, см. (4.106)).

$nF_{\mathbf{X}}(\mathbf{x})$ – n -мерная функция распределения (§ 4.3).

H_i – i -я гипотеза (§ 4.3.7).

$i = \overline{1, N}$ – последовательность чисел $1, 2, \dots, N$.

$I_A(x)$ – индикатор принадлежности x множеству A .

$I(A)$ – индикаторная функция.

$j = \sqrt{-1}$ – мнимая единица в преобразованиях типа Фурье.

J, \mathbf{J} – функционал (§ 4.5.8).

k_i – коэффициент компетентности (§ 4.4.7).

$I_N(\chi) = I_N(x_1, \dots, x_N)$ – отношение правдоподобия (§ 4.3.7).

$L_N(\chi | \theta)$ – функция правдоподобия (§ 4.3.7).

M – модуль в методе вычетов (§ 4.3.8).

$m_Y(x)$ – условное математическое ожидание, функция регрессии случайной величины Y по значениям (вдоль значений) x случайной величины X (§ 4.3).

$\tilde{m}_Y(x)$ – среднеквадратическая прямая регрессии.

N – число опытов, объем выборки.

$P(A)$ – вероятность события A (табл. 3.4, § 4.3).

$p(A)$ – конкретное значение вероятности $P(A)$ события A .

$P^*(A), p^*(A)$ – частость (частота) появления события A в опытах, оценка вероятности $P(A), p(A)$ (табл. 3.4, § 4.3).

$P\{A_1 | A_2\}$ – (условная) вероятность появления события A_1 при условии, что ему предшествовало появление события A_2 .

$P_F(T, \tau)$ – спектрограмма (§ 4.3.6.4, см. (4.108)).

$P_{\text{ср}}$ – средняя мощность сигнала (§ 3.6).

p_I, p_{II} – вероятности ошибок I и II рода (§ 4.3.7).

$Q(\lambda, t, \tau)$ – ансамблевые характеристики случайных функций (§ 4.3, Приложение 4).

$\hat{Q}, \hat{\theta}$ – оценки Q, θ .

$q(\mathcal{S}; \theta)$ – аналитическая модель характеристики $Q(\mathcal{S}; \theta)$ (§ 4.3.6).

$q(\lambda; \alpha)$ – характеристика сигнала $\xi(t)$ (§ 3.6).

$\mathcal{Q}(\mathcal{S}), Q(\mathcal{S})$ – эмпирические характеристики (§ 3.6, § 4.3).

$Q(\lambda; \alpha)$ – характеристика детерминированной функции $x(t)$ как модели сигнала $\xi(t)$.

Q_p – квантильные характеристики (табл. 4.4).

R, C, K, Z, S, Q – характеристики случайных элементов (теоретические, генеральных совокупностей; в тех случаях, когда можно спутать с \mathcal{R}, \mathcal{Q} и т. д.).

R_{XY} – информационный коэффициент связи (§ 4.3.5).

$R(\tau)$ – корреляционная функция.

$R(\lambda), I(\lambda)$ – действительная и мнимая составляющие спектральной плотности $X(\lambda)$ (§ 3.6).

\mathbf{R} – матрица дублирования спектра плана (§ 3.7).

R – Байесовский средний риск – среднее значение потерь (§ 4.3.7).

$r_{i(1, \dots, n)}$ – совокупный коэффициент корреляции (§ 4.3.5).

$r_{xx}(\tau), r_{xy}(\tau)$ – нормированные корреляционные функции $x(t), y(t)$ (§ 3.6).

$\mathcal{R}, \mathcal{Q}, K, Z$ – выборочные (статистические, эмпирические) характеристики – аналоги теоретических R, Q, K, Z .

s – число интервалов группирования (§ 3.3, § 4.3.3).

s_i – область выборочного пространства, соответствующая S_i -му состоянию объекта (§ 4.3.7).

s_x – среднеквадратическое значение сигнала $x(t)$ (§ 3.6).

$s(t), c(t)$ – сезонная и циклическая аддитивные составляющие модели нестационарного случайного процесса (§ 4.3.6.1).

S – экспериментальное отклонение, эмпирическое среднеквадратическое отклонение (табл. 3.4, § 4.3.3).

S^2 – выборочная дисперсия (§ 4.3.3, 4.3.4).

$S_A(p)$ – смешанная стратегия (§ 4.5.9).

$S(\lambda)$ – спектральная плотность мощности.

S, s, s_x – эмпирическое среднеквадратическое отклонение (СКО); среднеквадратическое отклонение сигнала, выборочное СКО случайной величины или функции (§ 3.6, § 4.3.6.3).

S_i – i -е состояние объекта (системы) (§ 4.3.7).

T – длина временного интервала $(0, T)$.

T_0 – период (§ 3.6).

U – универсальное множество (§ 4.4.3).

$(U, \mu_A(u))$ – нечеткое множество A с функцией принадлежности $\mu_A(u)$, $u \in U$ (§ 4.4.3).

U – условия (§ 2.1.2, 4.3.1).

W – коэффициент конкордации (§ 4.4.7).

$W(x) = W_X(x)$ – плотность распределения вероятностей случайной величины X .

$W_f(\lambda; \tau)$ – вейвлетное представление (§ 4.3.6.4).

x – символ (шкальное значение) измерительной шкалы (§ 3.1.3).

x – возможные значения экспериментальных данных, результатов измерения физической величины (рис. 3.4).

$x, y, z, x(t), y(t), z(t)$: 1) переменные детерминированных функций $y = f(x, z)$; 2) возможные значения случайных величин и функций X, Y, Z ; 3) возможные значения элементов выборок, исходных данных; 4) эмпирические значения X, Y, Z .

$x_1, x_2, \dots, x_i, \dots$ – возможные значения величины X , полученные в 1, 2, ..., i -м, ... опыте.

$x(t), y(t), z(t)$: 1) детерминированные модели сигналов $\xi(t)$ (§ 3.6); 2) траектории (реализации, выборочные функции) случайных функций $X(t), Y(t), Z(t)$.

$x(t)$ – детерминированная функция (§ 3.6, § 4.3).

$\overline{x(t)}$ – среднее (среднее арифметическое) $x(t)$.

\bar{x} – эмпирическое среднее, среднее арифметическое (табл. 3.4, § 4.3).

x_0^* – идеальная точка (§ 4.5.5).

x^* – оптимальное решение (§ 4.5).

$x_{(i)}$ – i -й элемент вариационного ряда (§ 3.3).

$x'_{[k]}$ – левое граничное значение k -го интервала группирования (§ 3.3).

$\dot{x}'_{[k_i]}$ – среднее значение k -го интервала группирования (§3.3).

$x'_{[l]}$ – левое граничное значение l -го интервала группирования (§ 4.3.4).

$X, Y, Z, X(t), Y(t), Z(t)$ – случайные величины, функции.

\tilde{X} – случайная величина со стандартным ($a = 0, \lambda = 1$) распределением вероятности (§ 4.3.8).

x, y, z – конкретные значения элементов выборок, эмпирических данных, случайных величин X, Y, Z .

x_{\min}, x_{\max} – минимальное и максимальное значения из множества значений x (§ 3.3).

x_p – квантиль порядка p (табл. 4.4).

X, Y, Z, W, E – факторы, переменные, помехи (§ 3.7).

$X(t)$ – случайная функция аргумента t (§ 4.3).

$X(\lambda) = \mathcal{F} \left\{ \left[x(t)e^{-j\lambda t} \right]_t \right\}$ – спектральное (частотное) представление $x(t)$

по Фурье (§ 3.6).

$X_d(\omega)$ – спектр дискретизованной функции (§ 3.6).

$\mathbf{X}, \mathbf{Y}, \mathbf{R}$ – обозначения векторов, матриц.

$\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_n(t))$ – векторная случайная функция (§ 4.3).

Γ – обобщенное представление длины непрерывного временного интервала или числа отсчетов (длины) временного ряда.

$\gamma_{XY}^2(\omega)$ – функция когерентности (§ 4.3.6.3).

Δ^2 – средний квадрат погрешности (§ 4.3.5).

Δt – шаг дискретизации временного аргумента t .

Δx – шаг, ширина интервала группирования (§ 3.3), шаг измерения, оценивания (§ 4.3.4), квантования по уровню.

$\Delta_{\hat{Q}}^2(\mathbf{\Theta})$ – средний квадрат отклонения оценки $\hat{Q}(\mathbf{\Theta})$.

$\Delta\omega_3$ – эффективная ширина спектра сигнала, функции (§ 3.6, (3.151)).

\mathbf{E}, ε – погрешность (§ 4.3).

$\varepsilon_{\hat{Q}(\mathbf{\Theta})}$ – смещение оценки $\hat{Q}(\mathbf{\Theta})$.

θ – параметр.

$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)$ – параметры моделей, характеристик (§ 3.7, § 4.3).

$\eta_{Y|X}, \eta_{X|Y}$ – корреляционные отношения (§ 4.3).

$I(A)$ – индикатор события A (§ 4.3.3).

$\mathbf{I}\{A\}$ – индикаторный оператор (§ 4.3.3).

K_{XY} – конкорреляционный момент (§ 4.3).

$K(\tau)$ – конкорреляционная функция.

λ – обобщенное обозначение частоты как аргумента функции (§ 3.6, § 4.3).

λ – аргумент характеристики $Q(\lambda)$, $Q(\lambda)$.

$\Lambda = (-\infty, \infty)$ – диапазон частот для функций непрерывного временного аргумента $t = t$ (§ 3.6, § 4.3).

$\Lambda = (-\pi, \pi)$ – диапазон частот для дискретных t (§ 3.6, § 4.3).

$\mathbf{M}\{\cdot\}$ – оператор усреднения по вероятностной мере, математического ожидания (§ 4.3.3, Приложение 1).

$\mathbf{M}(p, q)$ – средний выигрыш (§ 4.5.9).

$\mu_A(u)$ – функция принадлежности элемента $u \in U$ множеству A (§ 4.4.3).

$\mu_T\{\cdot\}$, $\mu_N\{\cdot\}$, $\mu_T\{\cdot\}$ – операторы усреднения по выборке (см. Приложение 1).

$\nu_{Y|X}, \nu_{X|Y}$ – коэффициенты нелинейности функций регрессии (§ 4.3).

Ξ – случайная величина с абсолютно непрерывным равномерным на интервале $(0,1)$ распределением (§ 4.3.8).

$\Xi(i\Delta t)$ – δ -коррелированная (некоррелированная) случайная последовательность (§ 4.3.8).

ξ_i – элементы выборки из ГС с равномерным на $(0,1)$ распределением, элементы базовой равномерно распределенной ППСЧ (§ 4.3.8).

$\xi(t), \psi(t), \zeta(t)$ – сигналы (§ 3.6).

$\mathbf{П}(n \times m)$ – платежная матрица (§ 4.5.9).

ρ_{ij} – парные коэффициенты корреляции (§ 4.3.6).

$\rho_{ij(1, \dots, n)}$ – частный коэффициент корреляции (§ 4.3.5).

ρ_{XY} – коэффициент корреляции случайных величин X и Y (§ 4.5).

$\mathbf{P}\{A\}$ – вероятностная мера случайного события A (табл. 3.4, § 4.3).

\mathbf{p}, \mathbf{R} – корреляционные матрицы (§ 4.3).

σ_X – среднеквадратическое отклонение случайной величины X (§ 4.3.3).

$\sigma_{\hat{Q}}(\mathbf{\Theta}), \sigma_{\hat{Q}}^2(\mathbf{\Theta}) = D_{\hat{Q}}(\mathbf{\Theta})$ – среднеквадратическое отклонение (σ) и дисперсия (σ^2, D) оценки $\hat{Q}(\mathbf{\Theta})$.

$\sigma_Y^2(x)$ – условная дисперсия (§ 4.3.5).

τ_0 – радиус корреляции (§ 3.6).

$\tau_{\text{эф}}$ – эффективная длительность сигнала, функции (§ 3.6, (3.17)).

$\tau_{\text{е}}$ – интервал корреляции (§ 3.6).

$\varphi(\chi)$ – критериальная статистика при проверке гипотез (§ 4.3.7).

χ – набор (множество) исходных данных (§ 2.1.2, рис. 3.4 и далее).

χ^2 – статистика хи-квадрат (§ 4.3).

χ_{XY} – коэффициент конкорреляции (§ 4.3).

$\psi(t)$ – вейвлет (§ 4.3.6.4).

$\omega_{\text{эф}}$ – эффективная полоса, интервал спектра сигнала, функции (§ 3.6, (3.18)).

$\mathfrak{F}\{f(x)|_x\}$ – обобщенный оператор интегрирования и суммирования (§ 4.3.7, (4.132)).

$\mathfrak{F}\{\}$ – оператор преобразования Фурье (§ 4.3.6.3, (3.10), (4.99), (4.100)).

\mathcal{M} – модель (§ 4.3.1).

$\mathcal{R}(\tau)$ – кепстр (§ 4.3.6.3, см. (4.104)).

\mathfrak{x} – значения (реализации) случайного элемента \aleph (§ 4.3.1).

\aleph – случайный элемент (§ 4.3.1).

\otimes – символ интеграла свертки.

П5.2. Принятые сокращения

АЦП – аналого-цифровой преобразователь (§ 3.6).

БХ-оценки – оценки, полученные по методу базовых характеристик (§ 4.3.7).

ВТ – вычислительная техника.

ВХ – вероятностная характеристика (§ 4.3).

ГОС – государственный образовательный стандарт (введение).

ГС – генеральная совокупность (§ 4.3.7).

ГП – геометрическое программирование (§ 4.5.2).

Данные – или сигналы, или данные, или знания-1 (анзния), или контент, или любая их смесь.

ДЛП – дробно-линейное программирование (§ 4.5).

ДОФ-оценка – непараметрическая оценка, полученная по методу дельтаобразных функций (§ 4.3.7).

ДП – динамическое программирование (§ 4.5.2).

ДПМ – дисперсионная плотность мощности (§ 4.3.6).

ДФ – дисперсионная функция (§ 4.3.6).

ЗХ-оценка – оценка параметров, полученная по методу значений характеристик (§ 4.3.7).

КА – корреляционный анализ (§ 3.6, § 4.3).

КК – коэффициент корреляции (§ 4.3).

ККК – коэффициент конкорреляции (§ 4.3).

ККФ – конкорреляционная функция (§ 4.3.6).

КМ – корреляционная матрица (§ 4.3).

- КО – корреляционное отношение (§ 4.3).
- КП – квадратичное программирование (§ 4.5.2).
- КПМ – конкорреляционная плотность мощности (§ 4.3.6).
- КСХ – корреляционно-спектральные характеристики (§ 4.3).
- КХ-оценка – оценка, полученная по методу квантильных характеристик (§ 4.3.7).
- ЛП – линейное программирование (§ 4.5).
- ЛР – линия регрессии (§ 4.3).
- ЛР-функция – случайная функция с линейной регрессией (§ 4.3.6).
- ММБ-оценка – оценка, полученная по методу минимума меры близости (§ 4.3.7).
- МП – математическое программирование (§ 4.5).
- МП-оценка – оценка, полученная по методу максимального правдоподобия (§ 4.3.7).
- МХК-оценка – оценка, полученная по методу минимума хи-квадрат (§ 4.3).
- Н-оценка – непараметрическая оценка (§ 4.3.7).
- НК-оценка – оценка, полученная по методу наименьших квадратов (§ 4.3).
- НККФ – нормированная конкорреляционная функция (§ 4.3.6).
- НКФ – нормированная корреляционная функция (§ 3.6, § 4.3.6).
- НМ-оценка – оценка, полученная по методу наименьших модулей (§ 4.3.7).
- НР-оценка – оценка, полученная по методу неортогональных разложений (§ 4.3.7).
- НП – нелинейное программирование (§ 4.5).
- ОМХК-оценка – оценка, полученная по обобщенному методу минимума хи-квадрат (§ 4.3.7).
- ОНК-оценка – оценка, полученная по обобщенному методу наименьших квадратов (§ 4.3.7).
- ОНМ-оценка – оценка, полученная по обобщенному методу наименьших модулей (§ 4.3.7).
- ОР-оценка – оценка, полученная по методу ортогональных разложений (§ 4.3.7).
- ПККФ – полуконкорреляционная функция (§ 4.3.6).
- П-оценка – параметрическая оценка (§ 4.3.7).
- ППСЧ – последовательность псевдослучайных чисел (§ 4.3.8).
- ПРВ – плотность распределения вероятностей (§ 4.3).
- ПСПЧ – псевдослучайная последовательность чисел (§ 4.3.8).
- ПСЧ – псевдослучайные числа (§ 4.3.8).
- РА – регрессионный анализ (§ 4.3).
- СКО – среднее квадратическое (среднеквадратическое) отклонение (§ 3.6, § 4.3).

СМО – система массового обслуживания (§ 4.6).

СП – стохастическое программирование (§ 4.5.3).

СПМ – спектральная плотность мощности (§ 4.3.6).

СПБП – сепарабельное программирование (§ 4.5.2).

СФ – случайная функция (§ 4.3.6).

ТМО – теория массового обслуживания (§ 4.6).

УГС – укрупненная группа специальностей (введение).

ФРВ – функция распределения вероятностей (§ 4.3).

ЦП – целочисленное программирование (§ 4.5.2).

ЦФ – целевая функция (§ 4.5).

ЧАКФ – частная автокорреляционная функция (§ 4.3.6).

ЧКК – частный коэффициент корреляции (§ 4.3.5).

ЧКФ – частная корреляционная функция (§ 4.3.6).

ЧХ-оценка – оценка параметров, полученная по методу числовых характеристик (§ 4.3.7).

ЭХ-оценка – непараметрическая оценка, полученная по методу эмпирических характеристик (§ 4.3.7).

Я-оценка – непараметрическая оценка, полученная по ядерному методу (§ 4.3.7).

ОГЛАВЛЕНИЕ

Введение	3
Глава первая. Исходные определения и понятия	5
§ 1.1. Что такое модель? Термины как модели определяемых ими понятий	5
§ 1.2. Что такое наука?	14
§ 1.3. Информация («рабочее» понятие) и ее носители	19
§ 1.4. Информатика и ее структуры	28
§ 1.5. Структура теоретической (формальной) информатики	32
Заключение	34
Вопросы для самоподготовки	35
Глава вторая. Моделирование как метод исследования объектов	37
§ 2.1. Моделирование объектов	37
2.1.1. Определение моделирования	37
2.1.2. Требования к модели	39
2.1.3. Виды моделей	46
§ 2.2. Необходимые понятия семиотики	51
§ 2.3. Понятие о системологии	52
2.3.1. Системология как наука и область практической деятельности. Системные дисциплины. Свойства систем	52
2.3.2. Системные принципы, законы и закономерности	57
§ 2.5. Виды объектов с точки зрения их модельного представления	61
§ 2.6. Системный подход к исследованию объектов	64
Заключение	66
Вопросы для самоподготовки	68
Глава третья. Экспериментирование как метод исследования и этап моделирования объектов	70
§ 3.1. Физические и математические величины	70
3.1.1. Необходимые определения	70

3.1.2. Виды физических величин.....	71
3.1.3. Измерительные шкалы	74
§ 3.2. Наблюдение и эксперимент.....	82
§ 3.3. Разновидности экспериментов	84
§ 3.4. Понятие о неполноте априорных и апостериорных (экспериментальных) Данных о моделируемом объекте	95
§ 3.5. Измерения как этап наблюдения, натурального экспериментирования и моделирования.....	97
§ 3.6. Временное и частотное представления сигналов и временных рядов.....	109
§ 3.7. Понятие о планировании наблюдений и экспериментов.....	125
§ 3.8. Технологические аспекты экспериментирования	137
§ 3.9. Правила экспериментирования	141
Заключение.....	146
Вопросы для самоподготовки.....	146
Глава четвертая. Формальный аппарат описания и исследования структур, состояний и поведения объектов: куюмодные аналитические модели.....	149
§ 4.1. Вводные замечания	149
§ 4.2. Детерминированный аппарат. Законовые модели	150
§ 4.3. Стохастический (вероятностно-статистический) аппарат. Модели объективно обнаруживаемых закономерностей	154
4.3.1. Условия применимости. Постановки задач.....	154
4.3.2. Виды стохастических экспериментальных Данных	161
4.3.3. Дескриптивный анализ.....	162
4.3.4. Анализ и идентификация распределения значений.....	172
4.3.5. Анализ взаимосвязей случайных величин и векторов.....	175
4.3.6. Представление и анализ результатов стохастических экспериментов, описываемых случайными функциями (процессами, последовательностями).....	213
4.3.6.1. Модели случайных функций.....	213
4.3.6.2. Характеристики связи: корреляционный и регрессионный анализ случайных функций	216
4.3.6.3. Характеристики связи: спектральный анализ случайных функций	218
4.3.6.4. Характеристики связи: вейвлет-анализ и динамический спектральный анализ	223
4.3.7. Апостериорный статистический аппарат (аппарат математической статистики)	232
4.3.8. Аппарат статистической имитации случайных элементов. Метод Монте-Карло	271

§ 4.4. Другие разновидности формального аппарата описания и исследования объектов.....	291
4.4.1. Вводные замечания.....	291
4.4.2. Байесовский стохастический аппарат.....	292
4.4.3. Аппарат нечетких (размытых) множеств.....	297
4.4.4. Интервальный аппарат.....	304
4.4.5. Аппарат теории катастроф, динамического хаоса и самоорган- низации.....	307
4.4.6. Фракталы.....	316
4.4.7. Аппарат экспертного описания и исследования объектов.....	322
§ 4.5. Аппарат поиска оптимальных результатов.....	328
4.5.1. Вводные замечания.....	328
4.5.2. Детерминированные оптимизационные модели.....	329
4.5.3. Стохастические оптимизационные методы.....	334
4.5.4. Нечеткие и интервальные оптимизационные модели.....	336
4.5.5. Многокритериальная оптимизация.....	337
4.5.6. О корректности постановок оптимизационных задач.....	341
4.5.7. Практические примеры оптимизационных задач.....	342
4.5.8. Вариационные модели.....	343
4.5.9. Состязательные модели. Теория игр. Теория принятия оптимальных решений в условиях неопределенности, конфликтов.....	352
§ 4.6. Аппарат теории массового обслуживания (очередей).....	362
Заключение.....	372
Вопросы для самоподготовки.....	373
Библиографический список.....	377
Приложения.....	379
Приложение 1. Операторы усреднения.....	379
Приложение 2. Свойства вероятностных характеристик.....	381
Приложение 3. Некоторые неравенства и предельные теоремы теории вероятностей.....	398
Приложение 4. Примеры вероятностных характеристик.....	404
Приложение 5. Часто используемые обозначения и сокращения.....	408

Губарев Василий Васильевич

ВВЕДЕНИЕ В ТЕОРЕТИЧЕСКУЮ ИНФОРМАТИКУ

Часть 1

Учебное пособие

Редактор *Н.А. Лукашова*
Выпускающий редактор *И.П. Брованова*
Корректор *И.Е. Семенова*
Дизайн обложки *А.В. Ладьяжская*
Компьютерная верстка *Н.В. Гаврилова*

Подписано в печать 24.06.2014. Формат 60 × 84 1/16. Бумага офсетная
Тираж 100 экз. Уч.-изд. л. 24,41. Печ. л. 26,25. Изд. 288/13. Заказ №
Цена договорная

Отпечатано в типографии
Новосибирского государственного технического университета
630073, г. Новосибирск, пр. К. Маркса, 20