

Министерство сельского хозяйства Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего образования
ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ
АГРАРНЫЙ УНИВЕРСИТЕТ ИМЕНИ А.А. ЕЖЕВСКОГО

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
И АЛГОРИТМИЗАЦИЯ ЗАДАЧ ТЕПЛОЭНЕРГЕТИКИ**

*Учебное пособие по дисциплине «Математическое моделирование
и алгоритмизация задач теплоэнергетики»
для студентов очной и заочной форм обучения направлений подготовки
13.04.01 Теплоэнергетика и теплотехника (уровень магистратуры),
13.04.02 Электроэнергетика и электротехника (уровень магистратуры)
(методические указания и задания для выполнения контрольных работ)*

Молодежный 2021

УДК 519.85/.86:621.311.22 (075.8)

М 34

Печатается по решению методического совета энергетического факультета Иркутского ГАУ (протокол № 5 от 19 января 2021 г.).

Составители: Кудряшев Г.С., Третьяков А.Н.

Рецензент: доцент кафедры электрооборудования и физики Иркутского государственного аграрного университета им. А.А. Ежевского, кандидат технических наук, доцент Логинов А.Ю.

Математическое моделирование и алгоритмизация задач теплоэнергетики : учебное пособие по дисциплине «Математическое моделирование и алгоритмизация задач теплоэнергетики» для студентов очной и заочной форм обучения направлений подготовки 13.04.01 Теплоэнергетика и теплотехника (уровень магистратуры), 13.04.02 Электроэнергетика и электротехника (уровень магистратуры) : (методические указания и задания для контрольных работ) / Иркут. гос. аграр. ун-т им. А. А. Ежевского ; сост.: Г. С. Кудряшев, А. Н. Третьяков. – Молодежный : ИрГАУ, 2021. – 306 с. : ил. – Текст : электронный

Учебное пособие предназначено для изучения курса «Математическое моделирование и алгоритмизация задач теплоэнергетики». Основной целью учебного пособия является оказание помощи студентам при выполнении контрольной работы.

© Кудряшев Г.С., Третьяков А.Н., 2021
© Иркутский ГАУ им. А. А. Ежевского, 2021

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
1. ПОНЯТИЕ МОДЕЛИРОВАНИЯ. СПОСОБЫ ПРЕДСТАВЛЕНИЯ МОДЕЛЕЙ	7
2. ЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ	33
3. НЕЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ	39
4. ДИНАМИЧЕСКИЕ СИСТЕМЫ	41
5. ДИНАМИЧЕСКИЕ РЕГРЕССИОННЫЕ МОДЕЛИ, ЗАДАННЫЕ В ВИДЕ ПЕРЕДАТОЧНОЙ ФУНКЦИИ	50
6. МОДЕЛЬ В ВИДЕ ФИЛЬТРА КАЛЛМАНА	54
7. МОДЕЛЬ ДИНАМИЧЕСКОЙ СИСТЕМЫ В ВИДЕ ФУРЬЕ ПРЕДСТАВЛЕНИЯ (МОДЕЛЬ СИГНАЛА)	56
8. МОДЕЛЬ ДИНАМИЧЕСКОЙ СИСТЕМЫ В ВИДЕ ФУРЬЕ ПРЕДСТАВЛЕНИЯ (МОДЕЛЬ ОБЪЕКТА)	61
9. ОЦЕНКА КАЧЕСТВА МОДЕЛИ	65
10. ЧИСЛЕННЫЕ МЕТОДЫ ИНТЕГРИРОВАНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ. МЕТОД ЭЙЛЕРА	68
11. ПОСТРОЕНИЕ МОДЕЛИ ДИНАМИЧЕСКОЙ СИСТЕМЫ В ВИДЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ И РАСЧЕТ	77
12. УРАВНЕНИЯ ВЫСШИХ ПОРЯДКОВ	91
13. СТРУКТУРНО ПЕРЕСТРАИВАЕМЫЕ МОДЕЛИ	92
14. УТОЧНЕННЫЙ МЕТОД ЭЙЛЕРА	95
15. МЕТОД РУНГЕ-КУТТЫ	98
16. МЕТОДЫ ПРОГНОЗА И КОРРЕКЦИИ (ИТЕРАЦИОННЫЕ МЕТОДЫ)	10
17. МОДЕЛИРОВАНИЕ СИСТЕМ С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ	10
18. МОДЕЛИРОВАНИЕ СИСТЕМ С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ ПРИ ПЕРЕМЕЩЕНИИ МАТЕРИАЛЬНЫХ	10
19. УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ	11
20. ТЕХНОЛОГИЯ ИСПОЛЬЗОВАНИЯ ДИНАМИЧЕСКИХ МОДЕЛЕЙ	12
21. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ	14
22. ГЕНЕРАТОРЫ СЛУЧАЙНЫХ ЧИСЕЛ	15
23. МОДЕЛИРОВАНИЕ СЛУЧАЙНОГО СОБЫТИЯ. МОДЕЛИРОВАНИЕ ПОЛНОЙ ГРУППЫ НЕСОВМЕСТНЫХ СОБЫТИЙ	17
24. МОДЕЛИРОВАНИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ С ЗАДАНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ	17
25. МОДЕЛИРОВАНИЕ НОРМАЛЬНО РАСПРЕДЕЛЕННЫХ СЛУЧАЙНЫХ ВЕЛИЧИН	18
26. МОДЕЛИРОВАНИЕ СИСТЕМЫ СЛУЧАЙНЫХ ВЕЛИЧИН	19
27. РАСПРЕДЕЛЕНИЕ ПУАССОНА	19

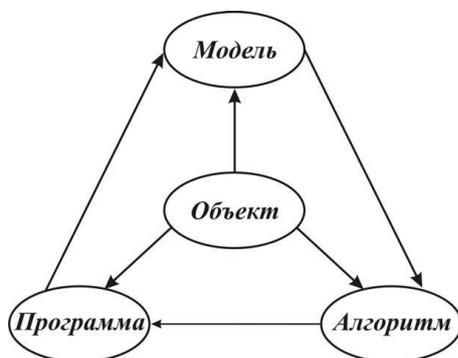
28. ПОТОК СЛУЧАЙНЫХ СОБЫТИЙ.....	20
29. ПОТОКИ С ПОСЛЕДЕЙСТВИЕМ (ПОТОКИ ЭРЛАНГА).....	20
30. МОДЕЛИРОВАНИЕ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ...	21
31. МОДЕЛИРОВАНИЕ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ И СИСТЕМ.....	23
32. ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ МОДЕЛИРУЮЩИХ АЛГОРИТМОВ.....	24
33. МОДЕЛИРОВАНИЕ МАРКОВСКИХ СЛУЧАЙНЫХ ПРОЦЕССОВ	25
34. ФИКСАЦИЯ И ОБРАБОТКА СТАТИСТИЧЕСКИХ РЕЗУЛЬТАТОВ	26
35. НЕФОРМАЛЬНЫЙ СИНТЕЗ.....	27
36. ЭКСПЕРТИЗА.....	28
Задания для студентов	29
Варианты задач для выполнения контрольной работы.....	30
ЛИТЕРАТУРА	30

ВВЕДЕНИЕ

Невозможно представить себе науку без широкого применения математического моделирования. Сущность этой методологии состоит в замене исходного объекта его математической моделью – и дальнейшем изучении модели с помощью реализуемых на компьютерах вычислительно-логических алгоритмов. Работа не с самим объектом (явлением, процессом), а с его моделью дает возможность безболезненно, относительно быстро и без существенных затрат исследовать его свойства и поведение в любых мыслимых ситуациях (преимущества теории). Вычислительные эксперименты с моделями объектов позволяют, опираясь на современные вычислительные методы и технические инструменты информатики, подробно и глубоко изучать объекты в достаточной полноте, недоступной чисто теоретическим подходам (преимущества эксперимента). Методология математического моделирования бурно развивается, охватывая все новые сферы – от разработки технических систем и управления ими до анализа сложнейших экономических и социальных процессов.

Элементы математического моделирования использовались с самого начала появления точных наук, и не случайно, что некоторые методы вычислений носят имена такие как Ньютон и Эйлер, а слово «алгоритм» происходит от имени средневекового арабского ученого Аль-Хорезми.

Постановка вопроса о математическом моделировании какого-либо объекта порождает четкий план действий. Его можно условно разбить на три этапа: модель – алгоритм – программа [9].



На первом этапе выбирается эквивалент объекта, отражающий в математической форме важнейшие свойства – законы, которым он подчиняется, связи, присущие составляющим его частям и т.д. Математическая модель исследуется теоретическими методами, что позволяет получить ванные предварительные знания об объекте.

Второй этап – выбор алгоритма для реализации модели на компьютере. Модель представляется в форме, удобной для применения численных методов, определяется последовательность вычислительных и логических операций, которые нужно произвести, чтобы найти искомые величины с заданной точностью. Вычислительные алгоритмы должны не искажать основные свойства модели и, следовательно, исходного объекта, быть экономичными и

адаптирующимися к особенностям решаемых задач и используемых компьютеров.

На третьем этапе создаются программы, переводящие модель и алгоритм на доступный компьютеру язык. К ним также предъявляются требования экономичности и адаптивности.

Создав «модель – алгоритм – программа», получаем инструмент, который сначала отлаживается, тестируется на пробных вычислительных экспериментах. После того как адекватность исходному объекту удостоверена, с моделью проводятся разнообразные и подробные опыты, дающие все требуемые качественные и количественные свойства и характеристики объекта.

1. ПОНЯТИЕ МОДЕЛИРОВАНИЯ. СПОСОБЫ ПРЕДСТАВЛЕНИЯ МОДЕЛЕЙ

Модель — аналог, прототип, шаблон, образец, используемый вместо оригинала для решения задач. Модель строится на основании ограниченного множества известных нам данных свойств, поведений об оригинале. Построение моделей и использование моделей производится с целью:

- получения неизвестных ранее данных, предсказания новых свойств и будущих поведений,
- извлечения пользы при реализации решений,
- систематизации известных данных.

Моделирование – способ, процесс замещения оригинала его моделью с последующим изучением свойств и поведения оригинала на модели.

Процесс моделирования состоит из:

- формализации (проектирование и настройка модели, систем моделей и моделей систем),
- моделирования (постановка различных задач и решение их на модели),
- интерпретации результатов моделирования, комплексирования с уже имеющимися реальными системами.

Модель вместо исходного объекта используется в случаях, когда эксперимент опасен, дорог, происходит в неудобном масштабе пространства и времени (долговременен, слишком кратковременен, протяжен...), невозможен, неповторим, ненагляден и т. д [9].

- эксперимент опасен — при деятельности в агрессивной среде вместо человека лучше использовать его макет; примером может служить луноход;
- дорог — прежде чем использовать идею в реальной экономике страны, лучше опробовать её на математической или имитационной модели экономики, просчитав на ней все «за» и «против» и получив представление о возможных последствиях;
- долговременен — изучить коррозию — процесс, происходящий десятилетия, — выгоднее и быстрее на модели;
- кратковременен — изучать детали протекания процесса обработки металлов взрывом лучше на модели, поскольку такой процесс скоротечен во времени;
- протяжен в пространстве — для изучения космогонических процессов удобны математические модели, поскольку реальные полёты к звёздам (пока) невозможны;
- микроскопичен — для изучения взаимодействия атомов удобно воспользоваться их моделью;
- невозможен — часто человек имеет дело с ситуацией, когда объекта нет, он ещё только проектируется. При проектировании важно не только

представить себе будущий объект, но и испытать его виртуальный аналог до того, как дефекты проектирования проявятся в оригинале. Сначала проектируют систему, потом её испытывают, потом снова корректируют проект и снова испытывают, и так до тех пор, пока проект не станет удовлетворять предъявляемым к нему требованиям. Процесс «проектирование-моделирование» цикличен. При этом цикл имеет вид спирали — с каждым повтором проект становится все лучше, так как модель становится все более детальной, а уровень описания точнее;

- «неповторим» — это достаточно редкий случай, когда эксперимент повторить нельзя; в такой ситуации модель — единственный способ изучения таких явлений. Пример — исторические процессы, — ведь повернуть историю вспять невозможно;

- «ненагляден» — модель позволяет заглянуть в детали процесса, в его промежуточные стадии; при построении модели исследователь как бы вынужден описать причинно-следственные связи, позволяющие понять все в единстве, системе. Построение модели дисциплинирует мышление. Важно: модель играет системообразующую и смыслообразующую роль в научном познании, позволяет *понять* явление, структуру изучаемого объекта. Не построив модель, вряд ли удастся понять логику действия системы. Это означает, что модель позволяет разложить систему на элементы, связи, механизмы, требует объяснить действие системы, определить причины явлений, характер взаимодействия составляющих.

Процесс моделирования есть процесс перехода из реальной области в виртуальную (модельную) посредством формализации, далее происходит изучение модели (собственно моделирование) и, наконец, интерпретация результатов как обратный переход из виртуальной области в реальную. Этот путь заменяет прямое исследование объекта в реальной области, то есть лобовое или интуитивное решение задачи. Итак, в самом простом случае технология моделирования подразумевает 3 этапа: формализация, собственно моделирование, интерпретация (рис.1.1).

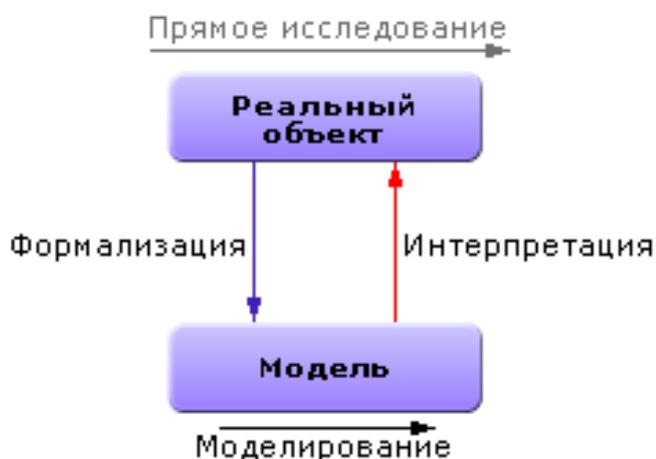


Рис.1.1. Процесс моделирования (базовый вариант)

Этапы повторяются вновь и вновь: формализация (проектирование), моделирование, интерпретация.

Весь цикл разработки показан на рис.1.14, где отражены методы, способы, приёмы, с помощью которых реализуется каждый из этапов.

Насколько аналог должен соответствовать исходному объекту?

Вариант 1: соответствие — 100%. Точность решения в этом случае максимальна, а ущерб от применения модели минимален. Но затраты на построение такой модели бесконечно велики, так как объект повторяется во всех своих деталях; фактически, создаётся точно такой же объект путём копирования его до атомов.

Вариант 2: соответствие — 0%. Модель совсем не похожа на реальный объект. Очевидно, что точность решения минимальна, а ущерб от применения модели максимален, бесконечен. Но затраты на построение такой модели нулевые.

Конечно, варианты 1 и 2 — это крайности. На самом деле модель создаётся из соображений компромисса между затратами на её построение и ущербом от неточности её применения. Это точка между двумя бесконечностями. То есть, моделируя, следует иметь в виду, что исследователь должен стремиться к оптимуму суммарных затрат, включающих ущерб от применения и затраты на изготовление модели (рис.1.2).

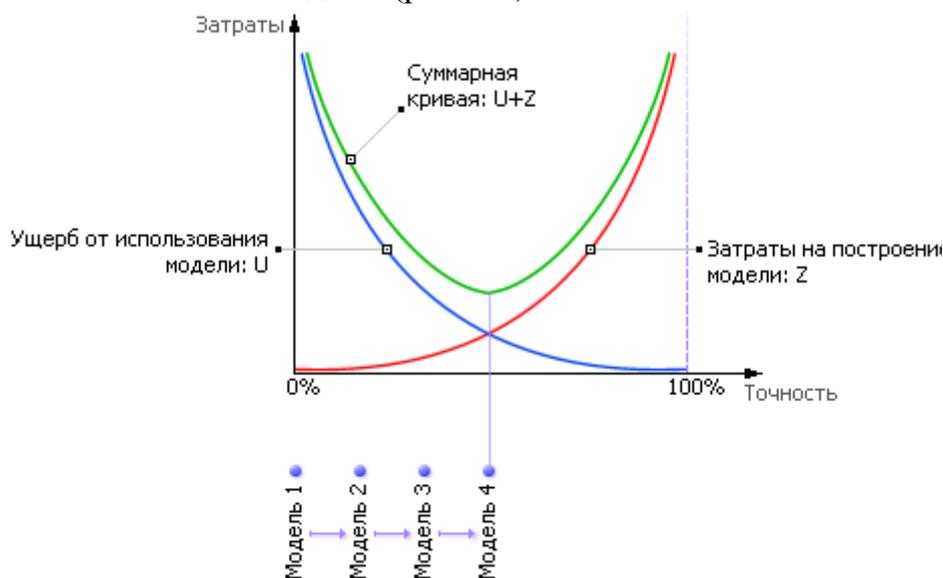


Рис.1.2. Соотношение суммарных затрат и точности для различных вариантов детализации прикладной модели

Просуммируйте две кривые затрат — получится одна кривая общих затрат. Найдите оптимум на суммарной кривой: он лежит между этими крайними вариантами. Видно, что неточные модели не нужны, но и абсолютная точность тоже не нужна, да и невозможна. Частое и распространённое заблуждение при построении моделей - требовать «как можно точнее» [9].

В модель включаются только существенные аспекты, представляющие объект, и отбрасываются *все остальные* (бесконечное большинство). Существенный или несущественный аспект описания определяют согласно цели исследования. То есть каждая модель составляется с какой-то целью. Начи-

ная моделирование, исследователь должен определить цель, отделив её от всех возможных других целей, число которых, по-видимому, бесконечно.

Указанная на рис.1.2 кривая является умозрительной и реально до начала моделирования построена быть не может. Поэтому на практике действуют таким образом: двигаются по шкале точности слева направо, то есть от простых моделей («Модель 1», «Модель 2») ко все более сложным («Модель 3», «Модель 4»). А процесс моделирования имеет циклический спиралевидный характер: если построенная модель не удовлетворяет требованиям точности, то её детализируют, дорабатывают на следующем цикле (см. рис.1.3).

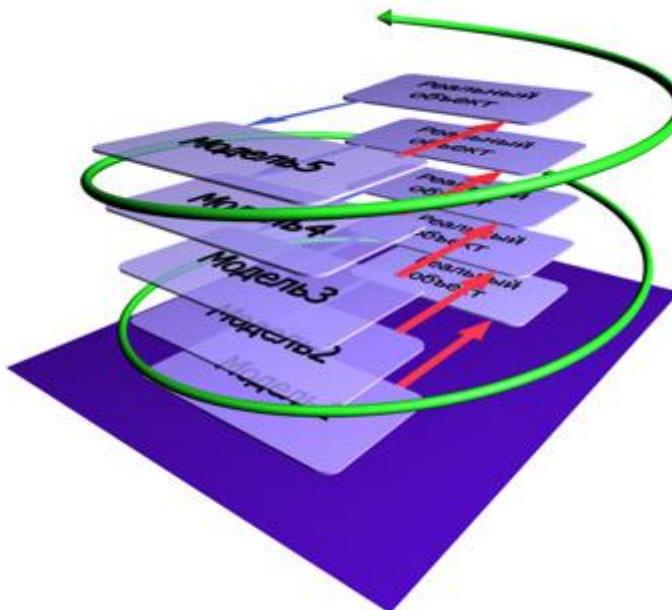


Рис. 1.3. Спиралевидный характер процесса проектирования и уточнения прикладных моделей

Улучшая модель, следят, чтобы эффект от усложнения модели превышал связанные с этим затраты. Как только исследователь замечает, что затраты на уточнение модели превышают эффект от точности при применении модели, следует остановиться, поскольку точка оптимума достигнута. Такой подход всегда гарантирует окупаемость вложений.

Двигаясь от варианта к варианту, исследователь совершенствует модель. Для построения и совершенствования моделей необходима их преемственность, средства отслеживания версий и так далее, то есть моделирование требует инструмента и опирается на технологию.

Инструмент — типовое средство, позволяющее достичь оригинальный результат и обеспечивающее сокращение затрат на выполнение промежуточных операций.

Технология — набор стандартных способов, приёмов, методов, позволяющий достичь результата гарантированного качества с помощью указанных инструментов за заранее известное время при заданных затратах, но при соблюдении пользователем объявленных требований и порядка.

Среда — совокупность рабочего пространства и инструментов на нем, поддерживающая хранение и изменение, преемственность проектов и интерпретирующая свойства объектов и систем из них.

Иногда модели пишут на языках программирования, но это долгий и дорогой процесс. Для моделирования можно использовать математические пакеты, но, как показывает опыт, в них обычно не хватает многих инженерных инструментов. Оптимальным является использование среды моделирования.

Модель, выполненная с учётом возможности её модернизации, конечно, имеет недостатки, например, низкую скорость исполнения кода. Но есть и неоспоримые достоинства. Видна и сохранена структура модели, связи, элементы, подсистемы. Всегда можно вернуться назад и что-то переделать. Сохранен след в истории проектирования модели (но когда модель отлажена, имеет смысл убрать из проекта служебную информацию). В конце концов, модель, которая сдаётся заказчику, может быть оформлена в виде специализированного автоматизированного рабочего места (АРМа), написанного уже на языке программирования, внимание в котором уже, в основном, уделено интерфейсу, скоростным параметрам и другим потребительским свойствам, которые важны для заказчика. АРМ, безусловно, вещь дорогая, поэтому выпускается он только тогда, когда заказчик полностью оттестировал проект в среде моделирования, сделал все замечания и обязуется больше не менять своих требований.

Моделирование является инженерной наукой, технологией решения задач. Это замечание - очень важное. Так как технология есть способ достижения результата с известным заранее качеством и гарантированными затратами и сроками, то моделирование, как дисциплина:

- изучает способы решения задач, то есть является инженерной наукой;
- является универсальным инструментом, гарантирующим решение любых задач, независимо от предметной области.

Смежными моделированию предметами являются: программирование, математика, исследование операций.

Программирование есть способ изложения алгоритма в языковой форме.

Алгоритм - один из способов представления мысли, процесса, явления в искусственной вычислительной среде, которой является компьютер. Специфика алгоритма состоит в отражении последовательности действий. Моделирование может использовать программирование, если моделируемый объект легко описать с точки зрения его поведения.

Отличие между алгоритмом и моделью.

Алгоритм - это процесс решения задачи путём реализации последовательности шагов, тогда как **модель** - совокупность потенциальных свойств объекта. Если к модели поставить вопрос и добавить дополнительные условия в виде исходных данных (связь с другими объектами, начальные условия, ограничения), то она может быть разрешена исследователем относительно

неизвестных. Процесс решения задачи может быть представлен алгоритмом (но известны и другие способы решения). Вообще примеры алгоритмов в природе неизвестны, они суть порождение человеческого мозга, разума, способного к установлению плана. Алгоритм - это и есть план, развёрнутый в последовательность действий. Следует различать поведение объектов, связанное с естественными причинами, и промысел разума, управляющий ходом движения, предсказывающий результат на основе знания и выбирающий целесообразный вариант поведения.

Модель + вопрос + дополнительные условия = задача.

Математика — наука, предоставляющая возможность исчисления моделей, приводимых к стандартному (каноническому) виду. Наука о нахождении решений аналитических моделей средствами формальных преобразований.

Исследование операций — дисциплина, реализующая способы исследования моделей с точки зрения нахождения наилучших управляющих воздействий на модели (синтез). По большей части имеет дело с аналитическими моделями. Помогает принимать решения, используя построенные модели.

Проектирование - процесс создания объекта и его модели; **моделирование** - способ оценки результата проектирования; моделирования без проектирования не существует.

Смежными дисциплинами для моделирования можно признать электротехнику, экономику, биологию, географию и другие в том смысле, что они используют методы моделирования для исследования собственного прикладного объекта (например, модель ландшафта, модель электрической цепи, модель денежных потоков и т.д.).

Рядом стоят дисциплины «Компьютерная графика» и «Модели и методы искусственного интеллекта» (см. рис.1.4).

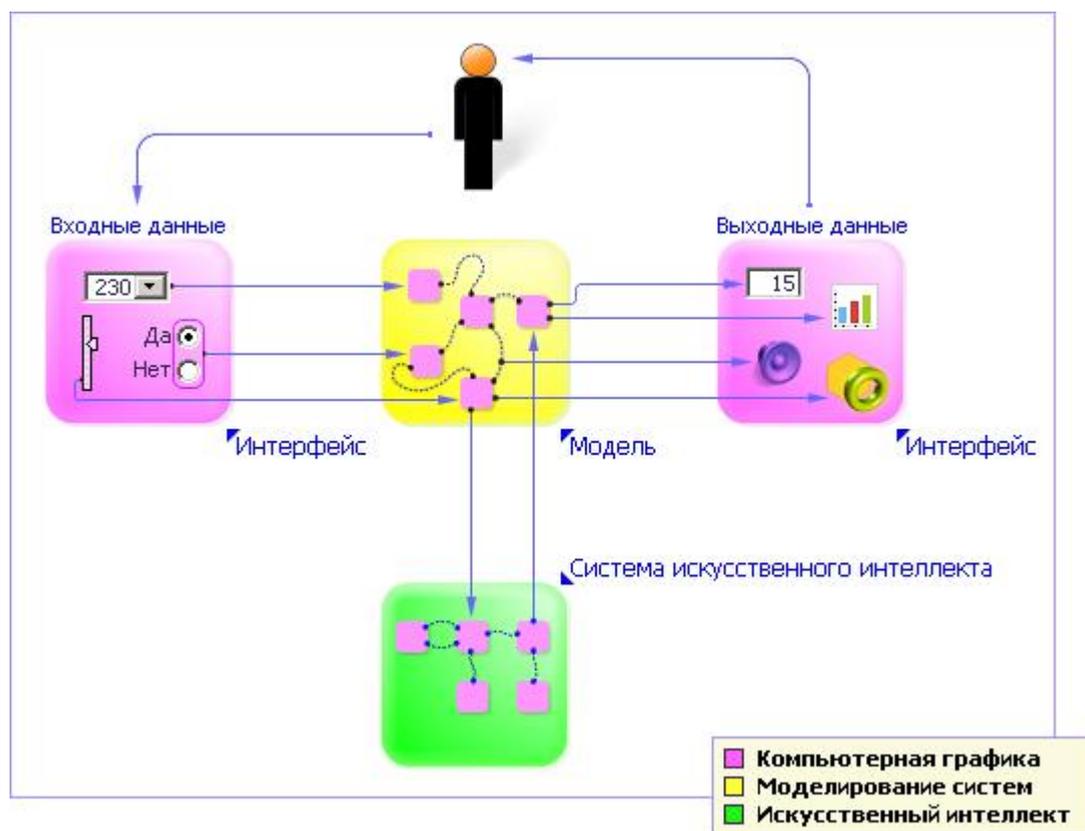


Рис. 1.4. Подсистемы при проектировании комплексных моделей

Компьютерная графика помогает организовать удобный естественный интерфейс для управления моделью, для наблюдения за её реакциями. Важно понимать, что пользователь взаимодействует с моделью не напрямую, а именно через интерфейс: с одной стороны, он посылает ей исходные (входные) данные (например, с помощью окон ввода, кнопок, движков, командной строки и т.д.), с другой - смотрит на результат работы модели, то есть воспринимает посредством интерфейса выходные данные [9].

Искусственный интеллект подразумевает построение высших моделей (например, адаптивных, которые умеют самонастраиваться, умеют создавать друг друга и т.д.). Подразумевается, что модель интеллекта в состоянии сама строить модели прикладных объектов и систем; объяснение того, как это делается, даётся в курсе «Модели и методы искусственного интеллекта». Вместе с тем заметим, что ряд исследователей, говоря об искусственном интеллекте, имеют в виду применение моделей (обучения, воспроизведения, языка и т.д.) для изучения и имитации одной из самых сложных систем во Вселенной - человека.

Искусственный интеллект - достаточно большая модель, которая содержит обширную информацию об окружающем мире и мета-модели, умеющие её достраивать. Мета-модели имеют большое подобие с имитируемым ими человеком.

В зависимости от носителя различают модели: натурные, мысленные, математические, имитационные, графические, фотографические и так далее. Каждая из моделей обладает различной способностью к прогнозу свойств

объекта. Например, по фотографии человека в анфас вряд ли можно верно представить, как выглядит его затылок. Приближение в виде трёхмерной модели — намного лучше, но можно ли с её помощью определить, когда, например, у виртуального человека вырастут волосы длиной 50 см? Имитационная модель ещё более информативна. Но наибольшей ценностью обладают модели, пригодные для решения задач, то есть обладающие прогностическими свойствами, умеющие отвечать на вопросы. Следует различать два понятия — «модель» и «задача». Модель связывает переменные между собой законами. Эти законы действуют независимо от того, какая сейчас задача стоит перед нами. Модель объективна, она подобна миру, который нас окружает, и содержит в себе информацию об этом. Структура мира (в общем смысле) неизменна, фундаментальна, модель, следовательно, тоже. Удобно ставить вопросы к модели, которая содержит нужную информацию о мире. Задача — это совокупность вопроса и модели. Можно к модели задавать все новые и новые вопросы и при этом не менять модель, но менять задачу.

Модель — способ нахождения ответов на вопросы. Чтобы ответить на поставленный вопрос, модель должна быть преобразована по правилам, обеспечивающим её эквивалентность, к виду, соответствующему ответу на вопрос. Это означает, что модель должна быть сформирована по правилам определённой алгебры. А процедура, которая помогает применить такие правила к модели, называется методом.

Рассмотрим пример.

Модель падения тела под углом к горизонту содержит информацию о координатах траектории, заданных в осях (x, y) : $y = -x^2 + 4 \cdot x - 3$ (координаты тела в полете) — см. рис. 1.5.

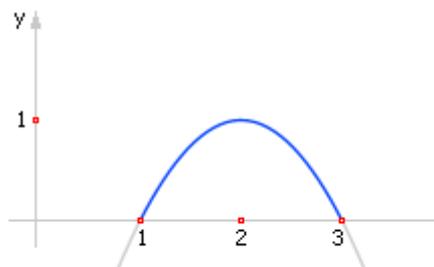


Рис. 1.5. Траектория движения тела, брошенного под углом к горизонту

Модель связывает две переменные y и x законом $f(y, x) = 0$. Модель может быть расширена некоторыми исходными данными, например, так: $y = -x^2 + 4 \cdot x - 3$, $y = 0$ (интересуют не все возможные значения y , а только точки на поверхности Земли).

$y = 0$ — это тоже закон, но более мелкого масштаба. Такие уравнения могут появляться и исчезать в зависимости от исследуемой проблемы. Обычно их называют **гипотезами**.

Вопрос: $x = ?$

Теперь модель и вопрос вместе образовали задачу:

$$y = -x^2 + 4 \cdot x - 3,$$

$$y = 0,$$

$$x = ?$$

Трактовать задачу можно так: при каких значениях x тело окажется на поверхности Земли?

Модель подразумевает, что исследователь может решать с её помощью прямые и обратные задачи.

Прямая задача не требует алгебраических преобразований, достаточно только арифметических подстановок: $x = 2$, $y = -x^2 + 4 \cdot x - 3$, $y = ?$. Ответ: $y = 1$. То есть, если на вход модели подать значение 2, то на выходе модели будет значение 1 — см. рис. 1.6.

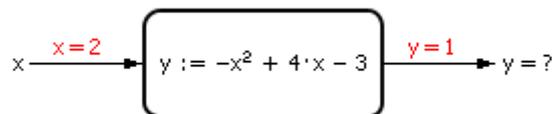


Рис. 1.6. Модели для решения прямой задачи

Обратная задача: $y = 0$, $y = -x^2 + 4 \cdot x - 3$, $x = ?$. Ответ: $x = 1$, $x = 3$. То есть ответ говорит: чтобы на выходе модели обеспечить значение 0, надо, чтобы на вход модели было подано значение 1 (или 3).

И в первом, и во втором случае мы в разной мере преобразовывали модель, но всегда так, чтобы на входе у неё была известная величина, а на выходе — неизвестная.

В первом варианте $y := -x^2 + 4 \cdot x - 3$.

Во втором варианте модель преобразуется к виду: $0 = -x^2 + 4 \cdot x - 3$. Здесь мы опустили ряд преобразований, известных из курса средней школы, а именно:

$$D := b^2 - 4 \cdot a \cdot c, \text{ где } a = -1, b = 4, c = -3.$$

$$x := (-b \pm \text{sqrt}(D)) / (2 \cdot a).$$

$$x := 1 \text{ или } x := 3.$$

Преобразования происходили с учётом правил алгебры. Если бы правила алгебры были нам неизвестны, то решить обратную задачу нам бы не удалось. А значит, не удалось бы ответить на поставленный вопрос: « $x = ?$ ».

Способность модели преобразовываться с помощью алгебры даёт возможность в дальнейшем использовать её многократно для решения различных задач, делать на ней прогнозы.

Сравните: телефонный справочник — это тоже своеобразная модель, но какие прогнозы вы можете сделать, какие обратные задачи решить? Как вычислить фамилию абонента по номеру телефона? Какую алгебру вы используете?

Поэтому, создавая модель, следует обязательно думать о том, какой алгеброй она будет преобразовываться. Создавать алгебру следует параллельно с моделью или использовать уже готовую алгебру и не отходить при построении модели от её правил.

Ещё один тип задач, который приходится решать на моделях — задачи настройки модели.

Приведём пример. При каких значениях параметра a модель $y = a \cdot x^2 + 4 \cdot x - 3$ обеспечит $y = 9$ при $x = 2$? Решаем систему уравнений:

$$\begin{aligned} y &= a \cdot x^2 + 4 \cdot x - 3 \\ y &= 9 \\ x &= 2 \end{aligned}$$

или

$$9 = a \cdot 2^2 + 4 \cdot 2 - 3$$

По правилам арифметики и алгебры, получим ответ: $a = 1$.

От показанного на рис. 1.7 структурного изображения модели можно перейти к другому, математическому, её виду: $Y = M(X)$.



Рис. 1.7. Структурное изображение модели в среде моделирования

Модель — закономерность, преобразующая входные значения в выходные. А как известно из математики, с выражением $Y = M(X)$ можно решить три вида задач, которые приведены в табл. 1.1.

Таблица 1.1.
Формы записи модели и типы решаемых задач

	Известно	Неизвестно	Решение
Прямая задача	X, M	Y	$Y = M(X)$
Обратная задача	Y, M	X	$X = M^{-1}(Y)$
Задача настройки модели	X, Y	M	$M = f(X, Y)$

Ряд моделей может быть недоопределён — это означает, что вариантов ответов много (два, три, сто или бесконечное множество). Если нужен один ответ, то проблему надо доопределять, дополнять условиями. «Недоопределён» означает, что можно произвольно, кроме гипотез, законов, ответа, потребовать дополнительно выполнение ещё каких-то условий. Возможно, при построении модели что-то не было учтено, не хватает каких-то законов. Рецепт понятен: модель надо достроить. Но может быть и по-другому. Решений много и есть, видимо, лучшие решения, и есть похуже. Тогда для нахождения лучшего решения следует сузить область решений, накладывая определённые ограничения, чтобы отсеять остальные. Такие задачи часто называют задачами управления [9].

Часть определений, которым надо безусловно удовлетворить, называются **ограничениями**.

Часть определений, относительно которых высказывают только пожелания («быть как можно больше или меньше»), называются **критериями**.

В целом получается обратная задача. А то, что надо определить - управляемая переменная. То есть интересуются: как следует изменить входной параметр (управление), чтобы обеспечить выполнение законов, не выйти за ограничения и чтобы при этом критерий принял наилучшее значение?

Пример. Модель: $y = -x^2 + 4 \cdot x - 3$. Вопрос: $x = ?$ Доопределение модели: y должен быть максимизирован, $x \geq 2.5$. Так как y должен быть максимизирован, то мы должны стараться двигаться вверх вдоль графика функции (рис. 1.8) и следить, чтобы значение x не стало меньше 2.5. Как видно из рисунка, значение y станет максимальным при $x = 2.5$. Ответ: $y = 0.75$, $x = 2.5$.

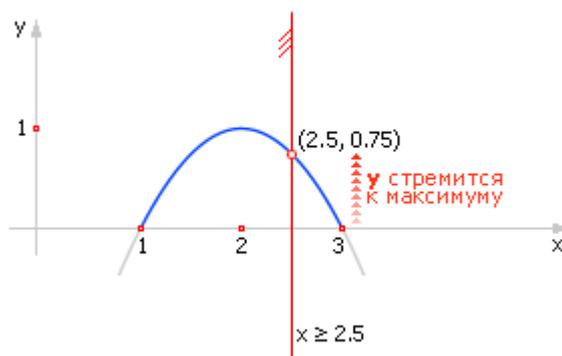


Рис. 1.8. Иллюстрация решения задачи с ограничениями

Создать модель бывает проще, чем сразу дать себе ответ на интересующий вопрос. Наверное, на практике вы замечали, что часто гораздо проще составить уравнения, чем угадать решение задачи. Например: решено разделить огромный шар размером с Землю на две половинки, полученную половинку снова поделить пополам и так далее. Попробуйте ответить на вопрос: сколько раз (n) надо провести такую операцию, чтобы размер делимой частички в результате достиг размера атома? Наверняка, сразу ответить на этот вопрос не удастся, интуиция подводит, придётся составить модель [9].

Пусть $D = 6\,400\text{ км} = 6\,400\,000\text{ м}$ — диаметр шара (Земли), а $d = 10^{-9}\text{ м}$ — диаметр атома. Тогда модель есть выражение: $2^n = D/d$ или $2^n = 6\,400\,000/10^{-9}$. Отсюда получаем: $2^n = 6.4 \cdot 10^{15}$ или $n = \log_2(6.4 \cdot 10^{15})$. Итак, приближённо, $n = 53$. Неожиданный результат, не правда ли?! Можно ли было его предугадать?

Тривиальные модели: $x = 5^\circ$; телефон друга Сидорова - 912-36-54. Такие модели не несут в себе прогностических свойств, поскольку на основе известной информации невозможно вычислить каким-либо образом другую информацию. Зная телефон одного друга Сидорова, невозможно вычислить телефон другого его друга. Это так называемые **пра-модели** (pra-model). Фактически это данные.

Заметим, что недооценка в современных условиях понятия моделирования ведёт к использованию в АРМах коммерческого назначения только данных. Именно поэтому такие АРМы не способны решать прогностические задачи и решают, в основном, только учётные задачи (см. рис. 1.9).

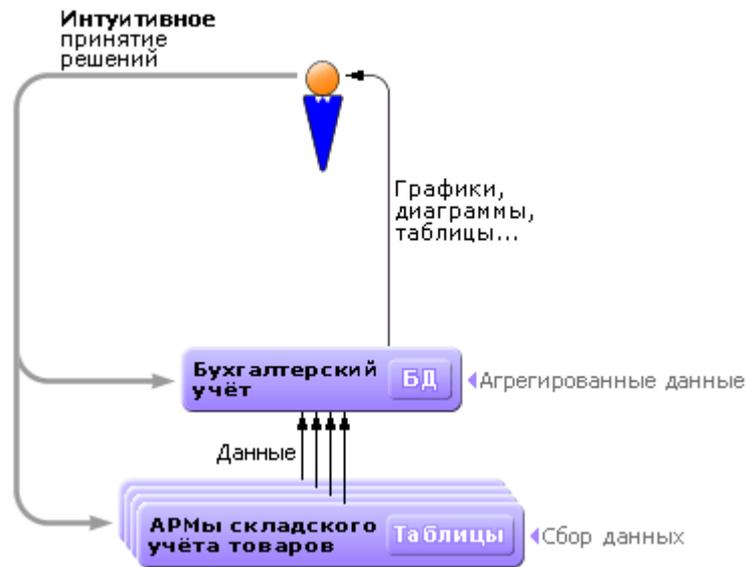


Рис. 1.9. Типовая схема взаимодействия АРМов в АСУ (без решения задач прогнозирования)

Чтобы проиграть ситуацию на предприятии на будущее, узнать, к чему приведёт то или иное решение, следует в состав АРМов включать модели (см. рис. 1.10).

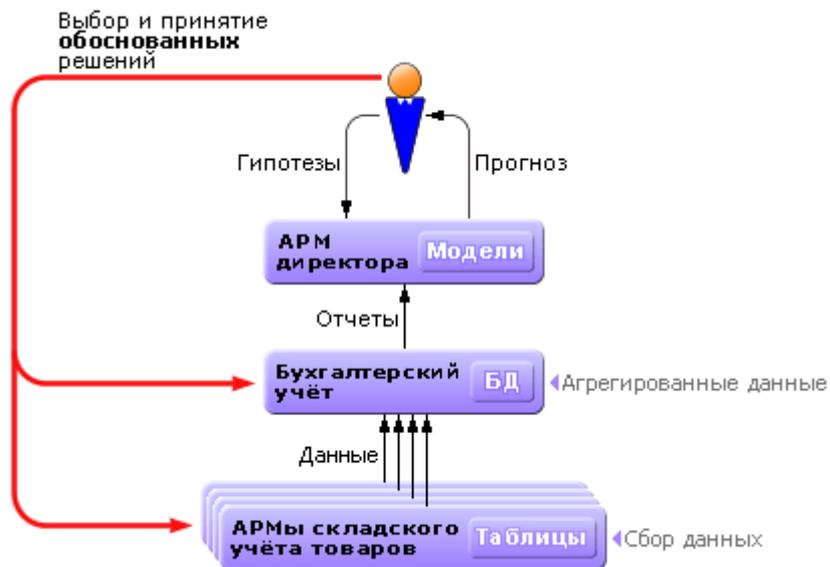


Рис. 1.10. Схема взаимодействия АРМов в АСУ (при постановке задач прогнозирования)

На рис. 1.11 показана пирамида моделей, различных по степени прогностичности.

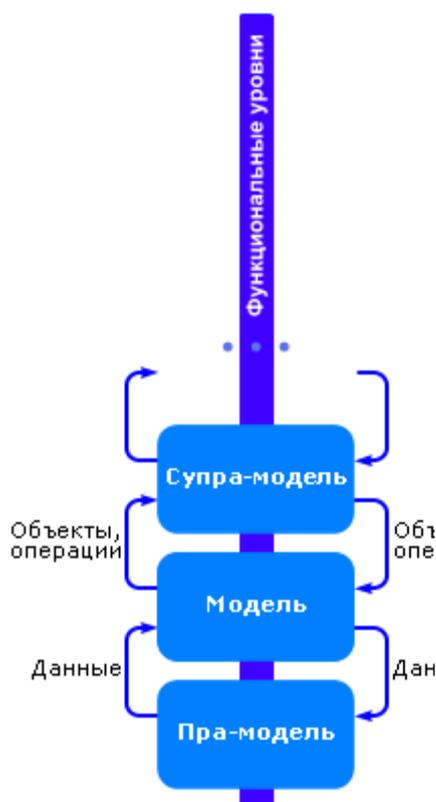


Рис. 1.11. Соотношение типов моделей по степени прогностичности

«Модель» «питается» информацией, структурированной по типу предыдущего уровня «Пра-модель», то есть она потребляет на входе *данные*, перерабатывает их и возвращает тоже *данные*, то есть модели более низкого уровня (пра-модели). Уровень «Супра-модель» потребляет на входе *модели* в виде объектов и операций, перерабатывает их и возвращает *модели* (примером таких супра-моделей могут служить грамматики, способные преобразовывать модели (уравнения). Более детально см. рис. 1.12). Данный принцип справедлив и для всех последующих (вышестоящих) уровней. Пирамида на рис. 1.11 представлена в виде функциональных уровней; это означает, что каждый последующий уровень мощнее предыдущего, то есть он позволяет получить больший, более мощный качественный результат.

Модели могут принимать различную форму, в зависимости от способа мышления исследователя, его взгляда на мир, используемой алгебры. Использование различных математических аппаратов впоследствии приводит к различным возможностям в решении задач.

Модели могут быть:

- феноменологические и абстрактные;
- активные и пассивные;
- статические и динамические;
- дискретные и непрерывные;
- детерминированные и стохастические;
- функциональные и объектные.

Феноменологические модели сильно привязаны к конкретному явлению. Изменение ситуации часто приводит к тому, что моделью воспользоваться в новых условиях достаточно сложно. Это происходит оттого, что при составлении модели её не удалось построить с точки зрения подобия внутреннему строению моделируемой системы. Феноменологическая модель передаёт внешнее подобие [9].

Абстрактная модель воспроизводит систему с точки зрения её внутреннего устройства, копирует её более точно. У неё больше возможностей, шире класс решаемых задач.

Активные модели взаимодействуют с пользователем; могут не только, как пассивные, выдавать ответы на вопросы пользователя, когда тот об этом попросит, но и сами активируют диалог, меняют его линию, имеют собственные цели. Все это происходит за счёт того, что активные модели могут самоизменяться.

Статические модели описывают явления без развития. Динамические модели прослеживают поведение систем, поэтому используют в своей записи, например, дифференциальные уравнения, производные от времени.

Дискретные и непрерывные модели. Дискретные модели изменяют состояние переменных скачком, потому что не имеют детального описания связи причин и следствий, часть процесса скрыта от исследователя. Непрерывные модели более точны, содержат в себе информацию о деталях перехода.

Детерминированные и стохастические модели. Если следствие точно определено причиной, то модель представляет процесс детерминировано. Если из-за неизученности деталей не удаётся описать точно связь причин и следствий, а возможно только описание в целом, статистически (что часто и бывает для сложных систем), то модель строится с использованием понятия вероятности.

Распределённые, структурные, сосредоточенные модели. Если параметр, описывающий свойство объекта, в любых его точках имеет одинаковое значение (хотя может меняться во времени!), то это система с сосредоточенными параметрами. Если параметр принимает разные значения в разных точках объекта, то говорят, что он распределён, а модель, описывающая объект, — распределённая. Иногда модель копирует структуру объекта, но параметры объекта сосредоточены, тогда модель — структурная.

Функциональные и объектные модели. Если описание идёт с точки зрения поведения, то модель построена по функциональному признаку. Если описание каждого объекта отделено от описания другого объекта, если описываются свойства объекта, из которых вытекает его поведение, то модель является объектно-ориентированной.

Каждый подход имеет свои достоинства и недостатки. Разные математические аппараты имеют разные возможности (мощность) для решения задач, разные потребности в вычислительных ресурсах. Один и тот же объект может быть описан различными способами. Инженер должен грамотно применять то или иное представление, исходя из текущих условий и стоящей перед ним проблемы.

Приведённая выше классификация является идеальной. Модели сложных систем обычно имеют комплексный вид, используют в своём составе сразу несколько представлений. Если удаётся свести модель к одному типу, для которого уже сформулирована алгебра, то исследование модели, решение задач на ней существенно упрощается, становится типовым. Для этого модель должна быть различными способами (упрощением, переобозначением и другими) приведена к каноническому виду, то есть к виду, для которого уже сформулирована алгебра, её методы. В зависимости от используемого типа модели (алгебраические, дифференциальные, графы и т. д.) на разных этапах её исследования используются различные математические аппараты.

Полный (расширенный) вариант схемы, представленной на рис. 1.13, см. на рис. 1.14. После прочтения всего курса лекций рекомендуется вернуться к рис. 1.14 и детально, на более глубоком уровне, ознакомиться с ним.

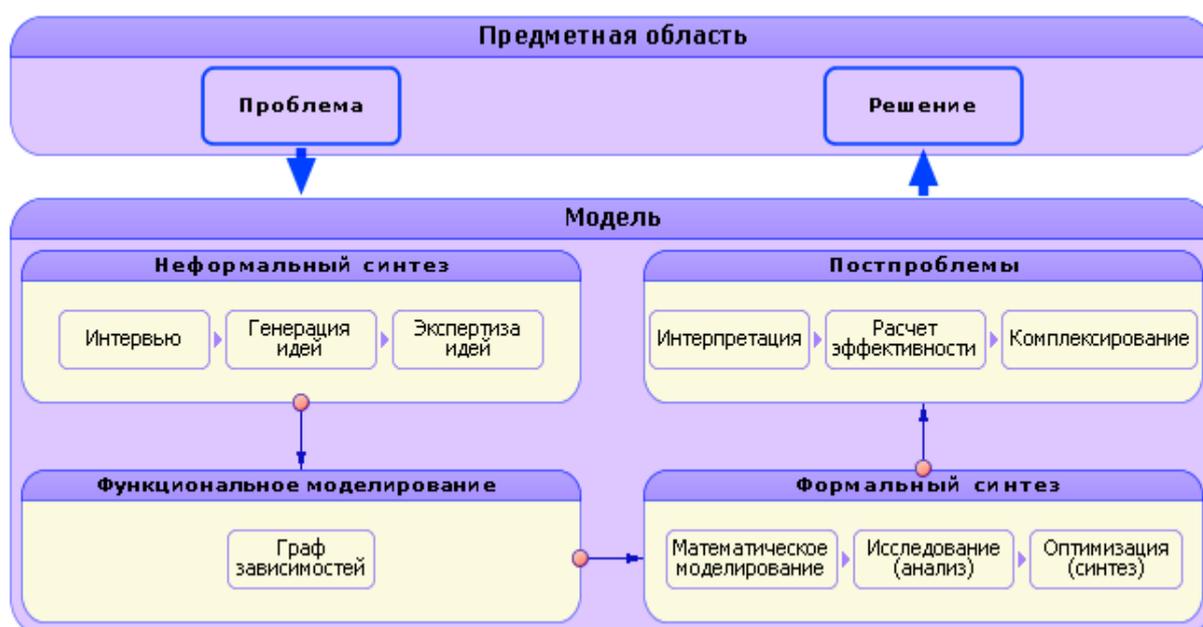


Рис. 1.13. Схема процесса моделирования

На рис. 1.15 представлены этапы построения модели.

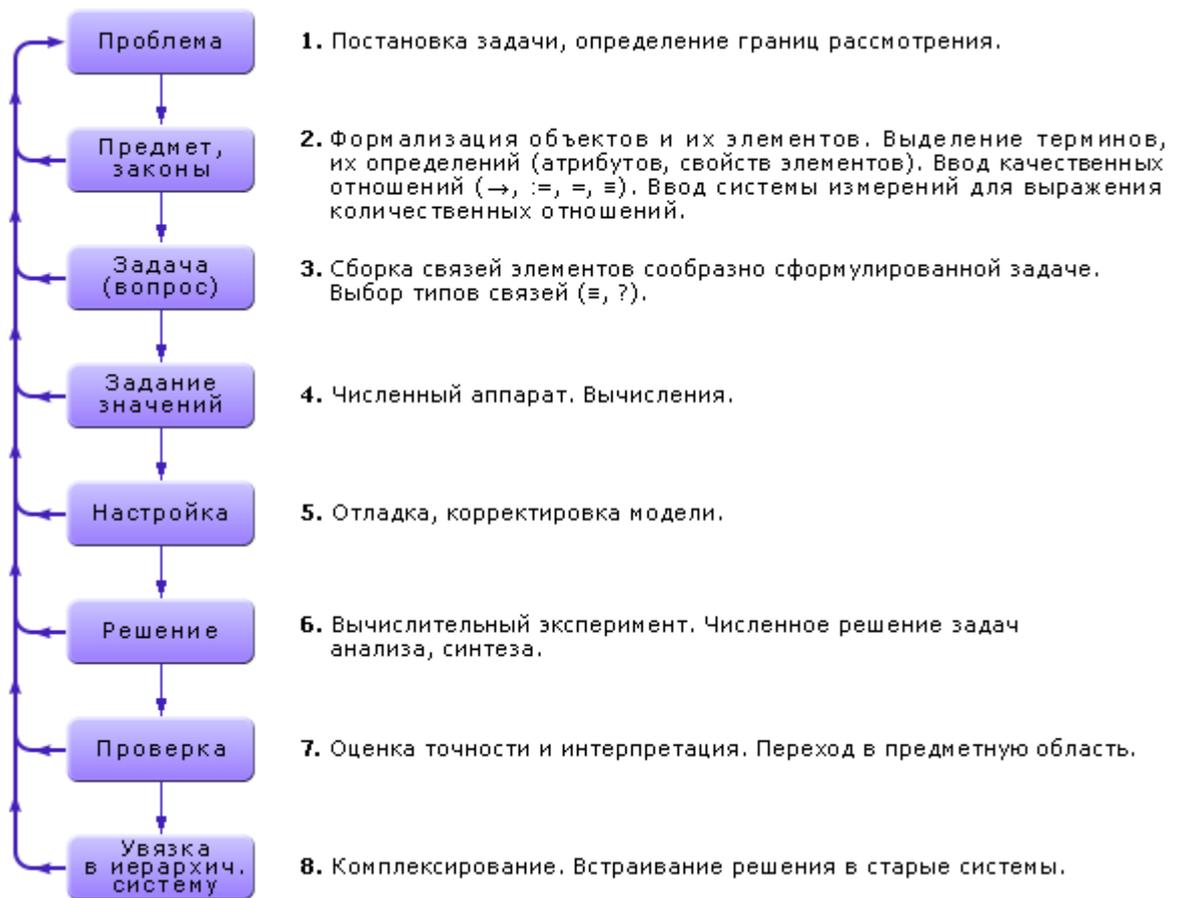


Рис. 1.15. Этапы моделирования

Спираль, которая была рассмотрена на рис. 1.3, представлена на рис. 1.15 как виток. Возможно возвращение с каждого этапа на более ранний (или более ранние) при обнаружении ошибки. Спираль имеет достаточно сложный вид, прошита дополнительными связями.

Моделирование, - это технология решения проблем, задач. Но у каждой технологии все-таки есть граница, за которой она менее эффективна. Такая граница есть и здесь. Посмотрите снова на рис. 1.13. Очевидно, что первые этапы решают менее формализованные задачи, а последующие — все более формальные. Методы первых этапов менее формализованы, а последующих — более формальные, мощные. Самые трудные и ответственные этапы для моделировщика — первые. Здесь необходимо больше интуитивных решений. И ошибка на более ранних этапах больше сказывается на дальнейших решениях, возвращаться и переделывать приходится гораздо больше, чем на последних этапах. Удачные решения на первых этапах вызывают пристальный интерес системотехников, наука моделирования проявляет к ним повышенное внимание. Формальные методы легко автоматизируются, то последние этапы схемы поддержаны программными продуктами и легко доступны конечным пользователям, но наибольший интерес сегодня представляют программные продукты, поддерживающие первые этапы — системы, помогающие формализовать задачи, а также системы, обеспечивающие сквозное про-

ектирование, доведённое до моделирования и конечной реализации (автоматическое порождение кода по описанию проекта) [9].

Здесь можно упомянуть два направления. Первое - инструментальное. Проектировщику необходим инструмент для формального описания рассматриваемого им объекта. Известно несколько таких инструментов: RationalRose, «Аналитик», IDEF по технологии SADT, Stratum. Есть инструменты, подсказывающие решения, есть просто пассивные наборы, библиотеки. Одним из инструментов нахождения решений является технология АЛРИЗ; следуя её алгоритму, отвечая на вопросы этой технологии, можно гарантированно прийти к решению.

Второй путь — аналитические системы, выводящие из фактов знания. О них пойдёт речь в следующем нашем курсе - «Модели и методы искусственного интеллекта».

В качестве примера посмотрим, как можно обнаружить, а потом описать закономерность.

Нам нужно решить «Задачу о разрезаниях», то есть надо предсказать, сколько потребуется разрезов в виде прямых линий, чтобы разделить фигуру (рис.1.16) на заданное число кусков (для примера достаточно, чтобы фигура была выпуклой).

Попробуем решить эту задачу вручную.

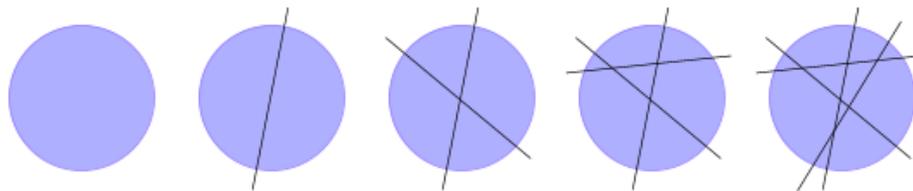


Рис. 1.16. Задача о разрезании фигуры на заданное число кусков

Из рис. 1.16 видно, что при 0 разрезах образуется 1 кусок, при 1 разрезе образуется 2 куска, при двух — 4, при трёх — 7, при четырёх — 11. Можете ли вы сейчас сказать наперёд, сколько потребуется разрезов для образования, например, 821 куска? По-моему, нет! Почему вы затрудняетесь? — Вам неизвестна закономерность $K = f(P)$, где K — количество кусков, P — количество разрезов. Как обнаружить закономерность?

Составим таблицу, связывающую известные нам числа кусков и разрезов.

Таблица 1.2.
Таблица соответствия разрезов и получающихся фрагментов фигуры

Разрезы	0	1	2	3	4	...
Куски	1	2	4	7	11	...

Рассмотрим разности между отдельными экспериментами, посмотрим, чем отличается результат одного эксперимента от другого. Поняв разницу, мы найдём способ перехода от одного результата к другому, то есть закон, связывающий K и P .

Таблица 1.3.
Таблица соответствия разрезов и получающихся фрагментов фигуры

Разрезы	0	1	2	3	4	...
Куски	1	2	4	7	11	...
Первые разности	-	$1 = 2 - 1$	$2 = 4 - 2$	$3 = 7 - 4$	$4 = 11 - 7$...

Уже кое-какая закономерность проявилась, не правда ли?
Вычислим вторые разности.

Таблица 1.4.
Таблица соответствия разрезов и получающихся фрагментов фигуры

Разрезы	0	1	2	3	4	...
Куски	1	2	4	7	11	...
Первые разности	-	$1 = 2 - 1$	$2 = 4 - 2$	$3 = 7 - 4$	$4 = 11 - 7$...
Вторые разности	-	$1 = 2 - 1$	$1 = 3 - 2$	$1 = 4 - 3$

Далее продолжать процедуру вычисления разностей смысла нет.

Функция f называется **производящей функцией**. Если она линейна, то первые разности равны между собой. Если она квадратичная, то вторые разности равны между собой. И так далее.

Функция f есть частный случай формулы Ньютона:

$$a + \frac{b \cdot n}{1} + \frac{c \cdot n(n-1)}{1 \cdot 2} + \frac{d \cdot n(n-1)(n-2)}{1 \cdot 2 \cdot 3} + \frac{e \cdot n(n-1)(n-2)(n-3)}{1 \cdot 2 \cdot 3 \cdot 4} + \dots$$

Коэффициенты a, b, c, d, e для нашей *квадратичной* функции f находятся в первых ячейках строк экспериментальной таблицы 1.5.

Таблица 1.5.
Таблица соответствия разрезов и получающихся фрагментов фигуры

Разрезы	0	1	2	3	4	...
Куски	$a = 1$	2	4	7	11	...
Первые разности		$b = 1$	$2 = 4 - 2$	$3 = 7 - 4$	$4 = 11 - 7$...
Вторые разности		-	$c = 1$	$1 = 3 - 2$	$1 = 4 - 3$...

Итак, закономерность есть, и она такова:

$$K = a + b \cdot p + c \cdot p \cdot (p-1)/2 = 1 + p + p \cdot (p-1)/2 = 0.5 \cdot p^2 + 0.5 \cdot p + 1.$$

Теперь, когда закономерность определена, можно решить обратную задачу и ответить на поставленный вопрос: сколько надо выполнить разрезов, чтобы получить 821 кусок? $K = 821, K = 0.5 \cdot p^2 + 0.5 \cdot p + 1, p = ?$

Решаем квадратное уравнение $821 = 0.5 \cdot p^2 + 0.5 \cdot p + 1$, находим корни: $p = 40$.

Поставить эксперимент оказалось затруднительно. Пришлось построить модель, найти закономерность между переменными. Модель получилась в виде уравнения. Добавив к уравнению вопрос и уравнение, отражающее известное условие, образовали задачу. Задача оказалась типового вида (канонического), то её удалось решить одним из известных методов. Поэтому задача оказалась решена.

Модель отражает причинно-следственные связи. Между переменными построенной модели действительно есть крепкая связь. Изменение одной пе-

ременной влечёт за собой изменение другой. «модель играет системообразующую и смыслообразующую роль в научном познании, позволяет понять явление, структуру изучаемого объекта, установить связь причины и следствия между собой». Модель позволяет определить причины явлений, характер взаимодействия её составляющих. Модель связывает причины и следствия через законы, то есть переменные связываются между собой через уравнения или выражения [9].

Математика это способ изучения объекта, явления, один из нескольких возможных способов мышления.

Гипотезу о связи переменных A и B надо вносить самому исследователю, извне, сверх того. А как это делает человек? Посоветовать внести гипотезу легко, но как научить этому, объяснить это действие, а значит, опять-таки как его формализовать?

Носит это рассуждение имя Геделя, который доказал теорему о неполноте — нельзя доказать правильность некоторой теории (модели) в рамках этой же теории (модели). Посмотрим на рис. 1.12. Модель более высокого уровня преобразует эквивалентно модель более низкого уровня из одного вида в другой. Или генерирует модель более низкого уровня по эквивалентному опять же её описанию. А вот саму себя она преобразовать не может. Модель строит модель. И эта пирамида моделей (теорий) бесконечна.

Приведём пример, старую известную шутку из фольклора физиков.

Задача 1. Пусть два объекта (например, пешеход и велосипедист) движутся друг другу навстречу (рис. 1.17) со скоростями V_1 и V_2 соответственно. Необходимо узнать: когда и где встретятся эти объекты?

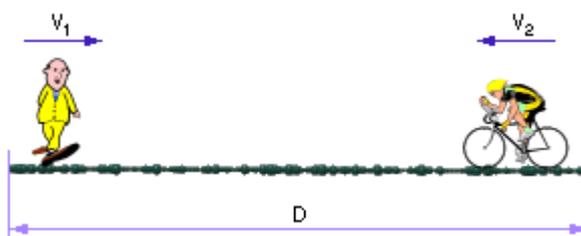


Рис. 1.17. Задача о встрече

Аналитический способ представления задачи 1

Аналитический явный способ

Эта модель весьма далека от реальности. Что-либо изучить на ней представляется проблематичным, так как из неё можно найти только время T и место встречи S . Идеализация заключается в том, что дорога считается идеально прямой, без уклонов и подъёмов, скорости объектов считаются постоянными, желания объектов не меняются, силы безграничны, отсутствуют помехи для движения, модель не зависит от величин D , V_1 , V_2 (они могут быть сколь угодно большими или малыми).

$$T_1 := D / (V_1 + V_2)$$

$$S_1 := V_1 \cdot T_1$$

Реальность обычно не имеет ничего общего с такой постановкой задачи, но за счёт большой идеализации (идеализации большого порядка) полу-

чается очень простая модель, которая может быть разрешена в общем виде (аналитически) математическими способами. Так формулируются чаще всего алгоритмические модели, где протянута цепочка вычислений от исходных данных к выходу. Поэтому мы применили в записи знак присваивания ($:=$). После вычисления правой части выражения её значение присваивается переменной, стоящей в левой части. Далее значение этой переменной применено в правой части следующего выражения. Схематически это выглядит так, как показано на рис. 1.18.

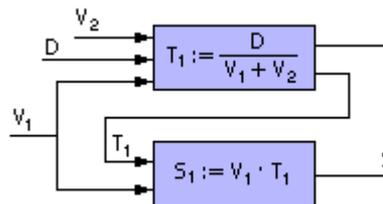


Рис. 1.18. Схема решения задачи о встрече (аналитический явный способ)

Аналитический неявный способ

В данной формулировке за счёт использования знака уравнивания получена связь переменных $f(T, V_1, V_2, D, S) = 0$ в виде системы уравнений. Устанавливая знак «?» на различные переменные, можно формулировать при необходимости целый ряд произвольных задач:

$$\begin{aligned} T_1 \cdot (V_1 + V_2) &= D \\ S_1 &= V_1 \cdot T_1 \\ T_1 &= ? \end{aligned}$$

При этом задачи формулируются пользователем и не предусматриваются специально моделировщиком. То есть модель *имеет вид объекта*. Мы получили более качественную модель. Идеализация её велика, но за счёт неявной формы записи появилась возможность изменения задачи, изучения на ней целого ряда проблем.

Имитационный способ представления задачи 1

При имитационном способе решения обязательным является наличие некоего счётчика, который позволяет моделировать процесс по шагам или по деталям процесса.

Имитационный алгоритмический способ

Повторяя пошагово расчёт в цикле, на каждом этапе работы алгоритма будем имитировать течение процесса (рис. 1.19). Обратите внимание, что процесс берётся не в целом, а как бы в деталях, по шагам. Переменная t является координатой, а значит, отслеживается счётчиком с шагом h . Идея имитации — продвигать пешехода и велосипедиста на величину $V \cdot h$ на каждом такте, где h — достаточно малая величина. Поскольку мы рассматриваем множество актов движения по отдельности, можно по ходу менять все переменные модели, например, V . Если путь пройден большой (S_1), то можно устроить привал ($V = 0$) на некоторое время. Остановка процесса имитации определяется суммой путей, пройденных велосипедистом и пешеходом навстречу друг другу, и сравнением её с расстоянием D [9].

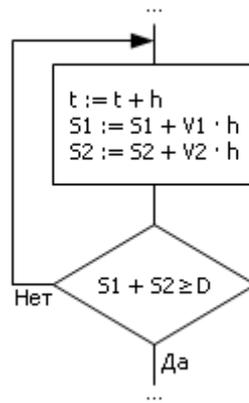


Рис. 1.19. Блок-схема решения задачи о встрече (имитационный алгоритмический способ)

На формально-математическом языке алгоритм выглядит так, как показано ниже.

$t := t + h \cdot e$
 $S_1 := S_1 + V_1 \cdot h \cdot e$
 $S_2 := S_2 + V_2 \cdot h \cdot e$
 $e := ed(D - (S_1 + S_2))$
 $f := \text{not}(e)$
 $\text{stop}(f)$

e — вспомогательный флаг;

f — флаг, показывающий, был ли пройден к текущему моменту t весь путь или нет;

$ed(x)$ — единичная функция: $ed(x) = 1$ при $x \geq 0$, иначе $ed(x) = 0$;

$\text{stop}(z)$ — функция останова вычислений при $z > 0$.

Имитационный геометрический способ

Решение может быть найдено геометрически. Для этого в осях (t, S) схемой, показанной на рис. 1.20, строятся траектории движения объектов.

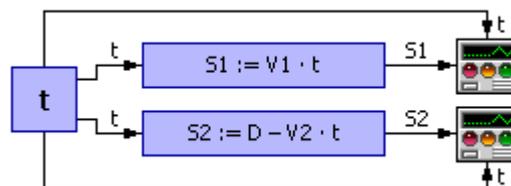


Рис. 1.20. Схема решения задачи о встрече (имитационный геометрический способ)

На рис. 1.21 вы видите картину, образованную двумя осциллограммами. Точка, в которой пересекаются осциллограммы, является предполагаемой точкой встречи двух объектов.

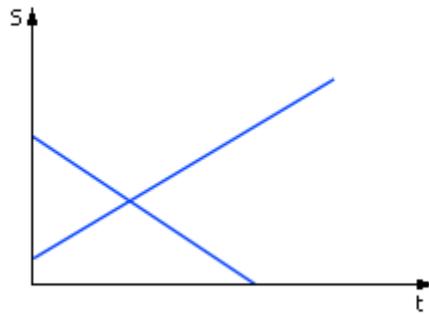


Рис. 1.21. Вид решения задачи о встрече (имитационный геометрический способ)

Имитационная статистическая постановка задачи

Главное отличие имитационных моделей от аналитических, которые мы рассмотрели выше, состоит в том, что имитационную модель можно постепенно усложнять, при этом результативность модели не падает.

Усложним задачу 1, введя в неё дополнительное условие. Представим, что на пути первого и/или второго объекта встретится помеха - пусть это будет участок железной дороги со шлагбаумом, который работает по случайному закону. Если шлагбаум открыт, то объект может переходить железную дорогу, в противном случае он не имеет права этого делать.

Промоделировать случайную работу шлагбаума можно с помощью генератора случайных чисел (ГСЧ). В различные моменты времени ГСЧ будет выдавать случайное число $r = 0$ или $r = 1$, это будет означать, что шлагбаум закрыт или, соответственно, открыт (см. рис. 1.22).



Рис. 1.22. Вид функции случайных помех

Частоту открывания шлагбаума можно контролировать, увеличивая или, наоборот, уменьшая число q , пересчитав случайное число r в z по формуле: $z := ed(q - r)$.

На рис. 1.23 дана иллюстрация усложнённой задачи 1.

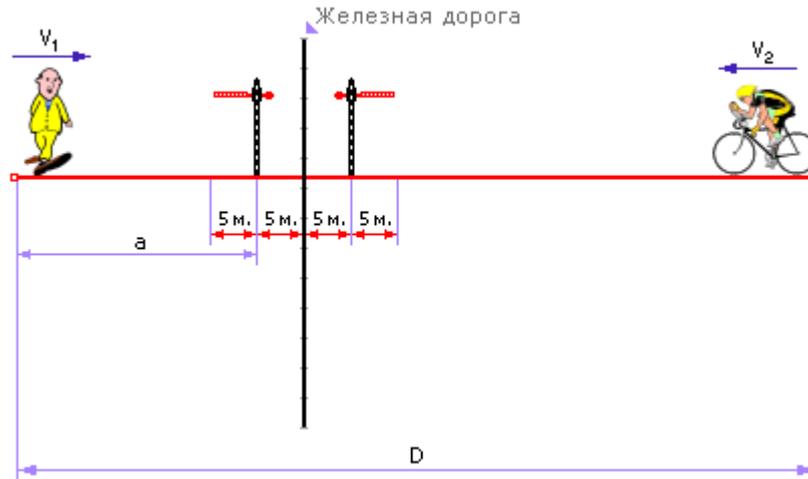


Рис. 1.23. Иллюстрация к усложнённой задаче о встрече
 На рис. 1.24 представлена алгоритмическая схема задачи.

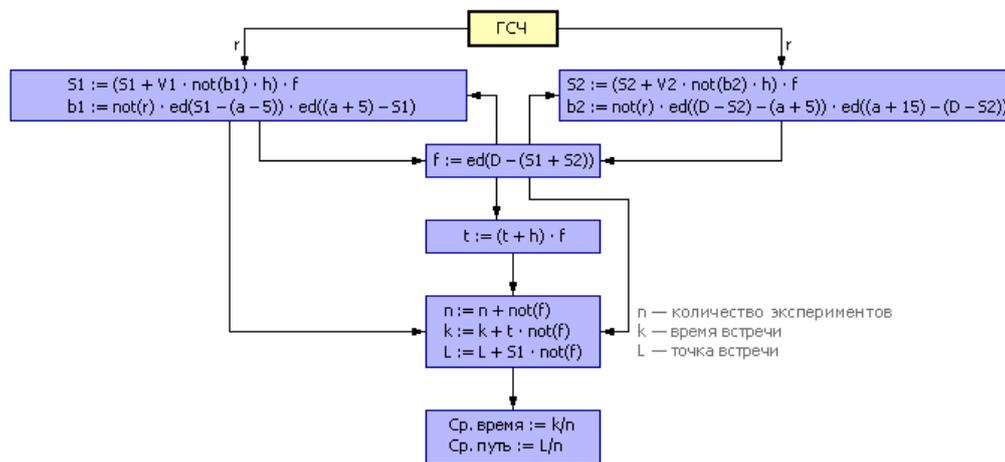


Рис. 1.24. Схема решения задачи о встрече

Условия $b1$ и $b2$ контролируют, находится ли первый и/или второй объект менее чем за 5 метров от шлагбаума, когда тот закрыт. $b1 = 1$ ($b2 = 1$) — это условие «не двигаться», если объект находится в зоне шлагбаума и шлагбаум закрыт; a — место нахождения шлагбаума, расстояние до шлагбаума от нуля; f — флаг встречи. Если $f = 0$, то встреча произошла и моделирование начинается снова с $t = 0$, $S1 = 0$, $S2 = 0$, а к статистическим счётчикам необходимо прибавить итоги эксперимента — номер эксперимента, время встречи, место встречи.

Поскольку алгоритм использует случайные числа в качестве исходных данных, придётся сделать несколько экспериментов и найти средние значения выходных величин. Результат одного эксперимента случаен и ни о чем не говорит. Среднее значение более информативно. Ещё более информативны сведения о первом и втором моменте — среднем и разбросе значений вокруг него (дисперсии) и так далее [9].

Имитационный критериальный способ

Этот способ — ещё один шаг к усложнению модели и приближению её к реальным условиям. Если во всех предыдущих случаях скорость V была известна, то теперь она — непредсказуема, даже статистически, что чаще всего и наблюдается в реальных условиях. Скорость V объекта принимается из некоторых дополнительных моделей. Причём эти модели могут быть с обратными связями. Например, скорость может зависеть от обстановки, степени достижения цели, плана, целесообразности поведения, знания местности, желания двигаться (психологический мотив). Такая постановка ведёт к адаптивным системам и системам искусственного интеллекта.

Например, если объект (пешеход) имеет возможность осмотреть местность с обзорной башни, то он может заранее просчитать свой путь и затем следовать ему. Без обзорной башни пешеход может зайти в безвыходный тупик или бесконечно долго перебирать варианты пути. Для адаптивных систем вводится критерий для оценки перспективности выбора направления движения.

Модель — способ замещения реального объекта, используемый для его исследования, когда натуральный эксперимент невозможен, дорог, опасен, долговременен.

Примеры. Поскольку исследование Луны небезопасно для человека, для этой цели используют луноход как модель исследователя; поскольку реальные эксперименты над экономикой страны дороги по своим последствиям, то используют математические модели экономики для изучения последствий управляющих решений; поскольку процесс обработки металлов взрывом скоротечен во времени, то его изучают на модели в увеличенном масштабе времени, а процесс коррозии — в уменьшенном; атом изучается в увеличенном масштабе пространства, а космогонические процессы — в уменьшенном масштабе пространства; поскольку при проектировании объекта его попросту не существует, то исследование будущих свойств объекта ведётся на модели.

Модель несёт системообразующую и смыслообразующую роль в научном познании. На модели изучают неизвестные свойства предметов. Модель стремится как можно более ярко выразить структуру явления, его главные аспекты. Модель является концентрированным выражением сущности предмета или процесса, выделяя только его основные черты.

Знания — это модели окружающего мира, фиксируемые человеком в его мозгу или на технических носителях. Модели обладают повышенной наглядностью, выделяя главные аспекты сущности, и активно используются в процессах познания и обучения. Человек, решая, как ему поступить в той или иной ситуации, всегда пытается представить себе последствия решения, для этого он проигрывает ситуацию, представляет её себе мысленно, строя модель в голове. Компьютер является усилителем для производства данной деятельности, инструментом информационной технологии. Компьютерные модели ускоряют процесс исследования, делают его более точным.

Алгоритмы — знания, выстраиваемые человеком (или, шире, разумным существом) в цепочку так, чтобы соединить исходное состояние с желаемым, целью; это один из вариантов ряда мероприятий, шагов, приводящих к цели.

Модели — это основа разумной мыслительной деятельности; модели играют роль базиса, а моделирование — роль инструмента для прогнозирования.

Процесс моделирования состоит из трёх стадий: формализации (переход от реального объекта к модели), моделирования (исследование и преобразования модели), интерпретации (перевод результатов моделирования в область реальности).

Модель есть зависимость F между входом X и выходом Y . Модель отражает закономерность $Y = F(X)$. Часто модель является законом. Модель верна в рамках допущенных при её построении гипотез. Поэтому модель ограничена некоторой областью и адекватна в ней.

Набор моделей образует научную дисциплину (механика, физика, горное дело и т. д.) Модель может быть расширена путём учёта в ней дополнительных параметров. Тогда область её применения становится шире.

Объект может быть представлен в виде аналитической или имитационной модели.

Аналитическое представление подходит лишь для очень простых и сильно идеализированных задач и объектов, которые, как правило, имеют мало общего с реальной (сложной) действительностью, но обладают высокой общностью. Аналитические модели обычно применяют для описания фундаментальных свойств объектов (поэтому ими так широко пользуется теоретическая физика), так как фундамент прост по своей сути. Сложные объекты редко удаётся описать аналитически.

Имитационное моделирование позволяет разлагать большую модель на части (объекты, «кусочки»), которыми можно оперировать по отдельности, создавая другие, более простые или, наоборот, более сложные модели. Таким образом, имитационное моделирование тяготеет к объектно-ориентированному представлению, которое естественным образом описывает объекты, их состояние, поведение, а также взаимодействие между ними. Имитационную модель можно постепенно усложнять и усложнять; аналитический способ этого не допускает или допускает, но с большими ограничениями.

Модель может быть соединена с другими моделями. Математически это означает совместное решение моделей (пересечение) и наложение тождеств на связываемые переменные. При связывании модели образуют систему, которая имеет определённую структуру (вложенную, параллельную, последовательную, смешанную, с обратными связями и т. д.)

Если к построенной модели добавить вопрос, то с помощью неё можно решить задачу, получить ответ, который заранее не очевиден. Если ответ заранее очевиден, то модель не строят. Обычно одна и та же модель годится для решения множества задач. Решая задачи, человек обычно строит модели

в своей голове и оживляет (интерпретирует) их там же или на искусственном носителе (в среде моделирования).

Как было сказано ранее:

модель + вопрос + дополнительные условия = задача.

В виде условий могут быть любые дополнительные выражения: равенства, присваивания начальных данных, неравенства, цели, функционалы и т. д., имеющие смысл ограничений, условий, дополнительных связей.

В виде вопроса может служить одна (или несколько) из неизвестных переменных.

Задача доопределяет свободные переменные модели, сужает область возможных решений.

Задачи, решаемые на модели, делятся на прямые и обратные. Прямые задачи по заданному X находят Y путём подстановки X в уравнение $Y = F(X)$. Обычно такие задачи называют **задачами анализа**. Обратные задачи по заданному Y находят X путём нахождения обратной функции F^{-1} и подстановки $X = F^{-1}(Y)$. Обычно их называют **задачами синтеза**. Если найти обратную функцию F^{-1} в явном виде затруднительно, то составляют вычислительные схемы для численного определения X . Часто к этому виду приводятся задачи управления объектами.

Ещё одной специфической задачей является нахождение по заданным X и Y зависимости F или её коэффициентов. Это **задача построения (или настройки) модели**.

Роль модели

Построив модель, исследователь может:

- прогнозировать свойства и поведение объекта как внутри области, в которой построена модель, так и (при обоснованном применении) за её пределами (прогнозирующая роль модели);
- управлять объектом, отбирая наилучшие воздействия путём испытания их на модели (управляющая роль);
- познавать явление или объект, модель которого он построил (познавательная роль модели);
- получать навыки по управлению объектом путём использования модели как тренажёра или игры (обучающая роль);
- улучшать объект, изменяя модель и испытывая её (проектная роль).

Адекватность модели

Поскольку модель является выражением конечного ряда и только важнейших для конкретного исследования аспектов сущности, то она не может быть абсолютно идентичной моделируемому объекту. Кроме этого, реальный объект бесконечен для познания, поэтому нет смысла стремиться к бесконечной точности при построении модели. Для выяснения необходимой степени адекватности обычно строят ряд моделей, начиная с грубых, простых моделей и двигаясь ко все более сложным и точным. Как только затраты на построение очередной модели начинают превышать планируемую отдачу от

модели, то уточнение модели прекращают. Первоначальные шаги производятся в каком-либо существующем универсальном моделирующем пакете. После одобрения модели под неё пишется специализированный пакет. Необходимость в этом возникает в случае, если функционирование модели в универсальной среде моделирования не удовлетворяет требованиям быстродействия (или каким-то другим).

Моделирование — прикладная инженерная наука класса технологических. Моделирование — дисциплина, ставящая целью построение моделей и их исследование посредством собственных универсальных методов, а также специфических методов смежных с ней наук (математика, исследование операций, программирование).

2. ЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ

В целях исследований часто бывает удобно представить исследуемый объект в виде ящика, имеющего входы и выходы, не рассматривая детально его внутренней структуры. Конечно, преобразования в ящике (на объекте)

происходят (сигналы проходят по связям и элементам, меняют свою форму и т. п.), но при таком представлении они происходят скрыто от наблюдателя.

По степени информированности исследователя об объекте существует деление объектов на три типа «ящиков»:

- «**белый ящик**»: об объекте известно все;
- «**серый ящик**»: известна структура объекта, неизвестны количественные значения параметров;
- «**черный ящик**»: об объекте неизвестно ничего.

Черный ящик условно изображают как на рис. 2.1.



Рис. 2.1. Обозначение черного ящика

Значения на входах и выходах черного ящика можно наблюдать и измерять. Содержимое ящика неизвестно [9].

Задача состоит в том, чтобы, зная множество значений на входах и выходах, построить модель, то есть определить функцию ящика, по которой вход преобразуется в выход. Такая задача называется **задачей регрессионного анализа**.

В зависимости от того, доступны входы исследователю для управления или только для наблюдения, можно говорить про активный или пассивный эксперимент с ящиком.

Например, перед нами стоит задача определить, как зависит выпуск продукции от количества потребляемой электроэнергии. Результаты наблюдений отобразим на графике (см. рис. 2.2). Всего на графике n экспериментальных точек, которые соответствуют n наблюдениям.

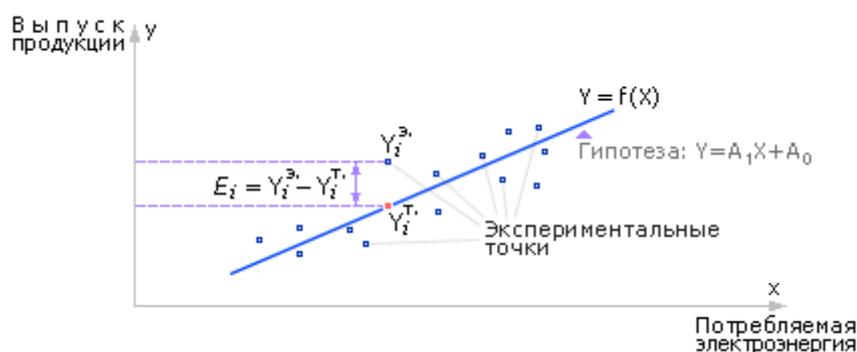


Рис. 2.2. Графический вид представления результатов наблюдения над черным ящиком

Для начала предположим, что мы имеем дело с черным ящиком, имеющим один вход и один выход. Допустим для простоты, что зависимость между входом и выходом линейная или почти линейная. Тогда данная модель будет называться **линейной одномерной регрессионной моделью**.

1) Исследователь вносит гипотезу о структуре ящика

Рассматривая экспериментально полученные данные, предположим, что они подчиняются линейной гипотезе, то есть выход Y зависит от входа X линейно, то есть гипотеза имеет вид: $Y = A_1X + A_0$ (рис. 2.2).

2) Определение неизвестных коэффициентов A_0 и A_1 модели

Линейная одномерная модель (рис. 2.3).

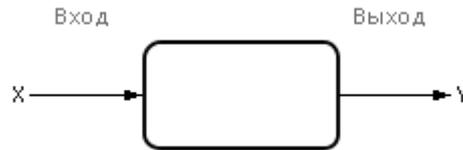


Рис. 2.3. Одномерная модель черного ящика

Для каждой из n снятых экспериментально точек вычислим ошибку (E_i) между экспериментальным значением ($Y_i^{\text{Эксп.}}$) и теоретическим значением ($Y_i^{\text{Теор.}}$), лежащим на гипотетической прямой $A_1X + A_0$ (см. рис. 2.2):

$$E_i = (Y_i^{\text{Эксп.}} - Y_i^{\text{Теор.}}), i = 1, \dots, n;$$

$$E_i = Y_i - A_0 - A_1 \cdot X_i, i = 1, \dots, n.$$

Ошибки E_i для всех n точек следует сложить. Чтобы положительные ошибки не компенсировали в сумме отрицательные, каждую из ошибок возводят в квадрат и складывают их значения в суммарную ошибку F уже одного знака:

$$E_i^2 = (Y_i - A_0 - A_1 \cdot X_i)^2, i = 1, \dots, n.$$

$$F(A_0, A_1) = \sum_{i=1}^n E_i^2$$

Цель метода — минимизация суммарной ошибки F за счет подбора коэффициентов A_0, A_1 . Другими словами, это означает, что необходимо найти такие коэффициенты A_0, A_1 линейной функции $Y = A_1X + A_0$, чтобы ее график проходил как можно ближе одновременно ко всем экспериментальным точкам. Поэтому данный метод называется **методом наименьших квадратов**.

$$F(A_0, A_1) = \sum_{i=1}^n E_i^2 = \sum_{i=1}^n (Y_i - A_0 - A_1 X_i)^2 \Rightarrow \min_{A_0, A_1}$$

Суммарная ошибка F является функцией двух переменных A_0 и A_1 , то есть $F(A_0, A_1)$, меняя которые, можно влиять на величину суммарной ошибки (см. рис. 2.4).

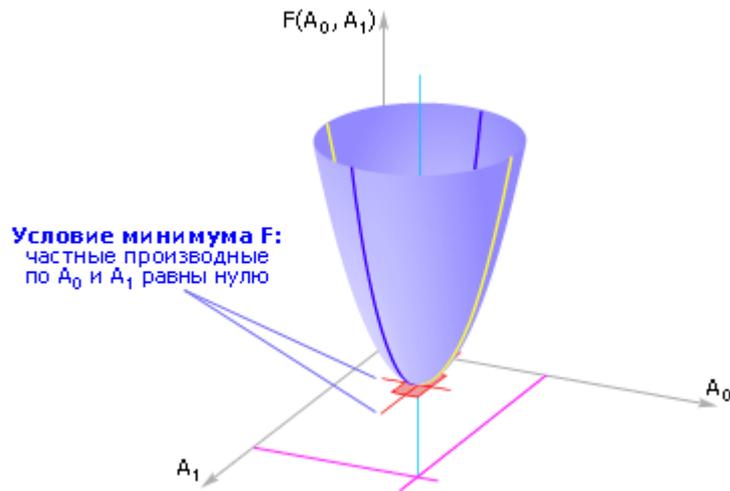


Рис. 2.4. Вид функции ошибки

Чтобы суммарную ошибку минимизировать, найдем частные производные от функции F по каждой переменной и приравняем их к нулю (условие экстремума):

$$\frac{\partial F}{\partial A_0} = -2 \sum_{i=1}^n (Y_i - A_0 - A_1 X_i) = 0$$

$$\frac{\partial F}{\partial A_1} = -2 \sum_{i=1}^n (Y_i - A_0 - A_1 X_i) X_i = 0$$

После раскрытия скобок получим систему из двух линейных уравнений:

$$\sum_{i=1}^n A_0 + A_1 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i$$

$$A_0 \sum_{i=1}^n X_i + A_1 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$

Для нахождения коэффициентов A_0 и A_1 методом Крамера представим систему в матричной форме:

$$\begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix}$$

Решение имеет вид:

$$A_0 = \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i^2 - \sum_{i=1}^n X_i Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

$$A_1 = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

Вычисляем значения A_0 и A_1 .

3) Проверка

Чтобы определить, принимается гипотеза или нет, нужно, во-первых, рассчитать ошибку между точками заданной экспериментальной и полученной теоретической зависимости и суммарную ошибку:

$$E_i = (Y_i^{\text{Эксп.}} - Y_i^{\text{Теор.}}), i = 1, \dots, n$$

$$F(A_0, A_1) = \sum_{i=1}^n E_i^2$$

И, во-вторых, необходимо найти значение σ по формуле $\sigma = \sqrt{\frac{F}{n}}$, где F — суммарная ошибка, n — общее число экспериментальных точек.

Если в полосу, ограниченную линиями $Y^{\text{Теор.}} - S$ и $Y^{\text{Теор.}} + S$ (рис. 2.5), попадает 68,26% и более экспериментальных точек $Y_i^{\text{Эксп.}}$, то выдвинутая нами гипотеза принимается. В противном случае выбирают более сложную гипотезу или проверяют исходные данные. Если требуется большая уверенность в результате, то используют дополнительное условие: в полосу, ограниченную линиями $Y^{\text{Теор.}} - 2S$ и $Y^{\text{Теор.}} + 2S$, должны попасть 95,44% и более экспериментальных точек $Y_i^{\text{Эксп.}}$.

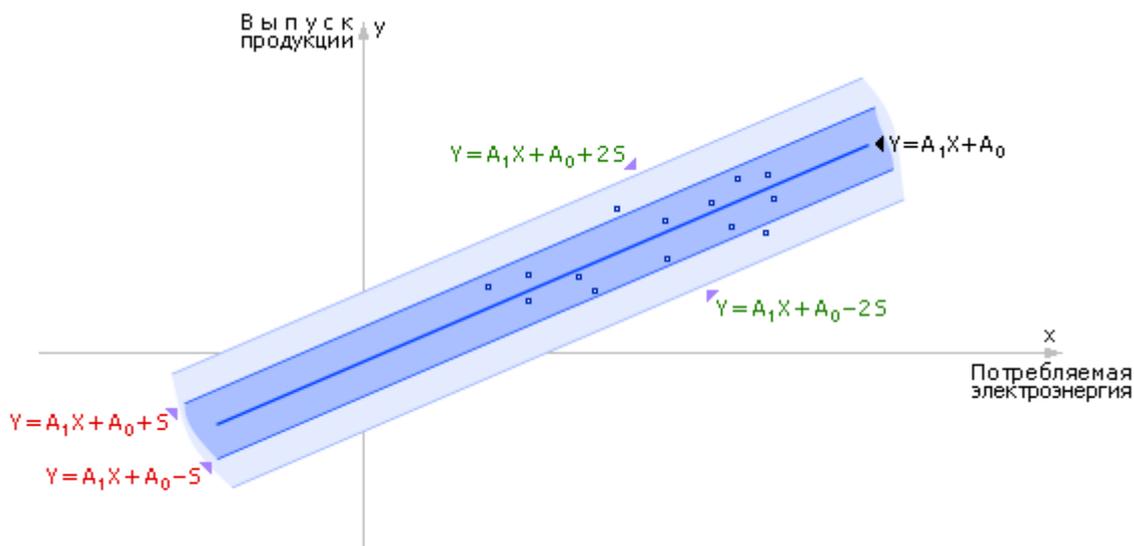


Рис. 2.5. Исследование допустимости принятия гипотезы

Расстояние S связано с σ следующим соотношением:

$$S = \sigma / \sin(\beta) = \sigma / \sin(90^\circ - \arctg(A_1)) = \sigma / \cos(\arctg(A_1)),$$

что проиллюстрировано на рис. 2.6.

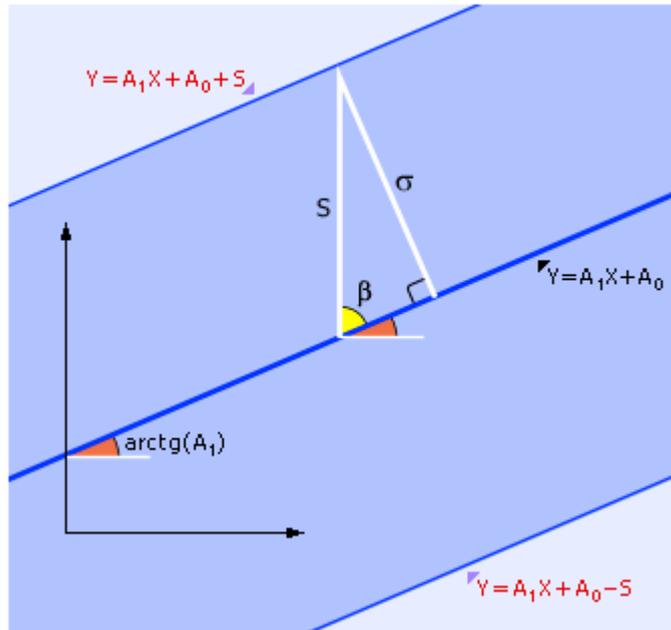


Рис. 2.6. Связь значений σ и S

Условие принятия гипотезы выведено из нормального закона распределения случайных ошибок (см. рис. 2.7). P — вероятность распределения нормальной ошибки [9].

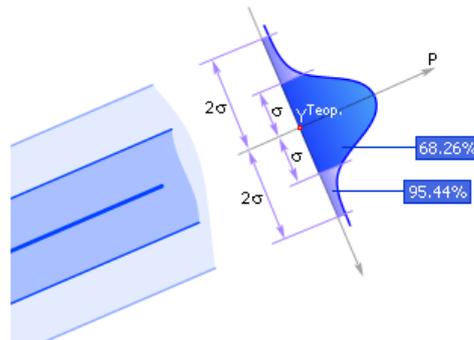


Рис. 2.7. Иллюстрация закона нормального распределения ошибок

Наконец, приведем на рис. 2.8 графическую схему реализации одномерной линейной регрессионной модели.

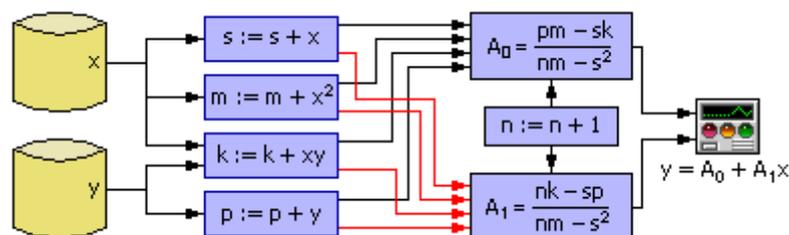


Рис. 2.8. Схема реализации метода наименьших квадратов в среде моделирования

Линейная множественная модель

Предположим, что функциональная структура ящика снова имеет линейную зависимость, но количество входных сигналов, действующих одновременно на объект, равно m (см. рис. 2.9):

$$Y = A_0 + A_1 \cdot X_1 + \dots + A_m \cdot X_m.$$

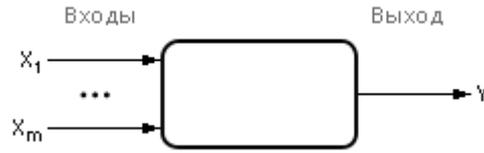


Рис. 2.9. Обозначение многомерного черного ящика на схемах

Так как подразумевается, что мы имеем экспериментальные данные о всех входах и выходах черного ящика, то можно вычислить ошибку между экспериментальным ($Y_i^{\text{Эксп.}}$) и теоретическим ($Y_i^{\text{Теор.}}$) значением Y для каждой i -ой точки (пусть, как и прежде, число экспериментальных точек равно n):

$$E_i = (Y_i^{\text{Эксп.}} - Y_i^{\text{Теор.}}), i = 1, \dots, n;$$

$$E_i = Y_i - A_0 - A_1 \cdot X_{1i} - \dots - A_m \cdot X_{mi}, i = 1, \dots, n.$$

Минимизируем суммарную ошибку F :

$$F(A_0, A_1, \dots, A_m) = \sum_{i=1}^n E_i^2 \Rightarrow \min_{A_0, A_1, \dots, A_m}$$

Ошибка F зависит от выбора параметров A_0, A_1, \dots, A_m . Для нахождения экстремума приравняем все частные производные F по неизвестным A_0, A_1, \dots, A_m к нулю:

$$\frac{\partial F}{\partial A_j} = 0, j = \overline{0, m}$$

Получим систему из $m + 1$ уравнения с $m + 1$ неизвестными, которую следует решить, чтобы определить коэффициенты линейной множественной модели A_0, A_1, \dots, A_m . Для нахождения коэффициентов методом Крамера представим систему в матричном виде:

$$\begin{pmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} & \dots & \sum_{i=1}^n X_{mi} \\ \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{1i}X_{1i} & \sum_{i=1}^n X_{2i}X_{1i} & \dots & \sum_{i=1}^n X_{mi}X_{1i} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{1i}X_{2i} & \sum_{i=1}^n X_{2i}X_{2i} & \dots & \sum_{i=1}^n X_{mi}X_{2i} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n X_{mi} & \sum_{i=1}^n X_{1i}X_{mi} & \sum_{i=1}^n X_{2i}X_{mi} & \dots & \sum_{i=1}^n X_{mi}X_{mi} \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ \dots \\ A_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_i X_{1i} \\ \sum_{i=1}^n Y_i X_{2i} \\ \dots \\ \sum_{i=1}^n Y_i X_{mi} \end{pmatrix}$$

Вычисляем коэффициенты A_0, A_1, \dots, A_m .

Далее, по аналогии с одномерной моделью (см. 3). «Проверка»), для каждой точки вычисляется ошибка E_i ; затем находится суммарная ошибка F и значения σ и S с целью определить, принимается ли выдвинутая гипотеза о линейности многомерного черного ящика или нет.

При помощи подстановок и переобозначений к линейной множественной модели приводятся многие нелинейные модели.

3. НЕЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ

Полиномиальная множественная регрессионная модель

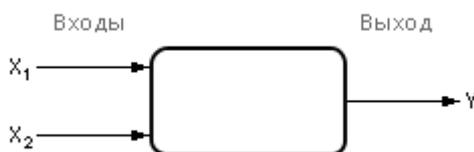


Рис. 3.1. Обозначение двумерной модели черного ящика на схемах

Если черный ящик имеет, например, два входа, а зависимость выхода от входов напоминает квадратичную, то целесообразно выбрать такую гипотезу:

$$Y = A_0 + A_1 \cdot X_1 + A_2 \cdot X_2 + A_3 \cdot X_1 \cdot X_2 + A_4 \cdot X_1 \cdot X_1 + A_5 \cdot X_2 \cdot X_2.$$

Обозначим: $Z_1 = X_1 \cdot X_2$; $Z_2 = X_1 \cdot X_1$; $Z_3 = X_2 \cdot X_2$ и подставим эти выражения в предыдущую формулу:

$$Y = A_0 + A_1 \cdot X_1 + A_2 \cdot X_2 + A_3 \cdot Z_1 + A_4 \cdot Z_2 + A_5 \cdot Z_3.$$

Таким образом, данная задача сведена к линейной множественной модели. А модель черного ящика теперь выглядит так, как показано на рис. 3.2.

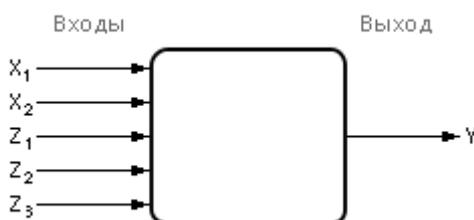


Рис. 3.2. Преобразованная модель черного ящика

Мультипликативная регрессионная модель

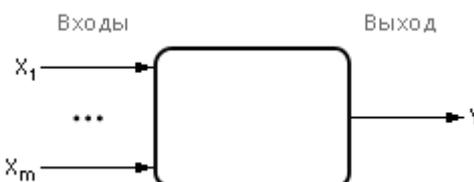


Рис. 3.3. Обозначение модели многомерного черного ящика на схемах

$$Y = A_0 \cdot X_1^{A_1} \cdot X_2^{A_2} \cdot \dots \cdot X_m^{A_m}.$$

Прологарифмируем левую и правую части данного уравнения:

$$\ln(Y) = \ln(A_0) + A_1 \cdot \ln(X_1) + A_2 \cdot \ln(X_2) + \dots + A_m \cdot \ln(X_m).$$

Обозначим:

$$W = \ln(Y), B_0 = \ln(A_0), Z_1 = \ln(X_1), Z_2 = \ln(X_2), \dots, Z_m = \ln(X_m).$$

Получим:

$$W = B_0 + A_1 \cdot Z_1 + A_2 \cdot Z_2 + \dots + A_m \cdot Z_m.$$

То есть вновь осуществлен переход к линейной множественной модели.

Обратная регрессионная модель

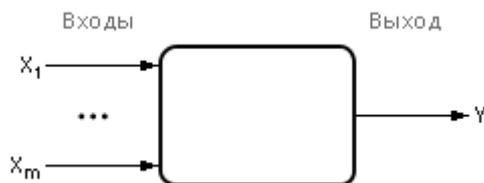


Рис. 3.4. Обозначение модели многомерного черного ящика на схемах

$$Y = k / (A_0 + A_1 X_1 + \dots + A_m X_m).$$

Заменим: $W = 1/Y$, $a_i = A_i/k$. И перейдем к линейной множественной модели:

$$W = a_0 + a_1 \cdot X_1 + \dots + a_m \cdot X_m.$$

Экспоненциальная модель

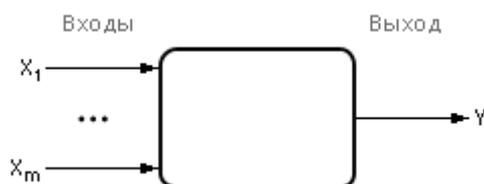


Рис. 3.5. Обозначение модели многомерного черного ящика на схемах

$$Y = e^{B_0 + B_1 X_1 + B_2 X_2 + \dots + B_m X_m}.$$

Прологарифмируем левую и правую части уравнения:

$$\ln(Y) = B_0 + B_1 \cdot X_1 + B_2 \cdot X_2 + \dots + B_m \cdot X_m.$$

Выполним замену $W = \ln(Y)$ и получим:

$$W = B_0 + B_1 \cdot X_1 + B_2 \cdot X_2 + \dots + B_m \cdot X_m.$$

Далее пользуемся выражением для линейной множественной модели.

4. ДИНАМИЧЕСКИЕ СИСТЕМЫ

До этого мы рассматривали статические модели, то есть случай, когда один эксперимент не зависит от другого. Можно сказать, что система не обладала памятью. То есть, в какой бы момент времени мы ни измеряли значение выходной величины, при одинаковом значении входного сигнала результат был один и тот же. Если каждый раз значение на выходе, при одном и том же входном значении, разное, то есть зависит от того, в какой последовательности подавались входные значения, то мы имеем дело с *динамической системой*.

Динамические системы, в отличие от статических, помнят свое прошлое состояние, то есть обладают памятью. Поэтому в записи модели динамических систем присутствует производная, связывающая прошлое состояние системы с настоящим. Чем большей памятью обладает система, тем больше состояний из прошлого влияют на настоящее, тем большая степень старшей производной используется в записи модели. В данной лекции рассматриваются динамические системы [9].

Задача 1. На входе и выходе черного ящика (рис. 4.1) имеются зависимости параметров X и Y от времени t . Задача состоит в том, чтобы адекватно определить черный ящик.



Рис. 4.1. Черный ящик, содержащий динамическую систему. Условное обозначение

Графики зависимостей $X(t)$ и $Y(t)$ могут быть самыми разными, например, такими, как показано на рис. 4.2.

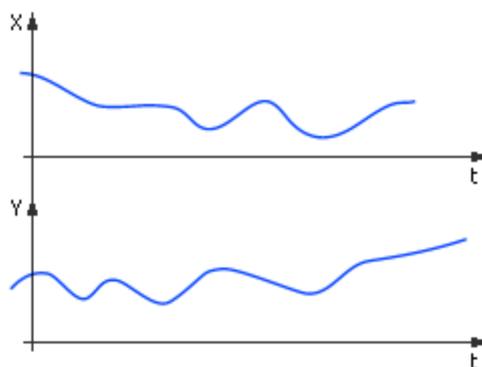


Рис. 4.2. Временные зависимости -входной и выходной сигналы

Поскольку моделирование систем подразумевает численные расчеты на компьютере, то аналоговый сигнал переводят в дискретный вид, для этого с определенной частотой исходный сигнал дискретизируют, как показано на рис. 4.3.

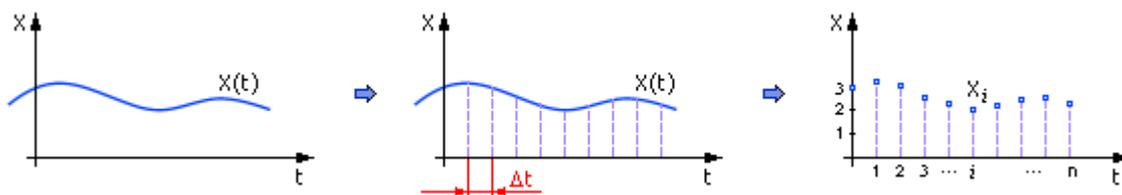


Рис. 4.3. Дискретизированный временной сигнал

По этим данным строят таблицу отсчетов (см. табл. 4.1, где $\Delta t = 0.1$).

Таблица 4.1.

Табличное представление временного сигнала

i	0	1	2	3	...	i	...	n
t	0	0.1	0.2	0.3	...	$\Delta t \cdot i$...	$\Delta t \cdot n$
x_i	3	3.2	3.1	2.6	...	x_i	...	x_n

Совокупность значений переменной в таблице, упорядоченных во времени, часто называют **динамическим рядом**. Естественно, часть информации при такой операции теряется. Чем меньше расстояние между отсчетами, чем больше частота дискретизации, тем меньше потери информации. Частоту дискретизации принимают такой, чтобы не потерять высокочастотные составляющие в сигнале, отдельные пики (см. также «Лекция 08. Модель динамической системы в виде Фурье представления (модель объекта)»).

Любая динамическая система характеризуется рядом параметров. Обычно параметрами называют коэффициенты при производных (первой, второй и т. д.) в записи модели. Чем большая степень старшей производной присутствует в записи модели, тем больший порядок динамической системы, тем глубже ее память, и тем больше коэффициентов (параметров) надо определить, чтобы идентифицировать систему.

Как определить параметры динамической системы? Сначала нужно оценить порядок динамической системы: он совпадает со степенью наибольшей из производных Y по отношению к t . Допустим, что на вход системы, до этого находившейся в нулевых начальных условиях, подали единичный сигнал $X(t)$, как показано на рис. 4.4.

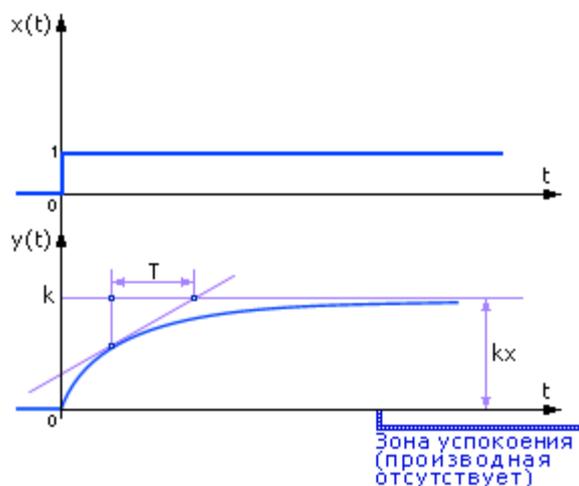


Рис. 4.4. Входной и выходной сигнал, для системы первого порядка

При нулевых начальных условиях, если входной сигнал отсутствует, выходной сигнал равен нулю, и говорят, что система находится в покое. Если подать на вход единичный (пробный) сигнал и удерживать его на входе достаточно долго, то система на выходе попытается подчиниться ему, начнет отклоняться от нулевого состояния. Ожидается, что система на выходе должна дойти до значения kx , то есть увеличить сигнал x в k раз (k — коэффициент усиления входного сигнала). Но, как видно, происходит это не сразу, а с некоторой задержкой, сигнал на выходе нарастает постепенно, инерционно. Насколько инерционно реагирует система, зависит от параметра T . Система достигнет значения kx на выходе и будет держать этот сигнал, пока держится на входе единичный сигнал. Переход от нуля до kx происходит во времени. Переход — процесс динамический, то есть в сигнале присутствует изменение, которое описывается производной, и выход оказывается меньше входа на некоторую величину f :

$$y = kx - f(dy/dt).$$

Когда система достигнет на выходе значения равного kx , то изменений не будет, значение производной станет равной нулю. $y = kx$.

$y = kx$ — частный случай инерционного звена.

Если на выходе будет наблюдаться экспоненциальный сигнал, то система будет называться системой первого порядка (или звеном первого порядка). Для ее описания достаточно одной производной (а в решении модели будет присутствовать один интеграл):

$$T \frac{dy}{dt} + y = kx$$

У такой системы два параметра — T и k .

Заметим, что один интеграл у линейных динамических систем всегда «порождает» одну экспоненту, двойной интеграл — сумму двух экспонент, и так далее [9].

Чтобы определить, является ли кривая экспонентой, в каждой ее точке проводится касательная до пересечения с линией установившегося уровня (на рис. 4.4 это линия $y(t) = k$); в случае, если кривая является экспонентой, величина T в любой точке будет постоянной.

Определить T , используя график, можно еще так. Проведите линию, параллельную оси t на уровне $0.95k$. Из точки, где эта линия пересечет экспоненту, опустите перпендикуляр на ось t . Отрезок от 0 до точки пересечения перпендикуляра с осью t будет равен $3T$.

T характеризует инерционность системы (память). При малой величине T система слабо зависит от предыстории и вход мгновенно заставляет измениться выход. При большом значении T система медленно реагирует на входной сигнал, а при очень большом значении T система выдает неизменный выходной сигнал, практически не реагируя на входные воздействия.

Коэффициент k характеризует способность системы к усилению (при $k < 1$ — к ослаблению) уровня входного сигнала. Чтобы определить коэффициент k на графике, достаточно дождаться успокоения сигнала на выходе системы и вычислить отношение уровня выходного сигнала к уровню входно-

го. Математически это означает, что все слагаемые, содержащие производные, равны нулю (система успокоилась, движения нет), а оставшееся слагаемое $Y = k \cdot X$ определяет значение k .

Звено первого порядка

Звено первого порядка обладает двумя параметрами: инерционностью T и коэффициентом усиления $k = Y(t = \infty)/X$.

Чем больше производных учитывается в записи модели, тем со звеном большего порядка мы имеем дело, тем больше коэффициентов при производных следует определить.

Введем понятие передаточной функции как модели динамической системы. По определению передаточная функция - это отношение выхода ко входу:

$$W = Y/X.$$

Передаточная функция звена первого порядка имеет вид:

$$W = k/(Tp + 1),$$

где « p » — символ дифференцирования, тождественно равный « d/dt ». Символ « p » также называется **алгебраизованным оператором дифференцирования**. Тогда, используя определение передаточной функции, имеем:

$$Y/X = k/(Tp + 1).$$

Далее получим:

$$(Tp + 1) \cdot Y = k \cdot X$$

или

$$T \cdot dY/dt + Y = k \cdot X$$

или

$$T \cdot \Delta Y/\Delta t + Y = k \cdot X.$$

В разностном виде уравнение можно записать как

$$T \cdot (Y_{i+1} - Y_i) + Y_i \cdot \Delta t = k \cdot X_i \cdot \Delta t.$$

Или, выразив настоящее через прошедшее:

$$Y_{i+1} = A \cdot X_i + B \cdot Y_i.$$

Здесь $A = k \cdot \Delta t/T$ и $B = 1 - \Delta t/T$ — весовые коэффициенты. A указывает на вес компоненты X , определяющей влияние внешнего мира на систему, B указывает на вес компоненты Y , определяющей память системы, влияние на ее поведение истории.

В частности, если $B = 0$, то $Y_{i+1} = A \cdot X_i$, и мы имеем дело с безынерционной системой $Y = k \cdot X$, мгновенно реагирующей на входной сигнал и увеличивающей его в k раз.

Если коэффициент $B = 0.5$, то есть $1 - \Delta t/T = 0.5$ или $\Delta t/T = 0.5$, то получаем, что коэффициент $A = k \cdot \Delta t/T = k \cdot 0.5$ и, следовательно, $Y_{i+1} = 0.5 \cdot k \cdot X_i + 0.5 \cdot Y_i$. При постоянном (единичном) входном сигнале X будет получен график, как на рис. 4.5.

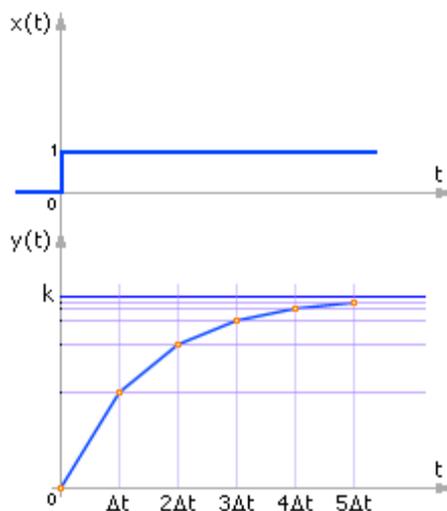


Рис. 4.5. Реакция звена первого порядка на единичный входной сигнал для дискретного случая

Экспонента, изображенная на графике, при большом n (в пределе $n = \infty$) стремится к значению входного (единичного) сигнала X , умноженного на коэффициент усиления k , что подтверждается расчетом:

$$Y_{n+1} = 0.5 \cdot k \cdot X_n + 0.5 \cdot Y_n = 0.5 \cdot k \cdot X_n + 0.5 \cdot (0.5 \cdot k \cdot X_{n-1} + 0.5 \cdot Y_{n-1}) = \dots = (0.5^1 + 0.5^2 + \dots + 0.5^{n+1}) \cdot k \cdot X_0 + 0.5^{n+1} \cdot Y_0 = 1 \cdot k \cdot X_0.$$

Напомним, что выражение $(0.5^1 + 0.5^2 + \dots + 0.5^{n+1})$ является геометрической прогрессией, сумма которой при $n = \infty$ равна 1. А стоящее при Y_0 выражение 0.5^{n+1} обращается в 0 при $n = \infty$.

Если еще усилить влияние прошлого ($B = 1$), то система начнет интегрировать саму себя (выход подан на вход системы), добавляя все время входной сигнал, что соответствует экспоненциальному неограниченному росту выходного сигнала: $Y_{i+1} = A \cdot X_i + Y_i$. По смыслу это соответствует положительной обратной связи. При $B = -1$ имеем модель: $Y_{i+1} = A \cdot X_i - Y_i$, по смыслу соответствующую отрицательной обратной связи. При определении модели требуется найти неизвестные коэффициенты k и T .

Звено второго порядка (колебательное звено)

Такие звенья описываются дифференциальным уравнением вида:

$$a_0 \frac{d^2 y}{dt^2} + a_1 \frac{dy}{dt} + a_2 y = b U$$

Если на вход звена подать единичную функцию Хэвисайда от времени $1[t]$, при нулевых начальных условиях системы, то реакция на выходе будет называться **переходной функцией (или переходной характеристикой)**, которую часто обозначают как $h(t)$. Сигнал $1[t]$ — это, в некотором смысле, эталонный испытательный сигнал. Существуют и другие эталонные испытательные сигналы. Например, бесконечный импульс нулевой длины (дельта-функция Дирака), гармонический сигнал, периодические прямоугольные импульсы.

Преобразуем по Лапласу это уравнение:

$$a_0 \cdot p^2 \cdot Y(p) + a_1 \cdot p \cdot Y(p) + a_2 \cdot Y(p) = b \cdot U(p)$$

или, иначе:

$$(a_0 \cdot p^2 + a_1 \cdot p + a_2) \cdot Y(p) = b \cdot U(p).$$

Определим передаточную функцию звена:

$$W(p) = \frac{Y(p)}{U(p)} = \frac{b}{a_0 p^2 + a_1 p + a_2} = \frac{b/a_2}{\frac{a_0}{a_2} p^2 + \frac{a_1}{a_2} p + 1} = \frac{k}{T^2 p^2 + 2\xi T p + 1}$$

Если записать уравнение без входного воздействия (нулевые входные воздействия $U = 0$) и сократить Y , то есть: $T^2 p^2 + 2\xi T p + 1 = 0$, то такое уравнение будет называться **характеристическим**, поскольку характеризует исключительно внутренние свойства звена. Обратите внимание, что в записи звена содержатся три параметра:

$$T^2 = \frac{a_0}{a_2}, \quad 2\xi T = \frac{a_1}{a_2}, \quad k = \frac{b}{a_2}$$

T — постоянная времени (в секундах);

ξ — коэффициент затухания (безразмерная величина);

k — передаточный коэффициент.

В зависимости от величины ξ звенья второго порядка классифицируются по видам:

- $\xi = 0$ — консервативное звено второго порядка;
- $0 < \xi < 1$ — колебательное звено второго порядка;
- $\xi \geq 1$ — апериодическое звено второго порядка.

Апериодическое звено 2-го порядка ($\xi \geq 1$)

Характеристическое уравнение звена следующее:

$$T^2 p^2 + 2\xi T p + 1 = 0.$$

И оно имеет действительные отрицательные корни:

$$p_{1,2} = \frac{-\xi \pm \sqrt{\xi^2 - 1}}{T}$$

Данное звено можно представить в виде последовательно соединенных звеньев с различными постоянными времени:

$$W(p) = \frac{k}{(T_1 p + 1)(T_2 p + 1)}$$

Тогда при $T_1 > T_2$ переходная характеристика звена имеет вид:

$$h(t) = k \left(1 - \frac{T_1}{T_1 - T_2} e^{-\frac{t}{T_1}} + \frac{T_2}{T_1 - T_2} e^{-\frac{t}{T_2}} \right)$$

То есть в решении присутствуют затухающие экспоненты. Типичное поведение звена с такими параметрами показано на **рис. 4.6**.

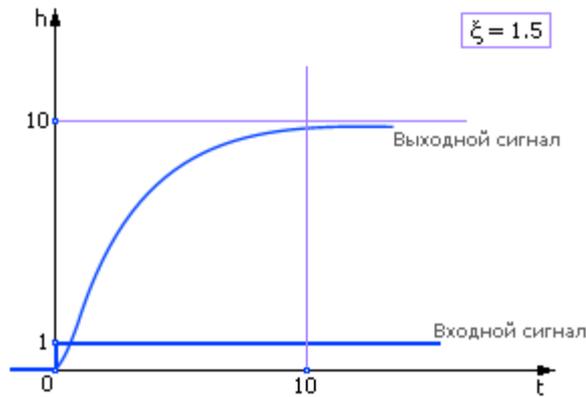


Рис. 4.6. Реакция аperiodического звена на единичный входной сигнал

В частном случае, когда $\zeta = 1$, оба корня будут одинаковыми, отрицательными:

$$p_{1,2} = \frac{-\zeta \pm \sqrt{\zeta^2 - 1}}{T} = -\frac{1}{T}$$

Колебательное звено 2-го порядка ($0 < \zeta < 1$)

Характеристическое уравнение звена следующее:

$$T^2 p^2 + 2\zeta T p + 1 = 0.$$

Корни разные, комплексно-сопряженные, с отрицательной вещественной частью:

$$p_{1,2} = \frac{-\zeta \pm \sqrt{\zeta^2 - 1}}{T} = \frac{-\zeta \pm j\sqrt{1 - \zeta^2}}{T} = a \pm jb$$

, где $a = -\zeta/T$, $b = \sqrt{1 - \zeta^2}/T$.

Так как корни мнимые, то в поведении звена присутствует колебательная составляющая. Именно за эту особенность поведения звено получило название колебательного (см. рис. 4.7 и рис. 4.8).

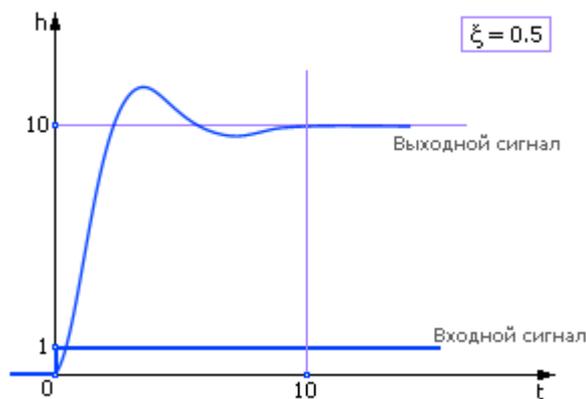


Рис. 4.7. Реакция колебательного звена на входной единичный сигнал ($\zeta = 0.5$)

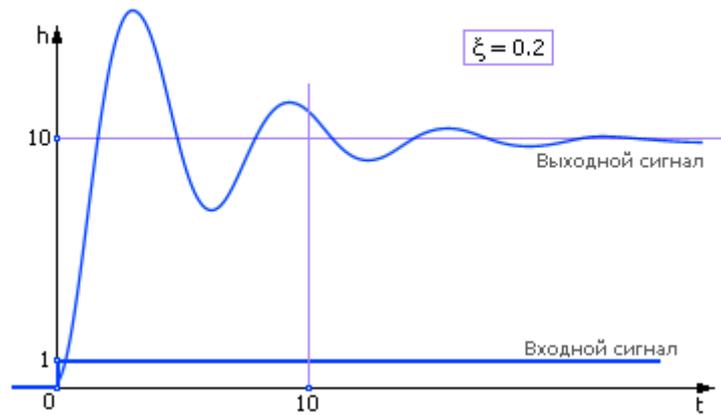


Рис. 4.8. Реакция колебательного звена на входной единичный сигнал ($\zeta = 0.2$)

Из графиков видно, что с ростом ζ колебательность звена уменьшается, исчезая при $\zeta \geq 1$

Переходная функция звена имеет вид:

$$h(t) = k \left[1 - \frac{e^{-\frac{\zeta}{T}t}}{\sqrt{1-\zeta^2}} \sin \left(\frac{\sqrt{1-\zeta^2}}{T} t + \arctg \frac{\sqrt{1-\zeta^2}}{\zeta} \right) \right] = k [1 - A e^{\alpha t} \sin(\omega_0 t + \varphi)]$$

где

$$A = \frac{1}{\sqrt{1-\zeta^2}}, \quad \omega_0 = \frac{\sqrt{1-\zeta^2}}{T}, \quad \varphi = \arctg \frac{\sqrt{1-\zeta^2}}{\zeta}$$

При малых ζ значение A приближается к 1, а значение φ — к 90° . По физическому смыслу ω_0 представляет собой собственную частоту колебаний.

Консервативное звено 2-го порядка ($\zeta = 0$)

Характеристическое уравнение звена следующее:

$$T^2 p^2 + 1 = 0.$$

Корни одинаковые, комплексно-сопряженные, с нулевой вещественной частью:

$$p_{1,2} = \frac{-\zeta \pm \sqrt{\zeta^2 - 1}}{T} = \frac{-\zeta \pm j\sqrt{1-\zeta^2}}{T} = a \pm jb$$

Так как корни чисто мнимые, то поведением звена являются незатухающие колебания ($\zeta = 0$), см. рис. 4.9.

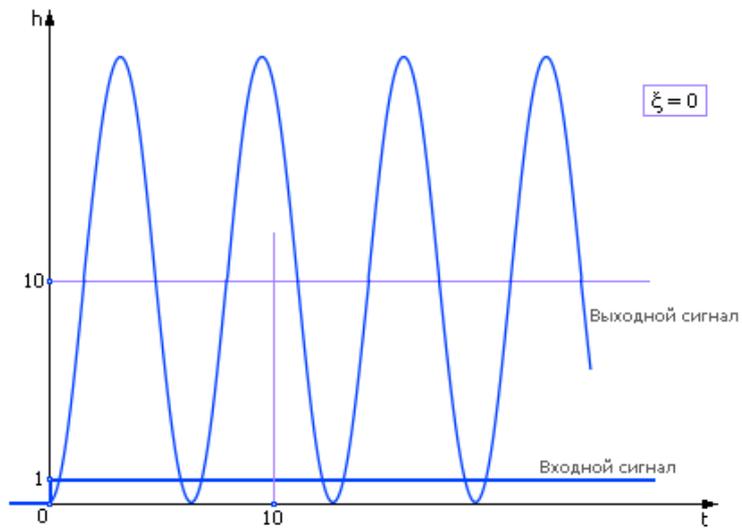


Рис. 4.9. Реакция колебательного звена на входной единичный сигнал ($\xi = 0$)

Переходная функция звена имеет вид: $h(t) = k \cdot (1 - \cos(t/T))$.

Из графика экспериментальным путем можно определить единственный параметр $T = T_0 / (2 \cdot \pi)$.

5. ДИНАМИЧЕСКИЕ РЕГРЕССИОННЫЕ МОДЕЛИ, ЗАДАННЫЕ В ВИДЕ ПЕРЕДАТОЧНОЙ ФУНКЦИИ

Построим регрессионную модель динамической системы на примере. Зададим модель в виде передаточной функции (см. рис.5.1).

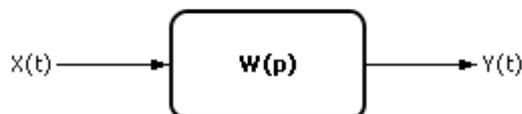


Рис. 5.1. Модель черного ящика в виде передаточной функции

Примерный вид динамических сигналов на входе и на выходе показан на рис.5.2. Ограничимся временем рассмотрения сигналов, равным T .

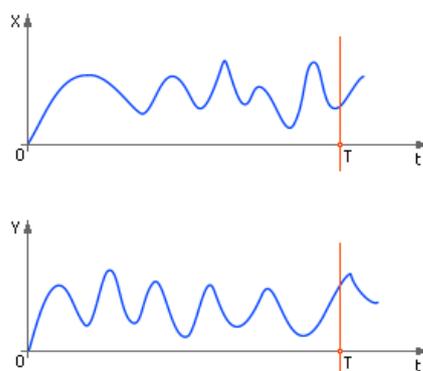


Рис. 5.2. Вид зависимости входного сигнала X от времени t и зависимости выходного сигнала Y от t для случая с непрерывным временем

После дискретизации, связанной с обработкой информации на цифровых машинах, эти сигналы будут выглядеть так, как показано на рис. 5.3. Обратите внимание на то, что отдельные отсчеты отстоят друг от друга на расстоянии Δt . Важно, что отсчеты стоят достаточно часто. Всего этих отсчетов — n , то есть $T = n \cdot \Delta t$,

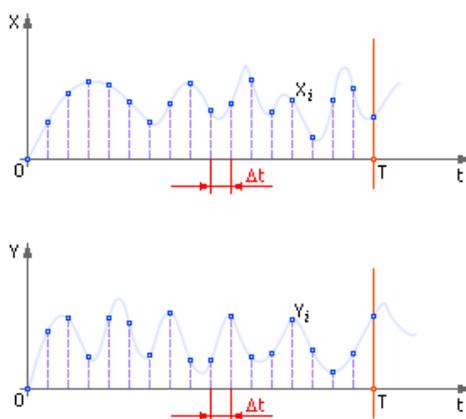


Рис. 5.3. Вид зависимости входного сигнала X от времени t и зависимости выходного сигнала Y от t для случая с дискретным временем

Допустим, что зависимости, представленные на рис.5.2 и рис.5.3, описываются передаточной функцией следующего вида (заметим, что, как и в лекции 02, вид зависимости выдвигается гипотетически и гипотеза должна быть, в конце концов, подтверждена или опровергнута) [9]:

$$W = \frac{A_1 + A_2 p}{p^2 + A_3 p + A_4}$$

Заменяя значок « p » на « d/dt » (обращаем ваше внимание, что такую замену можно производить только для случая нулевых начальных условий: $X(0) = 0$ и $Y(0) = 0$) и учитывая, что передаточная функция — это, по определению, отношение выхода к входу, то есть $W = Y/X$, получаем дифференциальное уравнение 2-го порядка:

$$\frac{d^2 Y}{dt^2} + A_3 \frac{dY}{dt} + A_4 Y = A_1 X + A_2 \frac{dX}{dt}$$

Дважды проинтегрируем это выражение и получим для некоторого произвольного момента времени t :

$$Y(t) + A_3 \int_0^t Y(\tau) d\tau + A_4 \int_0^t \int_0^\tau Y(\tau) d\tau dt = A_1 \int_0^t \int_0^\tau X(\tau) d\tau dt + A_2 \int_0^t X(\tau) dt$$

Коэффициенты A_1, A_2, A_3, A_4 требуется определить. Для этого выразим уравнение в разностном виде через суммы:

$$Y_m + A_3 \sum_{i=1}^m Y_i \Delta t_i + A_4 \sum_{j=1}^m \sum_{i=1}^j Y_i \Delta t_i \Delta \tau_j = A_1 \sum_{j=1}^m \sum_{i=1}^j X_i \Delta t_i \Delta \tau_j + A_2 \sum_{i=1}^m X_i \Delta t_i, \quad m = 1, \dots, n$$

где n — число экспериментальных точек. Заметим, что мы заменили интегралы и непрерывное течение времени — на суммы и дискретное представление времени. Чтобы вычислить суммы и двойные суммы экспериментально заданных зависимостей x и y , воспользуемся для удобства табл. 5.1 (чтобы излишне не загромождать таблицу, мы не стали дописывать в суммах Δt_i и $\Delta \tau_j$).

Таблица 5.1.
Таблица исходных данных и вспомогательных расчетов

i	X_i	Y_i	$\sum_{i=1}^m X_i$	$\sum_{i=1}^m Y_i$	$\sum_{j=1}^m \sum_{i=1}^j X_i$	$\sum_{j=1}^m \sum_{i=1}^j Y_i$
1	X_1	Y_1	X_1	Y_1	X_1	Y_1
2	X_2	Y_2	$X_1 + X_2$	$Y_1 + Y_2$	$2X_1 + X_2$	$2Y_1 + Y_2$
3	X_3	Y_3	$X_1 + X_2 + X_3$	$Y_1 + Y_2 + Y_3$	$3X_1 + 2X_2 + X_3$	$3Y_1 + 2Y_2 + Y_3$
...
m	X_m	Y_m	$X_1 + X_2 + \dots + X_m$	$Y_1 + Y_2 + \dots + Y_m$	$mX_1 + \dots + X_m$	$mY_1 + \dots + Y_m$
...
n	X_n	Y_n	$X_1 + X_2 + \dots + X_n$	$Y_1 + Y_2 + \dots + Y_n$	$nX_1 + \dots + X_n$	$nY_1 + \dots + Y_n$

Ошибку в некоторой m -ой точке можно записать так:

$$E_m = Y_m + A_3 \sum_{i=1}^m Y_i \Delta t_i + A_4 \sum_{j=1}^m \sum_{i=1}^j Y_i \Delta t_i \Delta \tau_j - A_1 \sum_{j=1}^m \sum_{i=1}^j X_i \Delta t_i \Delta \tau_j - A_2 \sum_{i=1}^m X_i \Delta t_i$$

Как и ранее, ошибка показывает, насколько отходит теоретическое значение Y_m от экспериментального значения.

Суммарная ошибка (вносимая всеми точками), которую надо минимизировать, будет:

$$F = \sum_{m=1}^n E_m^2 \Rightarrow \min_{A_1, A_2, A_3, A_4}$$

Величина ошибки зависит от значений параметров A_1, A_2, A_3, A_4 . Поэтому F является функцией от четырех переменных: $F(A_1, A_2, A_3, A_4)$. Чтобы найти минимум функции F , доставляемый за счет параметров A_1, A_2, A_3, A_4 , надо взять частные производные F по каждому из параметров и приравнять каждую производную к нулю. В результате получаем систему линейных уравнений:

$$\begin{cases} \frac{\partial F}{\partial A_1} = 0 \\ \frac{\partial F}{\partial A_2} = 0 \\ \frac{\partial F}{\partial A_3} = 0 \\ \frac{\partial F}{\partial A_4} = 0 \end{cases}$$

Получено четыре уравнения с четырьмя неизвестными A_1, A_2, A_3, A_4 . Из решения системы уравнений вычисляем неизвестные коэффициенты и дополняем ими модель, где коэффициенты уже определены как числа:

$$\frac{d^2 Y}{dt^2} + A_3 \frac{dY}{dt} + A_4 Y = A_1 X + A_2 \frac{dX}{dt}$$

Задача определения коэффициентов модели решена. Разумеется, как и ранее, необходимо сравнить получаемое из этой модели решение Y теоретическое с Y , заданным экспериментально, и вычислить ошибку F . Далее проверить ее значение по критерию - допустимо ли значение вычисленной ошибки, или гипотезу о виде модели требуется сменить на более точную.

Считая что коэффициенты модели теперь нам известны, построим для заданного примера реализацию, имитирующую поведение системы, описанной передаточной функцией. Для этого воспользуемся уже однажды полученной формулой:

$$Y_m + A_3 \sum_{i=1}^m Y_i \Delta t_i + A_4 \sum_{j=1}^m \sum_{i=1}^j Y_i \Delta t_i \Delta \tau_j = A_1 \sum_{j=1}^m \sum_{i=1}^j X_i \Delta t_i \Delta \tau_j + A_2 \sum_{i=1}^m X_i \Delta t_i, \quad m = 1, \dots, n$$

или

$$Y_m = \underbrace{A_1 \sum_{j=1}^m \sum_{i=1}^j X_i \Delta t_i \Delta \tau_j}_w + \underbrace{A_2 \sum_{i=1}^m X_i \Delta t_i}_v - \underbrace{A_3 \sum_{i=1}^m Y_i \Delta t_i}_m - \underbrace{A_4 \sum_{j=1}^m \sum_{i=1}^j Y_i \Delta t_i \Delta \tau_j}_n, \quad m = 1, \dots, n$$

Реализация модели представлена на рис. 5.4.

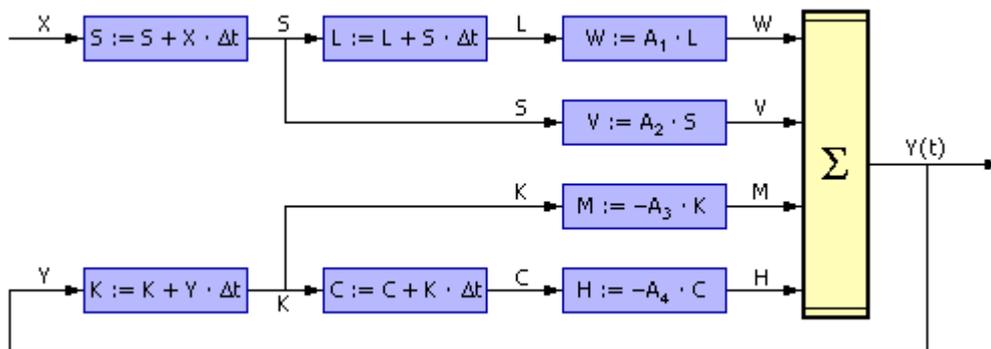


Рис. 5.4. Реализация передаточного звена после определения коэффициентов регрессионной модели
 При переходе от интеграла к численному суммированию мы воспользовались методом прямоугольников. Разбив площадь под кривой y на ряд прямоугольников одинаковой ширины Δt (см. рис. 5.5), получаем, что площадь i -го прямоугольника равна $y_i \cdot \Delta t$, а S - сумма площадей всех n прямоугольников - будет *приблизительно* равна площади под кривой (интегралу от функции y). Очевидно, что приближение тем точнее, чем меньше значение Δt .

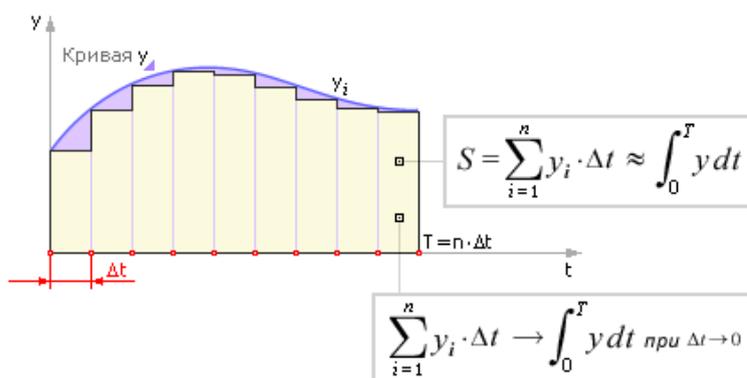


Рис. 5.5. Применение метода прямоугольников для численного вычисления интегралов

6. МОДЕЛЬ В ВИДЕ ФИЛЬТРА КАЛЛМАНА

Каллманом была доказана теорема о том, что любой динамический сигнал может быть представлен в виде:

$$Y_i = A_1 \cdot X_i + A_2 \cdot X_{i-1} + \dots + B_1 \cdot Y_{i-1} + B_2 \cdot Y_{i-2} + \dots + C.$$



Рис. 6.1. Графическое представление фильтра Каллмана на схемах

Идея фильтра Каллмана заключается в том, что выход системы в i -ый момент времени определяется входным сигналом, его предысторией и предысторией самого состояния системы [9].

Чем больше имеется членов ряда, то есть чем больше переменных Y учитывается в записи модели, тем глубже память системы. Заметим, что наличие члена Y_{i-1} в модели динамической системы соответствует наличию первой производной, Y_{i-2} — второй производной и т. д.

Допустим, известны следующие экспериментальные данные: состояния сигналов X_i и Y_i в n временных точках (табл. 6.1).

Таблица 6.1.
Таблица экспериментальных данных

i	X_i	Y_i
1	X_1	Y_1
2	X_2	Y_2
...
$n-1$	X_{n-1}	Y_{n-1}
n	X_n	Y_n

Поскольку для каждой экспериментальной точки X_i надо указать ее соседей, задаваемых рядом, то удобно отсчеты представить в расширенной таблице, используемой для расчета (см. табл. 6.2).

Таблица 6.2.
Таблица экспериментальных данных и промежуточных расчетов

i	X_i	X_{i-1}	...	Y_i	Y_{i-1}	Y_{i-2}
m	X_m	X_{m-1}	...	Y_m	Y_{m-1}	Y_{m-2}
$m+1$	X_{m+1}	X_m	...	Y_{m+1}	Y_m	Y_{m-1}
$m+2$	X_{m+2}	X_{m+1}	...	Y_{m+2}	Y_{m+1}	Y_m
...

Находим ошибку между значением экспериментально снятой точки и теоретическим ее значением (гипотезой):

$$E_m = Y_m - A_1 \cdot X_m - A_2 \cdot X_{m-1} - \dots - B_1 \cdot Y_{m-1} - B_2 \cdot Y_{m-2} - \dots - C.$$

Суммарная ошибка F (сумма берется по всем экспериментальным точкам) должна быть минимизирована относительно определяемых переменных $A_1, A_2, \dots, B_1, B_2, \dots, C$:

$$F = \sum_{i=1}^m E_i^2 \Rightarrow \min_{A_1, A_2, \dots, B_1, B_2, \dots, C}$$

После взятия частных производных от F по $A_1, A_2, \dots, B_1, B_2, \dots, C$, приравнивания их к нулю и составления системы уравнений получается линейная множественная регрессионная модель, из которой определяются неизвестные коэффициенты $A_1, A_2, \dots, B_1, B_2, \dots, C$ модели.

Поскольку коэффициенты модели определены, построим реализацию (см. рис. 6.2), имитирующую поведение системы, описанной фильтром Калмана.

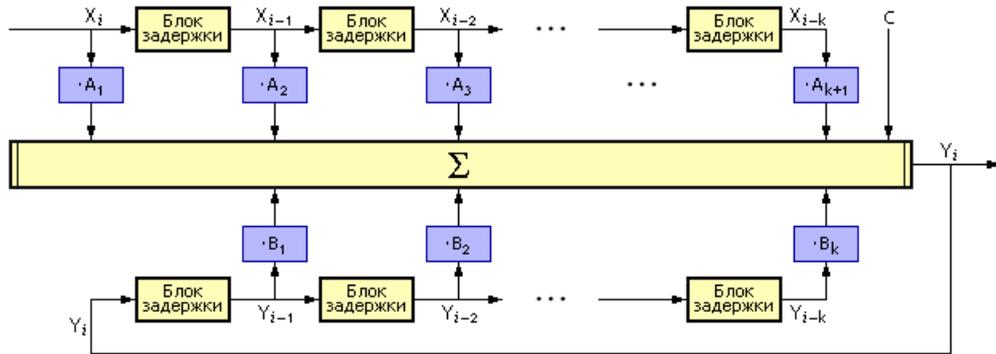


Рис. 6.2. Вариант технической реализации фильтра Калмана

«Блок задержки» в представленной реализации необходим для того, чтобы сдвинуть сигнал на такт и получить соседний отсчет для следующей переменной ряда модели. В случае реализации блока задержки, первый способ может быть основан на перезаписи информации из одной переменной (ячейки) в другую, на что требуется один такт. Таким образом, можно организовать задержку сигнала на любое число тактов. Например, задержка сигнала X относительно Y будет составлять 3 такта, если выполнить следующую последовательность операций: $A1 := X; A2 := A1; Y := A2$.

Во втором способе задержка организуется при помощи массива: на каждом такте нужно, чтобы цифры были перемещены в соседние ячейки.

На рис. 6.3 приведена схема настройки (автоматического нахождения коэффициентов).

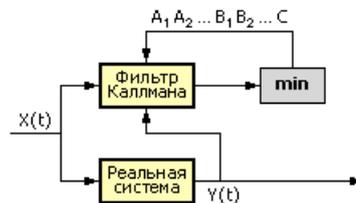


Рис. 6.3. Схема автоматической настройки коэффициентов модели «на ходу»

На рис. 6.4 приведена схема проверки фильтра Калмана.

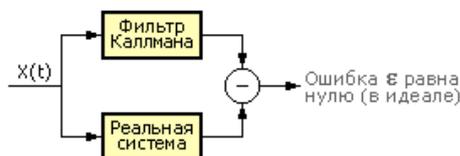


Рис. 6.4. Схема проверки работы модели фильтра Калмана

7. МОДЕЛЬ ДИНАМИЧЕСКОЙ СИСТЕМЫ В ВИДЕ ФУРЬЕ ПРЕДСТАВЛЕНИЯ (МОДЕЛЬ СИГНАЛА)

Этот способ моделирования динамических систем основывается на том, что в любом сигнале присутствуют гармонические составляющие. В зависимости от частоты, составляющие называются **гармониками** (первая, вторая и так далее). Сумма гармоник с соответствующими весами составляет **модель сигнала**.

В некотором сигнале присутствует сумма трех гармоник: $3 \cdot \cos(t) + 2 \cdot \cos(3t) + 0.5 \cdot \cos(5t)$. Это значит, что в сигнале присутствует первая гармоника с амплитудой 3, третья гармоника с амплитудой 2, пятая гармоника с амплитудой 0.5. Сам суммарный сигнал выглядит так, как показано на рис. 7.1.

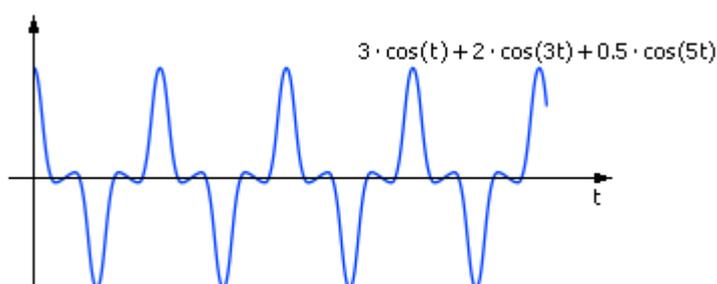


Рис. 7.1. Пример гармонического сигнала

Спектр этого сигнала показан на рис. 7.2. Ясно, что в нашем примере больший вес (амплитуду) в сигнале имеет (более других представлена) первая гармоника, наименьший вес имеет пятая гармоника [9].

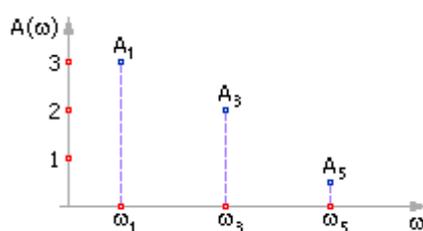


Рис. 7.2. Спектр гармонического сигнала

Любой сигнал, сколь сложен бы он ни был, может быть представлен суммой гармоник. Более простой сигнал представляется меньшим числом гармоник, более сложный - большим. Быстро меняющийся сигнал, содержащий резкие пики, имеет в своем составе гармоники высоких порядков. Чем больше гармоник представлено в модели сигнала, тем точнее, в общем случае, модель отражает реальный сигнал.

Пусть задан некий сигнал $X(t)$ (рис. 7.3).

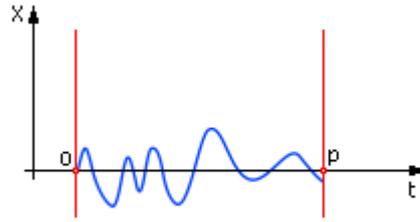


Рис. 7.3. Временной сигнал на входе преобразования Фурье

Определимся со временем рассмотрения сигнала: если сигнал *периодический*, то время рассмотрения равно периоду p сигнала; если сигнал *непериодический*, то периодом сигнала считается все время его рассмотрения.

$$A_0 = \frac{2}{p} \int_0^p X(t) dt$$

$$A_1 = \frac{2}{p} \int_0^p X(t) \cos\left(\frac{2\pi t}{p}\right) dt \quad B_1 = \frac{2}{p} \int_0^p X(t) \sin\left(\frac{2\pi t}{p}\right) dt$$

$$A_2 = \frac{2}{p} \int_0^p X(t) \cos\left(\frac{2\pi 2t}{p}\right) dt \quad B_2 = \frac{2}{p} \int_0^p X(t) \sin\left(\frac{2\pi 2t}{p}\right) dt$$

...

...

$$A_i = \frac{2}{p} \int_0^p X(t) \cos\left(\frac{2\pi i t}{p}\right) dt \quad B_i = \frac{2}{p} \int_0^p X(t) \sin\left(\frac{2\pi i t}{p}\right) dt$$

...

...

A_i и B_i — это веса соответствующих гармоник, присутствующих в сигнале; i — номер гармоники. Формулы их расчета называются **прямым преобразованием Фурье**.

Значение $2\pi \cdot i/p = \omega_i$ — это частота i -ой гармоники. Отметим также, что частота i -ой гармоники связана с частотой первой гармоники простым соотношением: $\omega_i = i \cdot \omega_1$.

Отметим важную особенность данного способа представления: вместо всего сигнала во всех его подробностях достаточно хранить вектор чисел, представляющих весовые коэффициенты составляющих его гармоник: $(A_0, A_1, A_2, \dots, B_1, B_2, \dots)$. То есть эти числа полностью характеризуют исходный сигнал, так как по ним сигнал можно полностью восстановить формулой **обратного преобразования Фурье**:

$$x(t) = \frac{A_0}{2} + A_1 \cos\left(\frac{2\pi t}{p}\right) + B_1 \sin\left(\frac{2\pi t}{p}\right) + A_2 \cos\left(\frac{2\pi 2t}{p}\right) + B_2 \sin\left(\frac{2\pi 2t}{p}\right) + \dots$$

$$+ A_i \cos\left(\frac{2\pi i t}{p}\right) + B_i \sin\left(\frac{2\pi i t}{p}\right) + \dots$$

Именно эти числа используются также при обработке сигнала в модели динамической системы. Изображение этих чисел на графике в зависимости от номера гармоники (частоты) называется **спектром сигнала** (рис. 7.4). Спектр показывает, насколько присутствует в сигнале соответствующая составляющая. Спектр - это частотная характеристика сигнала.

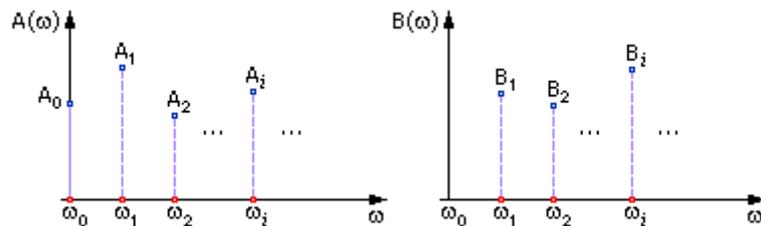


Рис. 7.4. Сигнал, представленный в частотной области на выходе преобразования Фурье, спектр сигнала

Здесь сигнал представлен в частотной области. Всегда по формулам прямого преобразования Фурье можно перейти из временной области в частотную, а по формулам обратного преобразования Фурье перейти из частотной области во временную. В какой области (частотной или временной) работать с сигналом в отдельный момент, решают из соображений удобства, наглядности и экономии вычислений. Заметим, что емкие с точки зрения вычислений операции интегрирования и дифференцирования сигнала во временной области заменяются на операции алгебраического сложения и умножения в частотной области, что с вычислительной точки зрения реализуется намного точнее и быстрее [9].

Система чисел A_i и B_i является **полной характеристикой сигнала**. Такой же полной характеристикой сигнала является система чисел S и φ , которые также образуют спектр (рис. 7.5). S — это амплитудно-частотная характеристика (АЧХ), φ — фазо-частотная характеристика (ФЧХ).

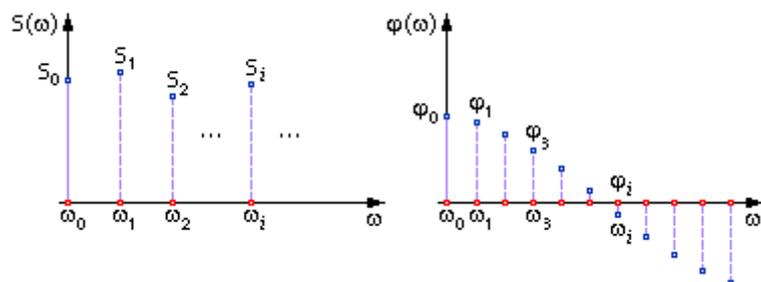


Рис. 7.5. Сигнал, представленный в частотной области, амплитудно-частотная и фазо-частотная характеристика сигнала (возможный вид)

Системы « A и B » и « S и φ » являются полностью равнозначными. Переход из системы « A и B » в систему « S и φ » производится по следующим формулам: $S_i = \sqrt{A_i^2 + B_i^2}$ — абсолютная амплитуда сигнала; $\varphi_i = \arctg(B_i/A_i)$ —

фаза сигнала, при сложении гармоник нужно учитывать сдвиг фаз (сдвиг фаз проиллюстрирован на рис. 7.8).

В случае с системой « S и φ » обратное преобразование Фурье имеет вид:

$$x(t) = \sum_{i=1}^k S_i \cdot \sin(\omega_i t + \varphi_i)$$

Рис. 7.6 и рис. 7.7 разъясняют смысл коэффициентов A и B разных гармоник. Эти коэффициенты — амплитуды синусов и косинусов соответствующих частот (гармоник). Во временной области графически они соответствуют размаху гармонических колебаний (рис. 7.6 и рис. 7.7); в частотной — высоте спектральной полосы на соответствующей частоте (рис. 7.4).

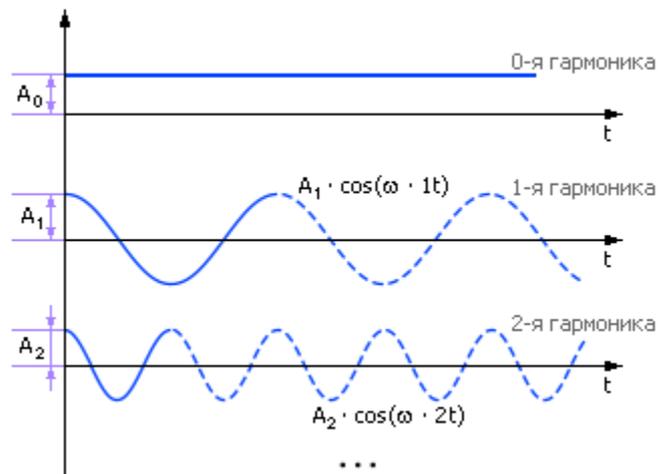


Рис. 7.6. Геометрическая иллюстрация параметров A и ω для косинусной составляющей гармонического сигнала

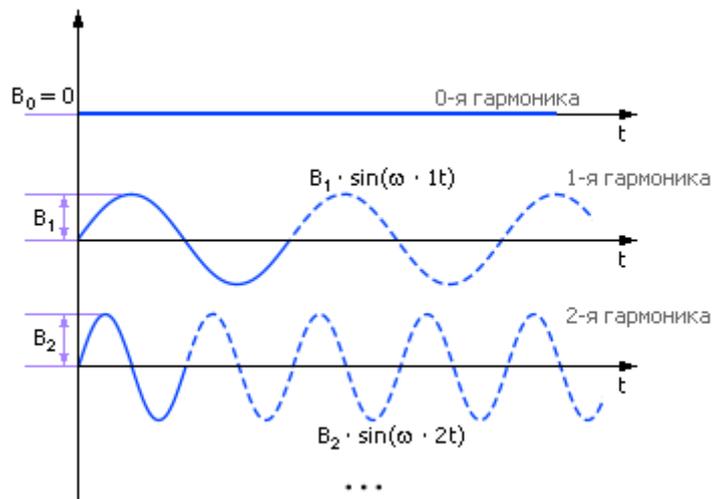


Рис. 7.7. Геометрическая иллюстрация параметров B и ω для синусной составляющей гармонического сигнала

Смысл чисел S_i и φ_i разъяснен на рис. 7.8.

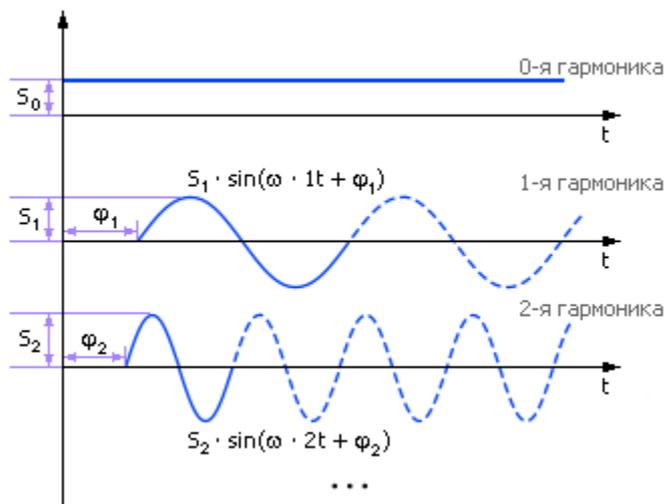


Рис. 7.8. Геометрическая иллюстрация параметров S и φ для составляющей гармонического сигнала

8. МОДЕЛЬ ДИНАМИЧЕСКОЙ СИСТЕМЫ В ВИДЕ ФУРЬЕ ПРЕДСТАВЛЕНИЯ (МОДЕЛЬ ОБЪЕКТА)

Пусть имеется входной динамический сигнал $X(t)$ и объект F , преобразующий этот сигнал в выходной $Y(t)$ (см. рис. 8.1). Если объект описывается дифференциальными уравнениями, то таким преобразованием является интегрирование входного сигнала и вычисление $Y(t)$. Интегрирование, как было ранее показано, - операция, требующая значительных вычислительных ресурсов и имеющая значительную погрешность при реализации на цифровых машинах [9].

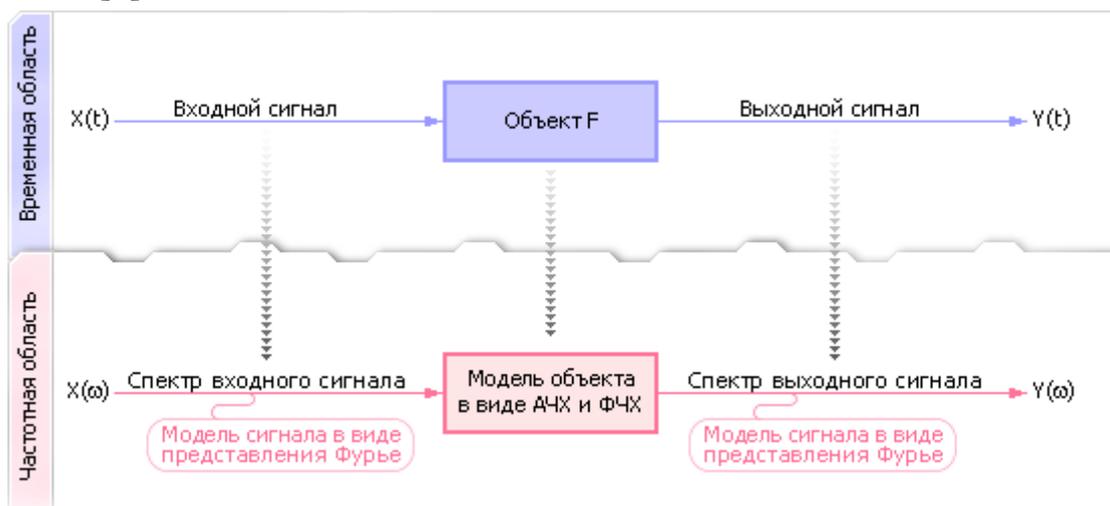


Рис. 8.1. Схема моделирования динамического объекта при переходе из временной области представления в частотную

Если перейти от описания входного сигнала во временной области к описанию в частотной области (см. рис. 8.1), а от дифференциальных уравнений перейти к частотной характеристике объекта, - то есть, фактически, заменить сигнал на частотную модель сигнала, а объект на частотную модель объекта, - то с вычислительной точки зрения процесс преобразования сигнала упростится. Конечно, полученный результат тоже будет частотной моделью выходного сигнала, которую для получения окончательного ответа придется сконвертировать во временную область $Y(t)$. Процесс такой конвертации из частотной области во временную и обратно называется **преобразованием Фурье** (есть и другие преобразования). Для тех объектов, для которых известна их модель в частотной области, такой подход достаточно просто реализуется на компьютере и позволяет достичь любой наперед заданной точности.

Модель объекта в частотном виде называется **передаточной функцией** или **АЧХ** (амплитудно-частотной характеристикой). Объекты, для которых известны АЧХ, обычно называют **типовыми звеньями** (усилительное звено, апериодическое, колебательное и т.д.). Пусть, для примера, характеристика объекта в частотной области следующая (см. рис.8.2).

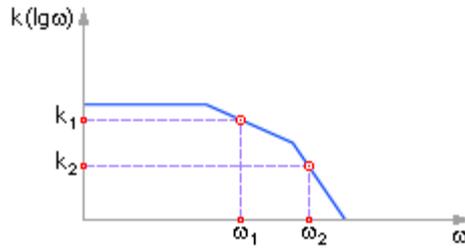


Рис. 8.2. АЧХ (возможный вид)

Амплитудно-частотная характеристика (АЧХ) показывает, насколько пропускается объектом на выход соответствующая гармоника. Значение k_i характеризует коэффициент усиления гармонического сигнала на определенной частоте ω_i .

Моделирование прохождения сигнала через объект в этом виде заключается в умножении коэффициента A_i гармоники с частотой ω_i входного сигнала $X(t)$ на коэффициент усиления k_i при той же гармонике с частотой ω_i в АЧХ: $A_i^* = A_i(\omega_i) \cdot k_i(\omega_i)$. (Для коэффициента B преобразование аналогично.) В результате получается коэффициент A_i^* выходной гармоники данной частоты ω_i . Процедура выполняется для всех частот, представленных во входном сигнале и АЧХ. После получения спектра выходного сигнала можно восстановить сигнал как временную зависимость с помощью формулы обратного преобразования Фурье.

Моделирование прохождения сигнала через динамический объект свелось к операции умножения двух переменных, точнее, к операции поэлементного умножения вектора одних переменных на вектор других переменных.

Схема преобразования показана на рис. 8.3.

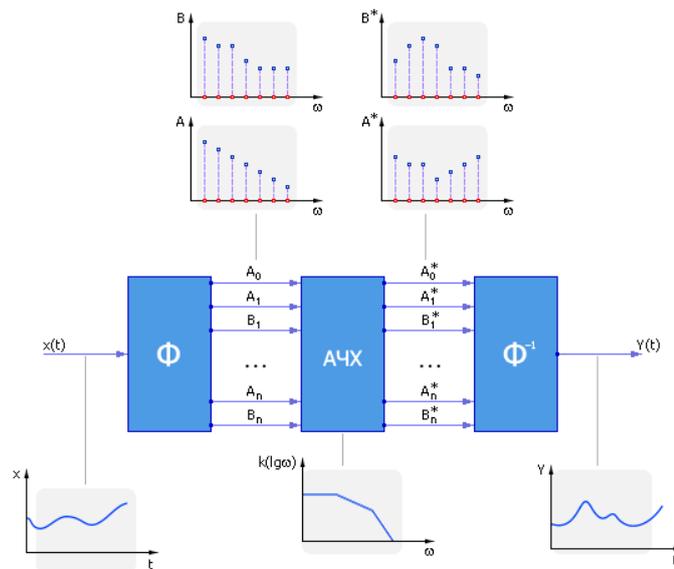


Рис. 8.3. Преобразование сигнала при использовании метода Фурье

Если бы временной сигнал проходил через звено, которое во временной области представлено дифференциальным уравнением, то пришлось бы его

интегрировать, что, конечно, приводит к погрешностям результата. В частотной области достаточно перемножить значения коэффициентов ряда Фурье сигнала и звена при одинаковых частотах. Очевидно, что достоинством метода является замена дифференциальных уравнений модели на алгебраические. **Разумеется, данный подход может быть использован только для объектов, у которых известен вид передаточной функции.** (Кстати, для неизвестных случаев АЧХ может быть получена **численным разложением в ряд.**)

В процессе моделирования набора объектов для преобразования сигнала (например, протяженных трактов радиоэлектронных устройств) иногда приходится применять прямое и обратное преобразование Фурье неоднократно. На практике последовательные блоки часто называют **каскадами**.

Пусть мы имеем радио-электронное устройство (РЭУ), состоящее из 5 блоков (см. рис. 8.4). Блоки 1, 2, 4, 5 — линейные и представлены соответствующими известными АЧХ; блок 3 — нелинейный, поэтому АЧХ для него неизвестна. Примером линейного блока может служить апериодическое звено, колебательное звено и т. д. (см. Лекцию 05. Динамические регрессионные модели, заданные в виде передаточной функции). Примером нелинейного блока может служить устройство ограничения сигнала (срез) по амплитуде.

Как видно из рис. 8.4, сначала входной сигнал $X(t)$ прямым преобразованием Фурье переводится в частотную область и проходит в виде спектра через АЧХ 1 и 2 *линейного* блока, затем обратным преобразованием Фурье сигнал после 2 блока переводится во временную область. Проходим *нелинейный* блок 3 во временном представлении. Результат работы блока 3 снова преобразуем прямым преобразованием Фурье в частотную область и проходим через АЧХ блоков 4 и 5. В конце полученный спектр преобразуется с помощью обратного преобразования Фурье во временную область, — вид сигнала, $Z(t)$, является результатом моделирования.

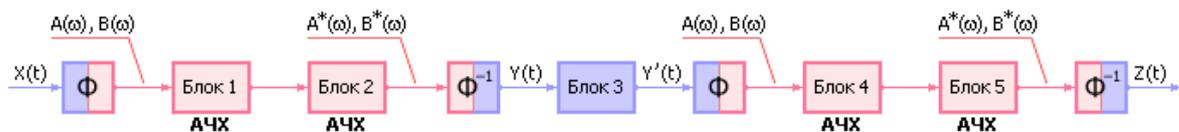


Рис. 8.4. Моделирование тракта, содержащего нелинейные блоки, с использованием метода Фурье

Метод, который мы рассмотрели, является одним из самых быстродействующих. Это связано с заменой операций интегрирования и дифференцирования, встречающихся в моделях динамических звеньев, на операции сложения и умножения при переходе в частотную область. Такая процедура обеспечивает точность и быстродействие модели [9].

Для метода важно, с какой частотой вы дискретизируете сигнал при разложении в ряд Фурье. Если частота дискретизации мала, то есть отсчеты в сигнале следуют редко, с большими интервалами, то часть сигнала остается потерянной, так как между отсчетами может оказаться резко возросший и

опавший пик, информация о котором пропадет. То есть говорят, что малая частота дискретизации срезает высокие частоты в сигнале. (Пик — это и есть высокочастотная составляющая, которая может быть потеряна).

По теореме Котельникова, чтобы не потерять соответствующую гармонику, требуется дискретизировать сигнал с частотой не менее чем в 2 раза большей, чем самая высокая частота из представленных в аналоговом сигнале:

$$2W_{\max} \leq W_{\text{дискр.}}$$

где $W_{\text{дискр.}} = 1/\Delta t_{\text{дискр.}}$ — частота дискретизации, W_{\max} — максимальная частота, присутствующая в сигнале.

9. ОЦЕНКА КАЧЕСТВА МОДЕЛИ

Оценка качества показывает, насколько теоретические вычисления по построенной модели отклоняются от экспериментальных данных. Наличие связи двух переменных называется **корреляцией**.

Если оценка качества применяется до исследования, то она решает задачу: есть ли связь между входом X и выходом Y и оценивает силу этой связи [9].

1. Линейный коэффициент корреляции

Линейный коэффициент корреляции указывает, есть ли между двумя рядами X и Y линейная зависимость и какой силы. Вычисляется по следующей формуле:

$$KR = \frac{m_{xy} - m_x m_y}{\sigma_x \sigma_y}$$

m_x , m_y , m_{xy} — математическое ожидание x , y , xy :

$$m_x = \frac{1}{n} \sum_{i=1}^n x_i, m_y = \frac{1}{n} \sum_{i=1}^n y_i, m_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i$$

Дисперсия σ_x^2 и σ_y^2 показывает, насколько разбросаны точки от средней величины:

$$\sigma_x^2 = \sum_{i=1}^n \frac{(x_i - m_x)^2}{n}, \sigma_y^2 = \sum_{i=1}^n \frac{(y_i - m_y)^2}{n}$$

Линейный коэффициент корреляции может иметь знак плюс или минус. Положительная его величина свидетельствует о прямой связи между X и Y . Чем ближе KR к $+1$, тем связь более тесная. Отрицательная величина его свидетельствует об обратной связи; в этом случае границей является -1 . Близость KR к нулю свидетельствует о слабой связи между X и Y (см. рис. 9.1).



Рис. 9.1.

2. Нелинейный коэффициент корреляции

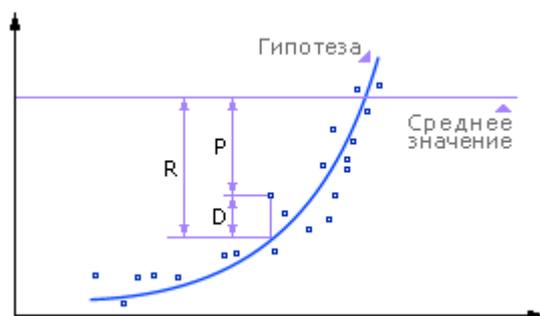


Рис. 9.2.

Нелинейный коэффициент корреляции вычисляется по следующей формуле:

$$\text{KNR} = \sqrt{P - \frac{D}{P}}$$

P — разброс между реальными точками и средней величиной:

$$P = \sum_{i=1}^n \frac{(y_i^P - y_i^C)^2}{n}$$

D — разброс между гипотетической кривой и реальными точками:

$$R = \sum_{i=1}^n \frac{(y_i^{\Gamma} - y_i^C)^2}{n}$$

R — разброс между гипотезой и средней величиной:

$$R = \sum_{i=1}^n \frac{(y_i^{\Gamma} - y_i^C)^2}{n}$$

3. Коэффициент корреляции двух динамических рядов

X и Y представляются в виде рядов z_i и u_i для того, чтобы исключить постоянную составляющую: $z_i = x_i - m_x$

$$u_i = y_i - m_y$$

$$r = \frac{\sum_{i=1}^n z_i u_i}{\sqrt{\sum_{i=1}^n z_i^2 \cdot \sum_{i=1}^n u_i^2}}$$

При $r \rightarrow 1$ имеет место тесная корреляция. При $r \rightarrow 0$ процессы взаимно ортогональны, корреляции нет, процессы не связаны друг с другом.

4. Корреляция внутри динамического ряда

Исследуется сила связи между прошлым и настоящим одного процесса. Для этого сигнал сравнивают с самим собой, сдвинутым во времени, и вычисляют коэффициент корреляции двух динамических рядов (см. п. 3).

5. Поиск периодичности ряда

Есть ли периодичность в динамическом ряду, можно выяснить, проделав прямое преобразование Фурье и рассмотрев спектр исследуемого сигнала. Об

этом рассказывается в лекции «Модель динамической системы в виде Фурье представления (модель сигнала)»

6. Зависимость динамики ряда Z от двух динамических факторов X и Y



Рис. 9.5.

Коэффициент множественной корреляции R:

$$R = \sqrt{\frac{r_{xz}^2 + r_{xy}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{xy}^2}}$$

$$r_{xz} = \frac{m_{xz} - m_x m_z}{\sigma_x \sigma_z} = \frac{m_{xy} - m_x m_y}{\sigma_x \sigma_y}$$

$$r_{yz} = \frac{m_{yz} - m_y m_z}{\sigma_y \sigma_z}$$

7. Связь двух признаков

Формула

$$K = \frac{ad - bc}{ad + bc}$$

где K — это коэффициент ассоциаций, позволяет выяснить, имеется ли какая-либо связь между двумя признаками. Если данный коэффициент близок к единице, то в этом случае можно говорить о существовании такой связи.

Пример. Попробуем с помощью данной формулы выяснить, есть ли связь между ростом и весом человека? Пусть в нашем распоряжении имеются данные о весе и росте 500 человек:

Таблица 9.1.

	Вес < 67 кг.	Вес > 67 кг.
Рост < 167 см.	a = 304 чел.	b = 17 чел.
Рост > 167 см.	c = 112 чел.	d = 67 чел.

По формуле: $K = (304 \cdot 67 - 17 \cdot 112) / (304 \cdot 67 + 17 \cdot 112) = 0.83$. Так как величина 0.83 близка к 1, то можно говорить о существовании определенной связи между весом и ростом.

10. ЧИСЛЕННЫЕ МЕТОДЫ ИНТЕГРИРОВАНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ. МЕТОД ЭЙЛЕРА

Пусть нам известна входная динамическая последовательность X (входной сигнал) и модель (способ преобразования входного сигнала в выходной сигнал). Рассматривается задача определения выходного сигнала $y(t)$ (см. рис. 10.1) [9].



Рис. 10.1. Структурная модель динамической системы с одним входом и одним выходом

Модель динамической системы может быть представлена дифференциальным уравнением. Основное уравнение динамики:

$$y' = f(x(t), y(t), t).$$

Известны начальные условия в нулевой момент времени t_0 : $y(t_0)$, $x(t_0)$. Чтобы определить выходной сигнал, заметим, что по определению производной:

$$y' \cong \operatorname{tg}(\alpha) = \frac{\Delta y}{\Delta t}$$

Нам известно положение системы в точке «1», требуется определить положение системы в точке «2». Точки отделены друг от друга расстоянием Δt (рис. 10.2). То есть расчет поведения системы производится по шагам. Из точки «1» мы скачком (дискретно) переходим в точку «2», расстояние между точками по оси t называется **шагом расчета Δt** .

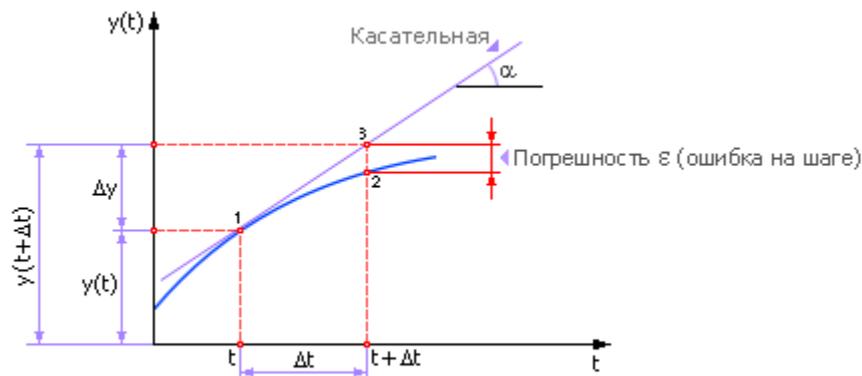


Рис. 10.2. Расчет будущего состояния системы методом Эйлера на одном шаге

Тогда:

$$y(t + \Delta t) \approx y(t) + \Delta y = y(t) + \operatorname{tg}(\alpha) \cdot \Delta t = y(t) + \frac{\Delta y}{\Delta t} \cdot \Delta t = y(t) + f(x, y, t) \cdot \Delta t$$

или

$$y(t + \Delta t) \approx y(t) + f(x, y, t) \Delta t$$

Состояние системы в будущем

Состояние системы в настоящем

Скорость, тенденция

Шаг

Приращение, изменение

Последняя формула называется **формулой Эйлера**.

Очевидно, чтобы узнать состояние системы в будущем $y(t + \Delta t)$, надо к настоящему состоянию системы $y(t)$ прибавить изменение Δy , прошедшее за время Δt .

$$\begin{aligned} \text{будущее} &= \text{настоящее} + \text{изменение} \\ \text{будущее} &= \text{настоящее} + \text{скорость} \cdot \text{шаг} \end{aligned}$$

Рассмотрим еще раз это важное соотношение, выведя его из геометрических соображений (рис. 10.3).

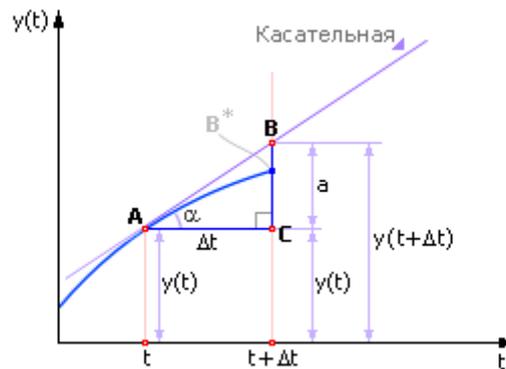


Рис. 10.3. Геометрическая иллюстрация метода Эйлера

Пусть А — точка, в которой состояние системы известно. Это «настоящее» состояние системы.

В точке А к траектории движения системы проведем касательную. Касательная — это производная функции $f(x(t), y(t), t)$ по переменной t . Производную в точке всегда легко вычислить, достаточно подставить известные переменные (в момент «Настоящее» они известны) в формулу $y' = f(x(t), y(t), t)$.

Заметим, что, по определению, производная связана с углом наклона касательной: $y' = \text{tg}(\alpha)$, значит, угол α легко вычислить ($\alpha = \text{arctg}(y')$) и провести касательную.

Проведем касательную до пересечения с линией $t + \Delta t$. Момент $t + \Delta t$ соответствует «будущему» состоянию системы. Проведем линию параллельно оси t от точки А до пересечения с линией $t + \Delta t$. Линии образуют прямоугольный треугольник АВС, один катет которого равен Δt (известен). Известен также угол α . Тогда второй катет в прямоугольном треугольнике АВС равен: $a = \Delta t \cdot \text{tg}(\alpha)$. Теперь легко вычислить ординату точки В. Она состоит из двух отрезков — $y(t)$ и a . Ордината символизирует положение системы в точке $y(t + \Delta t)$. То есть $y(t + \Delta t) = y(t) + a$ или далее $y(t + \Delta t) = y(t) + \Delta t \cdot \text{tg}(\alpha)$ или, подставляя дальше, имеем: $y(t + \Delta t) = y(t) + \Delta t \cdot y'$ и, наконец,

$y(t + \Delta t) = y(t) + \Delta t \cdot f(x(t), y(t), t)$. Снова мы получили формулу Эйлера (из геометрических соображений).

Эта формула может дать точные результаты только при очень малых Δt (говорят при $\Delta t \rightarrow 0$). При $\Delta t \neq 0$ формула дает расхождение между истинным значением y и расчетным, равное ε , поэтому в ней должен стоять знак приближенного равенства, либо она должна быть записана так:

$$y(t + \Delta t) = y(t) + \Delta t \cdot f(x(t), y(t), t) + \varepsilon.$$

И в самом деле. Взгляните еще раз на рис. 10.3. Будем мысленно сдвигать линию $t + \Delta t$ влево (фактически, будем приближать значение Δt к нулю). Как нетрудно видеть, расстояние $BB^* = \varepsilon$, — то есть ошибка — будет сокращаться. В пределе (при $\Delta t \rightarrow 0$) значение ошибки ε будет равно нулю.

Заменяя реальную кривую прямой (касательной) на отрезке Δt , мы вносим в решение ошибку, попадая в результате не в точку «2» (см. рис. 10.2), а рядом, в точку «3». Очевидно, что этот численный метод на каждом шаге имеет погрешность расчета ε .

Из рисунка видно, что чем меньше взять величину Δt , тем меньше будет ошибка расчета ε . То есть для расчета поведения системы на сколько-нибудь продолжительном отрезке времени (например, от t_0 до t_k), чтобы уменьшить ошибку на каждом шаге, шаги Δt делают по возможности малыми. Для достижения точки t_k отрезок $(t_k - t_0)$ делится на отрезки длиной Δt ; таким образом, всего получится $N = (t_k - t_0)/\Delta t$ шагов. В результате расчета придется формулу Эйлера применить для каждого шага, то есть N раз. Но следует иметь в виду, что ошибки ε_i на каждом i -ом шаге (в простейшем случае) складываются, а общая ошибка быстро накапливается (см. рис. 10.4). И в этом состоит существенный недостаток данного метода. Хотя с помощью этого метода можно получить (в численном виде) решение любого дифференциального уравнения (в том числе и неразрешимого аналитически). Уменьшая шаг, мы получаем более точные решения, но при этом не следует забывать, что увеличение числа шагов ведет к вычислительным затратам и снижению быстродействия. Кроме того, при большом числе итераций в расчет вносится другая существенная погрешность из-за ограниченной точности вычислительных машин и ошибок округления.

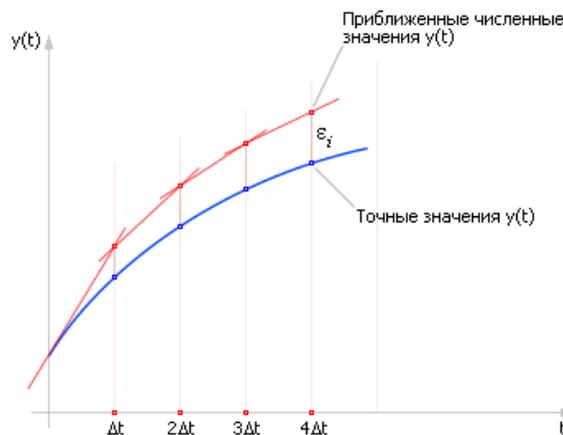


Рис. 10.4. Нарастание суммарной ошибки в методе Эйлера на ряде шагов

Задача 1. Дано дифференциальное уравнение $y' = 2ty$. Задано начальное положение системы: $y(0) = 1$. Требуется найти $y(t)$, то есть поведение системы на интервале времени t от 0 до 1.

Аналитический способ решения задачи 1

$$y' = 2ty.$$

Методом разделения переменных найдем:

$$\frac{y'}{y} = 2t$$

$$\int \frac{y'}{y} dt = \int 2t dt$$

Будем интегрировать от 0 до t_i , тогда согласно правилам интегрирования имеем:

$$\ln y(t) \Big|_0^{t_i} = t^2 \Big|_0^{t_i}$$

$$\ln y(t_i) - \ln y(0) = t_i^2$$

$$\ln y(t_i) - \ln 1 = t_i^2$$

$$y(t_i) = e^{t_i^2}$$

Полученное аналитическое решение характеризуется тем, что оно является абсолютно точным, но если уравнение окажется сколько-нибудь сложным, то решение не будет найдено вовсе. Аналитический путь решения не универсален.

Численный способ решения задачи 1

Численный способ решения предполагает, что расчет будет вестись по формуле Эйлера на ряде последовательных шагов. На каждом шаге решение имеет свою ошибку (см. рис. 10.2), поскольку на каждом шаге кривая заменяется прямым отрезком.

При алгоритмической реализации расчет реализуется циклом, в котором изменяется t (счетчик t) и y :

$$t := t + \Delta t$$

$$y := y + 2 \cdot t \cdot y \cdot \Delta t$$

Блок-схема при реализации метода на компьютере показана на рис. 10.5.

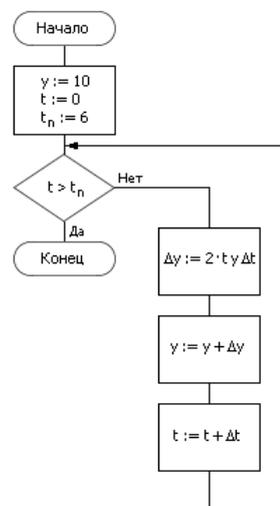


Рис. 10.5. Блок-схема реализации метода Эйлера

В реализации Стратум запись будет выглядеть так (наличие символа «~» при t):

$$t := t + \Delta t$$

$$y := y + 2 \cdot \sim t \cdot y \cdot \Delta t$$

Будем искать значение y рассмотренного ранее примера в численном виде на промежутке от $T = 0$ до $T = 1$. Возьмем число шагов $n = 10$, тогда шаг приращения Δt составит: $\Delta t = (1 - 0)/n = (1 - 0)/10 = 0.1$.

Таблица 10.1.

Численный расчет уравнения методом Эйлера и сравнение результата с точным решением на каждом шаге

i	t_i	$y_i = y_{i-1} + y'_{i-1} \cdot \Delta t$	$y'_i = 2t_i \cdot y_i$	$\Delta y_i = y'_i \cdot \Delta t$	$y_{i+1} = y_i + \Delta y_i$	$y_{\text{точн.}} = \exp(t_i^2)$
0	0,0	1	0	0	1	1
1	0,1	1	0,2	0,02	1,02	1,0101
2	0,2	1.02	0.408	0.0408	1.0608	1.0408
3	0,3	1.061	0.636	0.0636	1.1246	1.0942
4	0,4	1.124	0.900	0.0900	1.2140	1.1735
5	0,5	1.214	1.214	0.1214	1.3354	1.2840
6	0,6	1.336	1.603	0.1603	1.4963	1.4333
7	0,7	1.496	2.095	0.2095	1.7055	1.6323
8	0,8	1.706	2.729	0.2729	1.9789	1.8965
9	0,9	1.979	3.561	0.3561	2.3351	2.2479
10	1,0	2.335	4.669	0.4669	2.8019	2.7183

Обратите внимание на то, что рассчитанное численно значение (y_{i+1}) отличается от точного ($y_{\text{точн.}}$), и погрешность (разница столбцов y_{i+1} и $y_{\text{точн.}}$) в процессе расчета нарастает подобно тому, как было показано на рис. 10.4.

Теперь подсчитаем относительную погрешность σ для расчетного значения $y(1)$, полученного численно, в сравнении с теоретическим точным $y_{\text{теор.}}$ по следующей формуле:

$$\sigma = (1 - y_{\text{расч.}}/y_{\text{теор.}}) \cdot 100\%$$

и сравним σ при различных значениях Δt .

Если будем менять значение шага Δt , например, уменьшать шаг, то относительная погрешность расчета тоже будет уменьшаться. Вот что получится при вычислении значения $y(1)$ с разными значениями шага (см. табл. 10.2).

Таблица 10.2.

Зависимость погрешности расчета от размера шага Δt

Δt	$y_{\text{расч.}}(1)$	$y_{\text{теор.}}(1)$	σ
1/10	2.3346	2.7183	14%
1/20	2.5107	2.7183	8%
1/100	2.6738	2.7183	2%

Как видим, с уменьшением шага приращения Δt уменьшается величина относительной погрешности, а значит, повышается точность расчета.

Обратите внимание, что изменение шага в 10 раз (с 1/10 до 1/100) ведет к изменению величины ошибки примерно тоже в 10 раз (с 14% до 2%). При изменении шага в 100 раз ошибка примерно уменьшится тоже в 100 раз. Та-

ким образом, размер шага и ошибка для метода Эйлера связаны линейно. Хотите уменьшить в 10 раз ошибку — уменьшайте в 10 раз шаг и увеличивайте соответственно в 10 раз количество вычислений. Этот факт в математике принято обозначать символом $\varepsilon = O(\Delta t)$, а метод Эйлера называют методом первого порядка точности.

Поскольку в методе Эйлера ошибка достаточно велика и от шага к шагу накапливается, а точность пропорциональна количеству вычислений, то метод Эйлера обычно применяют для грубых расчетов, для оценки поведения системы в принципе. Для точных количественных расчетов применяют более точные методы.

Примечания

1. Каждый численный метод обладает точностью, поскольку результат отличается от теоретического. Точность метода зависит от величины шага. Различные методы имеют различную точность. Порядок зависимости точности от величины шага обозначают как $O(h)$. У метода Эйлера первый порядок точности, зависимость ошибки от величины шага линейна.
2. Если при уменьшении шага предел y_n стремится к значению $y_{\text{теор.}}$, то говорят, что метод обладает сходимостью. Исследователей интересует скорость сходимости метода.
3. Метод должен быть устойчив. Устойчивость связана с некоторой критической величиной шага. При проявлении неустойчивости наблюдается полное искажение качественной картины расчета, «разболтка» результата.
4. При выборе метода рекомендуется сначала добиться устойчивости, а внутри области устойчивости — сходимости результата. Устойчивость обеспечивает качественную картину. Сходимость обеспечивает количественный результат (см. также рис. 10.10).

Изложенное в пп. 1-4 поясним на примере.

Пример. Пусть

$$\begin{cases} \frac{dA}{dt} = (B - A), & A(0) = 8 \\ \frac{dB}{dt} = (A - B), & B(0) = 5 \end{cases}$$

Качественно это уравнения описывают процесс теплообмена между двумя телами, температуры которых в некоторый момент времени обозначим как A и B . Вообще A и B — переменные, меняющиеся во времени t . Найти поведение системы означает, что надо найти, как будут меняться температуры $A(t)$ и $B(t)$.

Интуитивно ясно, что при начальной разнице температур $A = 8$ и $B = 5$ температуры тел постепенно со временем должны выровняться, так как более горячее тело будет отдавать энергию более холодному, и его температура будет уменьшаться, а более холодное тело будет принимать энергию от более горячего, и его температура будет увеличиваться. Процесс теплообмена за-

кончится (то есть изменения прекратятся) тогда, когда температуры двух тел станут одинаковыми.

Проведем несколько расчетов поведения $A(t)$ и $B(t)$ с разной величиной шага Δt .

Будем брать различную величину шага Δt и находить соответствующие значения A и B во времени по следующим формулам Эйлера:

$$A^{\text{нов.}} = A^{\text{пред.}} + (B^{\text{пред.}} - A^{\text{пред.}}) \cdot \Delta t,$$

$$B^{\text{нов.}} = B^{\text{пред.}} + (A^{\text{пред.}} - B^{\text{пред.}}) \cdot \Delta t.$$

Расчет при $\Delta t = 2$ (табл. 10.3).

Таблица 10.3.
Изменение температур тел при численном расчете с шагом 2

№ шага	t	A	B
0	0	8	5
1	2	2	11
2	4	20	-7

Наблюдается явление «разболтки» (см. рис. 10.6). Неустойчивое решение. Из физических соображений очевидно, что так вести себя два тела в процессе теплообмена не могут.

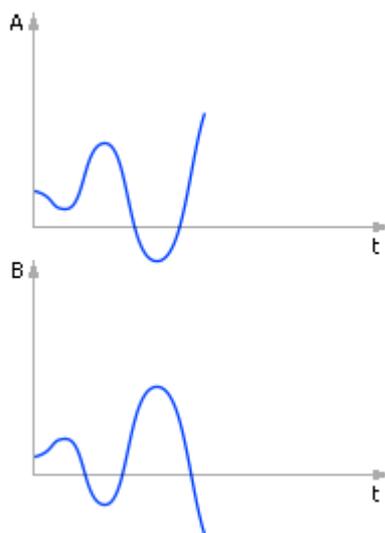


Рис. 10.6. Система ведет себя качественно неверно. Решение неустойчиво
Расчет при $\Delta t = 1$ (табл. 10.4).

Таблица 10.4.
Изменение температур тел при численном расчете с шагом 1

№ шага	t	A	B
0	0	8	5
1	1	5	8
2	2	8	5

Наблюдается поведение решения системы на границе устойчивости (см. рис. 10.7).

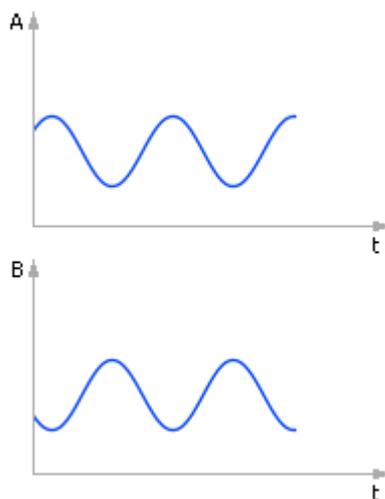


Рис. 10.7. Система ведет себя качественно неверно. Решение на грани устойчивости
 Расчет при $\Delta t = 0.5$ (табл. 10.5).

Таблица 10.5.
 Изменение температур тел при численном расчете с шагом 0.5

№ шага	t	A	B
0	0	8	5
1	0.5	6.5	6.5
2	1.0	6.5	6.5

Решение устойчиво, соответствует правильной качественной картине (см. рис. 10.8). Температуры тел постепенно сближаются, становятся со временем одинаковыми. Но решение пока имеет большую погрешность.

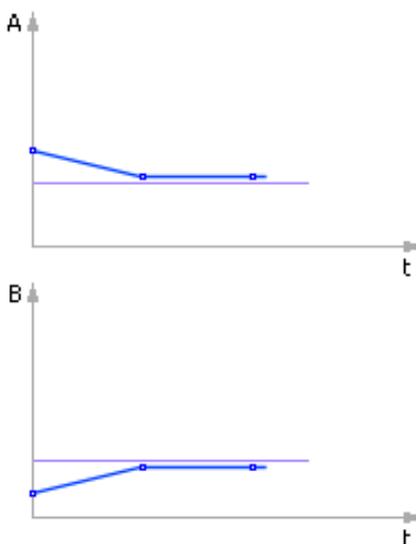


Рис. 10.8. Система ведет себя качественно правильно. Решение имеет большую погрешность
 Расчет при $\Delta t = 0.1$ (табл. 10.6).

Таблица 10.6.

Изменение температур тел при численном расчете с шагом 0.1

№ шага	t	A	B
0	0	8	5
1	0.1	7.7	5.3
2	0.2	7.46	5.54
3	0.3	7.27	5.73
4	0.4	7.12	5.88
5	0.5	7.00	6.00

Решение устойчиво. Решение более точно (см. рис. 10.9).

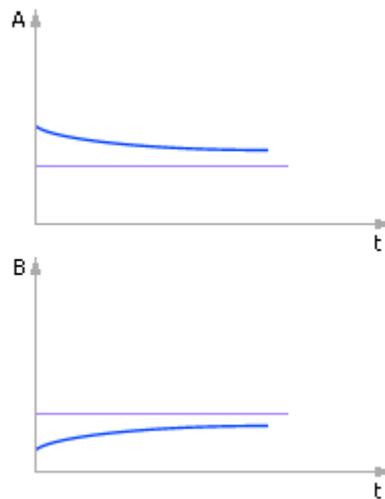


Рис. 10.9. Система ведет себя качественно верно. Количественно решение более точно

Роль изменения величины шага иллюстрирует рис. 10.10.

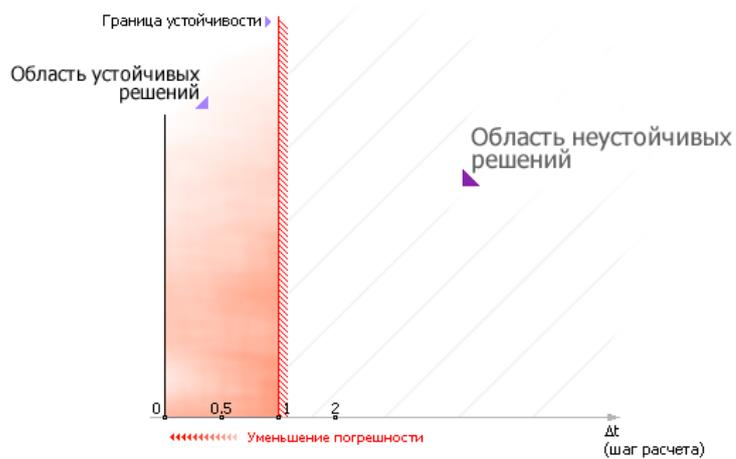


Рис. 10.10. Связь величины шага расчета с устойчивостью метода и его точностью

11. ПОСТРОЕНИЕ МОДЕЛИ ДИНАМИЧЕСКОЙ СИСТЕМЫ В ВИ-

ДЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ И РАСЧЕТ ЕЕ МЕТОДОМ ЭЙЛЕРА

Выполним построение модели динамической системы в виде дифференциальных уравнений и расчет ее методом Эйлера на примере.

Пример. Пусть исследуется система двух материальных тел **A** и **B** с различными теплофизическими свойствами (см. рис. 11.1). Система контактирует с опорой с температурой T_n и помещена во внешнюю среду с температурой T_c . Интересует протекание процесса изменения температур тел.

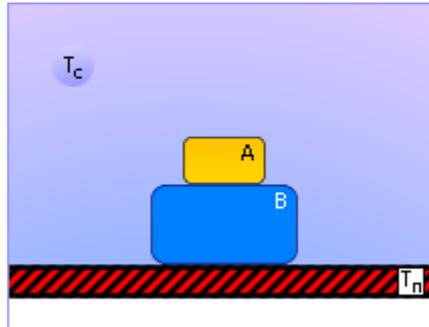


Рис. 11.1. Система взаимодействующих тел в задаче теплопроводности

В процессе жизни в системе изменяются (могут измениться) четыре показателя: температуры тел **A**, **B**, T_c , T_n . Значит, мы имеем дело с четырьмя переменными, зависящими от времени (поскольку переменные меняют свои значения со временем). Введем эти переменные: $X_1(t)$, $X_2(t)$, $X_3(t)$, $X_4(t)$.

Для построения математической модели данной системы отразим процесс теплопередачи в виде графа зависимостей (рис. 11.2).

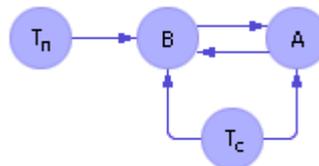


Рис. 11.2. Граф зависимости переменных системы

Если иметь в виду переменные $X_1(t)$, $X_2(t)$, $X_3(t)$, $X_4(t)$, то граф будет выглядеть так, как показано на рис. 11.3.

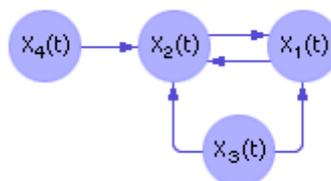


Рис. 11.3. Граф зависимости переменных модели

Стрелка от **A** к **B** обозначает изменение температуры $X_2(t)$ объекта **B** под влиянием объекта **A**. Понятно, что ряд стрелок (например, от **B** к T_c , от **A** к T_n и др.) отсутствует, то есть нет влияния одних параметров на другие: тело **B** не в состоянии сколько-нибудь существенно нагреть открытую атмосферу,

а тело **A** — массивную и потенциально бесконечную опору. Строго говоря, такое влияние есть, но оно настолько ничтожно, что разумно им пренебречь.

Поскольку переменных четыре, то нам необходимо, как минимум, четыре закона, описывающих их изменение. В общем виде, учитывая, от каких переменных зависит каждый показатель, получим:

- для тела **A** имеем зависимость температуры $X_1(t)$ от температуры тела **B** и температуры атмосферы T_c : $dX_1(t)/dt = f_1(X_2(t), X_3(t))$;
- для тела **B** имеем зависимость температуры $X_2(t)$ от температуры тела **A**, температуры атмосферы T_c и опоры T_n : $dX_2(t)/dt = f_2(X_1(t), X_3(t), X_4(t))$.

Стрелки, входящие в соответствующий кружок, указывают на количество влияющих параметров, а то, откуда они исходят, определяет конкретные названия переменных.

Для среды закон имеет вид: $X_3(t) = \text{const}$, то есть, температура атмосферы T_c не зависит от остальных составляющих данной системы и, соответственно, не изменяется. Для опоры закон имеет вид: $X_4(t) = \text{const}$, то есть, температура опоры T_n не зависит от остальных составляющих данной системы и, соответственно, не изменяется [9].

Система законов в первом приближении сформирована. Остается определить их конкретный вид: раскрыть, что из себя представляют значения выражений f_1 и f_2 . **Так как мы имеем дело с системой, зависящей от своего прошлого поведения на каждом последующем шаге, то мы применили для ее описания дифференциальные уравнения.**

Основной динамический закон для описания изменения переменной (уравнение движения) имеет вид:

$$dX(t)/dt = w(x(t), y(t), z(t), \dots).$$

Физический смысл записи таков. Производная в левой части уравнения, по определению, показывает, насколько изменяется X с изменением времени t . В инженерии подобное изменение называется **скоростью, темпом, тенденцией**. Итак, чтобы записать закон изменения переменной в дифференциальных уравнениях, надо указать скорость изменения переменных.

Сначала рассмотрим первое уравнение:

$$dX_1(t)/dt = f_1(X_1(t), X_2(t), X_3(t)).$$

Появление X_1 в правой части означает, что скорость изменения температуры зависит от собственного состояния тела. Что такое f_1 ? — это функция, связывающая переменные X_1, X_2, X_3 между собой. То есть, переменные соединены друг с другом знаками операций.

Взгляните на граф на рис. 11.3. Какие пары переменных взаимодействуют? Стрелки соединяют $X_1(t)$ с $X_2(t)$, $X_1(t)$ с $X_3(t)$, то есть имеет место два процесса, влияющих на скорость. Мы рассматриваем процессы теплообмена тел. Известно, что два процесса теплообмена независимы, то есть не управляют друг другом. Значит, результаты двух процессов можно складывать, они как бы накладываются друг на друга. Действительно, тепло, переданное от одного тела, складывается с теплом, переданным от другого. Таким образом, имеем:

$$dX_1(t)/dt = g_1(X_1(t), X_2(t)) + g_2(X_1(t), X_3(t)).$$

Раскроем структуру оставшихся выражений g_1 и g_2 . Очень удобно, что g_1 никак не зависит от g_2 и может рассматриваться отдельно. Такое разделение возможно, так как процессы g_1 и g_2 независимы. Процесс g_1 идет независимо от того, идет или нет процесс g_2 . Независимость процессов и линейность (аддитивность) выражений — понятия связанные. Так как процессы g_1 и g_2 независимы, то забудем на некоторое время о g_2 .

Какой знак нужно поставить между $X_1(t)$ и $X_2(t)$ в выражении g_1 ? Возможные варианты:

$$\begin{aligned} &X_1(t) + X_2(t); \\ &X_2(t) - X_1(t); \\ &X_1(t) - X_2(t); \\ &X_1(t) \cdot X_2(t); \\ &X_1(t) / X_2(t); \\ &X_2(t) / X_1(t); \\ &X_1(t) \wedge X_2(t); \\ &X_2(t) \wedge X_1(t) \end{aligned}$$

и далее более сложные, например, $X_1^2(t) \cdot \cos(X_2(t))/\exp(X_1(t))$. Исследователь начнет с наиболее простых выражений — природа построена просто. И только если простейшие выражения не удовлетворяют исследователя, он переходит к более сложным вариантам описания.

Описанное выше является принципом: «не вводи сущностей без надобности» (принцип Оккама).

Пусть: $dX_1(t)/dt = X_1(t) + X_2(t)$. Какие есть качественные варианты у этой физической системы?

- $X_1(t) > X_2(t)$. Тело **A** теплее тела **B**. Теплопоток при контакте двух тел направлен от **A** к **B**. Тело **A** отдает тепло телу **B**. То есть в процессе контакта значение $X_1(t)$ падает — уменьшается. (Нас интересует будущее именно $X_1(t)$, а не $X_2(t)$ — см. уравнение: $dX_1(t)/dt$). Посмотрим, так ли это в уравнении-гипотезе $dX_1(t)/dt = X_1(t) + X_2(t)$? Сумма $X_1(t) + X_2(t)$ может принимать как положительные, так и отрицательные значения, следовательно, значение $dX_1(t)/dt$ также может быть как положительным, так и отрицательным, а это, в свою очередь, значит, что $X_1(t)$ то растет, то падает. Противоречит физической картине, рассмотренной чуть выше: мы заключили, что при условии $X_1(t) > X_2(t)$ $X_1(t)$ может только лишь уменьшаться. Поэтому вариант гипотезы $dX_1(t)/dt = X_1(t) + X_2(t)$ неприемлем и надо пробовать другой.

Пусть теперь $dX_1(t)/dt = X_2(t) - X_1(t)$. Какие есть качественные варианты у этой физической системы?

- $X_1(t) > X_2(t)$. Тело **A** теплее тела **B**. Теплопоток при контакте двух тел направлен от **A** к **B**. Тело **A** отдает тепло телу **B**. То есть в процессе контакта значение $X_1(t)$ падает — уменьшается. Посмотрим, так ли это в уравнении? $X_1(t) > X_2(t)$, то есть $X_2(t) - X_1(t) < 0$, значит, $dX_1(t)/dt < 0$, следовательно, $X_1(t)$ падает. Вывод не противоречит физической кар-

тине. Значит, пока данный вариант приемлем и надо проверить его на остальных качественных ситуациях.

- $X_1(t) < X_2(t)$. Тело **A** холоднее тела **B**. Теплоток при контакте двух тел направлен от **B** к **A**. То есть в процессе контакта значение $X_1(t)$ растет — увеличивается. Посмотрим, так ли это в уравнении? $X_1(t) < X_2(t)$, то есть $X_2(t) - X_1(t) > 0$, значит, $dX_1(t)/dt > 0$, следовательно, $X_1(t)$ растет. Вывод не противоречит физической картине. Пока данный вариант приемлем и надо проверять его далее.
- $X_1(t) = X_2(t)$. Температура тела **A** равна температуре тела **B**. Теплоток при контакте двух тел равен нулю. То есть значение $X_1(t)$ не изменяется — тело **A** не отдает и не принимает тепло. Посмотрим, так ли это в уравнении? $X_1(t) = X_2(t)$, значит, $X_2(t) - X_1(t) = 0$, значит, $dX_1(t)/dt = 0$, значит, $X_1(t)$ не изменяется. Вывод не противоречит физической картине. Значит, данный вариант принимается, так как он правильно (пока только качественно!) отражает физическую картину во всех случаях.

Других вариантов существования системы нет, и рассмотрение оканчивается.

Забыв на некоторое время о g_1 , так же можно рассмотреть и g_2 , что предлагается проделать читателю самостоятельно. В конечном итоге мы получим:

$$dX_1(t)/dt = (X_2(t) - X_1(t)) + (X_3(t) - X_1(t)).$$

Так как, во-первых, у разных материалов разность температур влияет на скорость изменения температуры тела различным способом и, во-вторых, скорости двух процессов (у двух разных пар материалов) могут быть разными, то скорректируем модель, используя коэффициент теплопроводности, который играет роль усилителя (ослабителя) процессов. Это коэффициент влияния связи на объект. При $K = 0$ влияние отсутствует, связь отключается. При $K = 0.0001$ влияние слабое. При $K = 1000$ влияние связи огромно. Понятно, что коэффициент стоит при выражении процесса: $K \cdot (X_2(t) - X_1(t))$, где « \cdot » означает знак некоторой операции, а именно — операции умножения. Эта операция дает *зависимость* одного члена от другого (в нашем случае K от $X_2(t) - X_1(t)$).

При $K = 0$, какие бы значения ни принимало выражение $X_2(t) - X_1(t)$, результат $K \cdot (X_2(t) - X_1(t))$ дает 0. При $K = 0.1$ значение выражения принудительно ослабляется в 10 раз. То же самое верно и со стороны выражения $X_2(t) - X_1(t)$ — это естественно, ведь общее выражение $K \cdot (X_2(t) - X_1(t))$ симметрично. Значит, мультипликативная связь моделирует свойство *нелинейности* и взаимозависимости процессов (один может сводить на нет действие другого). Сложение таким свойством не обладает.

Модель имеет вид:

$$dX_1(t)/dt = K_{21} \cdot (X_2(t) - X_1(t)) + K_{31} \cdot (X_3(t) - X_1(t)).$$

В конце следует проверить размерности уравнения; размерность левой части должна совпасть с размерностью правой. Напомним только, что производная имеет размерность показателя X , деленного на единицу времени.

Теперь мы в состоянии синтезировать аналогично второе уравнение (рекомендуем проверить правильность данного уравнения самостоятельно):

$$dX_2(t)/dt = K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t)).$$

Уравнение изменения температуры атмосферы: $dX_3(t)/dt = 0$. То есть $X_3 = \text{const}$ (X_3 не изменяется) [9].

Уравнение изменения температуры опоры: $dX_4(t)/dt = 0$. То есть $X_4 = \text{const}$ (X_4 не изменяется).

Вся система уравнений в сборе имеет вид:

$$dX_1(t)/dt = K_{21} \cdot (X_2(t) - X_1(t)) + K_{31} \cdot (X_3(t) - X_1(t));$$

$$dX_2(t)/dt = K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t));$$

$$dX_3(t)/dt = 0;$$

$$dX_4(t)/dt = 0.$$

По физическим соображениям ясно: сколько тепла вытекает из **A** в **B**, столько же тепла поступает в **B** из **A**, то есть $K_{21} = K_{12}$.

К записи общей модели остается добавить конкретные значения коэффициентов теплопроводности ($K_{12} = K_{21}$, K_{31} , K_{32} , K_{42}) и начальное состояние системы:

$$X_1(0) = a,$$

$$X_2(0) = b,$$

$$X_3(0) = c,$$

$$X_4(0) = d,$$

где a, b, c, d — числа, указывающие температуру соответствующего объекта в момент времени $t = 0$.

Построена система обыкновенных дифференциальных уравнений в каноническом виде от единственной независимой переменной времени.

Система — потому что в наличии имеется несколько уравнений.

Обыкновенных — производная использована обычная, а не частная, так как использована одна переменная времени.

Дифференциальных — в уравнении встречается выражение производной dX/dt .

Уравнений — в выражениях имеется знак уравнивания.

В каноническом виде — производная не стоит под знаком какой-либо функции, встречается один раз и только в левой части уравнения. В правой части какие-либо производные отсутствуют.

Единственной — переменная, по которой берут производную, одна во всех уравнениях (время t).

Независимой переменной — переменная t не зависит более ни от каких переменных и изменяется сама по себе.

Любую динамическую систему можно привести к данному виду.

Выше описано подтверждение гипотез, содержащихся в модели данного примера, естественно, не доказывает абсолютной правильности принятой модели. Проверки могут быть продолжены. Если при очередной проверке гипотеза будет отвергнута, то модель следует снова уточнить. Одной из дополнительных проверок может быть, например, проверка на открытость системы. Выполним такую проверку.

Выше мы получили *открытую систему*, то есть такую, чье суммарное тепло не постоянно, а может изменяться. Это видно из асимметрии стрелок

на графе. Проверим это формально, сложим левые части всех уравнений и, отдельно, правые части. Слева мы имеем следующее:

$$dX_1(t)/dt + dX_2(t)/dt + dX_3(t)/dt + dX_4(t)/dt$$

или

$$d(X_1(t) + X_2(t) + X_3(t) + X_4(t))/dt$$

или

$$dX_{\text{системы}}(t)/dt.$$

В правой части мы имеем следующее:

$$K_{21} \cdot (X_2(t) - X_1(t)) + K_{31} \cdot (X_3(t) - X_1(t)) + K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t))$$

или (с учетом того, что $K_{21} = K_{12}$)

$$K_{31} \cdot (X_3(t) - X_1(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t)).$$

В итоге:

$dX_{\text{системы}}(t)/dt = K_{31} \cdot (X_3(t) - X_1(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t))$,
из чего следует, что левая часть явно не равна нулю и есть утечка или приток тепла в систему извне.

В случае закрытой системы уравнения имели бы вид:

$$dX_1(t)/dt = K_{21} \cdot (X_2(t) - X_1(t)) + K_{31} \cdot (X_3(t) - X_1(t));$$

$$dX_2(t)/dt = K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t));$$

$$dX_3(t)/dt = K_{13} \cdot (X_1(t) - X_3(t)) + K_{23} \cdot (X_2(t) - X_3(t)) + K_{43} \cdot (X_4(t) - X_3(t));$$

$$dX_4(t)/dt = K_{24} \cdot (X_2(t) - X_4(t)) + K_{34} \cdot (X_3(t) - X_4(t)).$$

При сложении это дает: $dX_{\text{системы}}(t)/dt = 0$ — уравнение закрытой системы, в которую извне нечего не притекает и из которой ничего не истекает. Хотя внутри системы происходят изменения (перераспределение тепла), суммарная температура системы неизменна. Ни первая модель (модель открытой системы), ни вторая модель (модель закрытой системы) не являются «плохими». Достигнуты различные цели, даны различные представления исследователя о процессах, о системе в целом.

Динамическая система, которую мы рассматривали выше, — это общий случай. Из нее всегда можно получить *статическую систему*, для чего нужно потребовать: $X_i(t + \Delta t) = X_i(t)$, в результате чего имеем: $dX_i(t)/dt = 0$. Иначе, запись $X_i(t + \Delta t) = X_i(t)$ означает, что прошлое равно настоящему, то есть состояние системы не меняется. Тогда от уравнений остается:
 $K_{21} \cdot (X_2(t) - X_1(t)) + K_{31} \cdot (X_3(t) - X_1(t)) = 0$
 $K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t)) = 0.$
 Это уравнения статики.

После построения базовой системы исследователь может вводить дополнительные конструктивные элементы в систему. Добавим, в схему нагреватель (см. рис. 11.4).

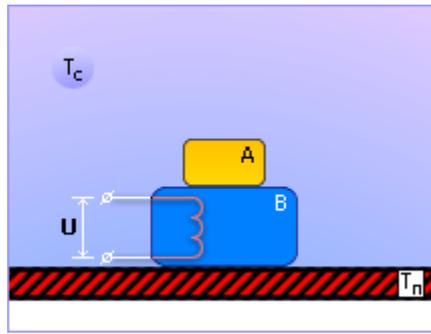


Рис. 11.4. Система взаимодействующих тел со слабым нагревателем в задаче теплопроводности

Граф изменится, так как на тело **В** будет действовать дополнительный объект — нагреватель. Обозначим его температуру переменной $X_5(t)$ (см. рис. 11.5).

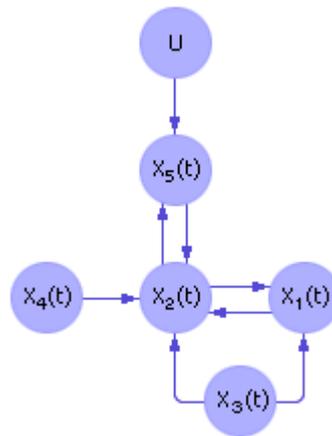


Рис. 11.5. Граф зависимости переменных модели со слабым нагревателем

Изменения в записи будут касаться только второго уравнения, описывающего изменение переменной X_2 во времени. Кроме того, нам понадобится уравнение нагревателя, так как в системе появилась дополнительная переменная $X_5(t)$ (вершина графа) и закон ее изменения должен быть определен.

При записи уравнения следует различать, к какому типу относится источник тепла — слабому или сильному.

Рассмотрим сначала гипотезу о слабом источнике. Обратите внимание: если нагреватель будет греть сильнее тела **В**, то, конечно, тело **В** нагревается от нагревателя, но если нагреватель в какой-то момент окажется холоднее тела **В**, то тело **В** само отдает тепло нагревателю, как бы странно это, на первый взгляд, ни казалось [9].

Так как такой источник энергии может не только нагревать другие тела, но и нагреваться от них сам, он называется **слабым**. То есть от услуг такого источника энергии можно отказаться, поскольку он находится в равном с другими телами положении, может влиять на них и может испытывать их влияние. На **сильный** источник другие тела воздействовать не могут, а сам он — может на них воздействовать, навязывая им свою волю безусловно.

Понятно, что в уравнении для тела **В** появится дополнительное слагаемое, аналогичное ранее рассмотренным:

$$dX_2(t)/dt = K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t)) + K_{52} \cdot (X_5(t) - X_2(t)).$$

Уравнение нагревателя может быть записано так:

$$dX_5(t)/dt = (K \cdot U(t) - X_5(t))/L + K_{25} \cdot (X_2(t) - X_5(t)),$$

где $U(t)$ — напряжение питания источника.

Начальные условия $X_5(0)$ и функция работы источника $U(t)$ должны быть заданы:

$$\begin{aligned} X_5(0) &= h, \\ U(t) &= f(t). \end{aligned}$$

Уравнение связывает причину (напряжение источника питания) и следствие (температура нагревателя). Свойства этого элемента таковы, что при выходе на рабочий режим нагреватель поддерживает постоянную температуру. При включении нагревателя рабочая температура устанавливается со временем, постепенно нарастая, при выключении источника нагреватель постепенно остывает. Процесс нагрева и остывания нагревателя инерционный. Фактически, для описания этих свойств нагревателя достаточно записать аperiодический закон (который мы ранее уже обсуждали в лекции 04) изменения его температуры $X_5(t)$ по отношению ко входу $U(t)$, что мы и сделали. В уравнении учтен коэффициент усиления K между входом $U(t)$ и выходом $X_5(t)$, инерционность процесса нагрева L и коэффициент теплопередачи K_{25} между телами $X_2(t)$ и $X_5(t)$.

Рассмотрим гипотезу о сильном нагревателе. Это случай, когда нагреватель устроен таким образом, что излученная им энергия не может отразиться телом **В**, или если энергия от тела **В** не может попасть к нагревателю, или если тело **В** не может «отказаться» от услуг нагревателя (см. рис. 11.6).

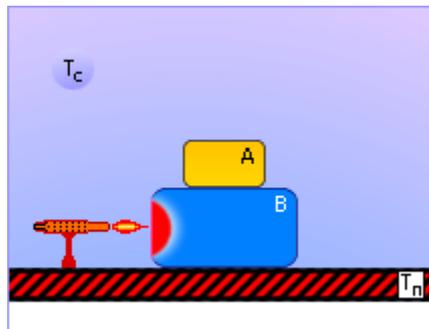


Рис. 11.6. Система взаимодействующих тел с нагревателем

На рис. 11.7 изображен граф зависимостей для данного случая.

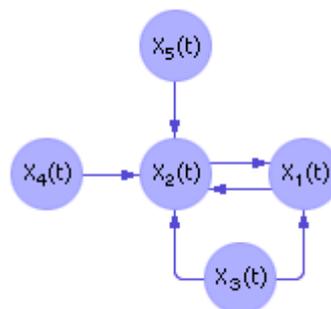


Рис. 11.7. Граф зависимости переменных модели

Неважно, нагрет ли нагреватель сильнее тела **В** или нет — тело **В** будет получать от него энергию в любом случае, причем, в том объеме, в каком эту энергию будет выделять нагреватель. В уравнении тела **В** появится дополнительное слагаемое:

$$dX_2(t)/dt = K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t)) + K_{52} \cdot X_5(t).$$

Уравнение нагревателя может быть записано так:

$$dX_5(t)/dt = f(t).$$

Начальные условия $X_5(0)$ и функция работы источника $f(t)$ должны быть заданы:

$$X_5(0) = h.$$

Вернемся к расчету движения во времени динамической системы, для которой есть все необходимые данные, есть начальное состояние ($X_i(0) = \text{const}$). Можно по формулам (с использованием метода Эйлера, см. лекцию 10) вычислить скорость ее изменения и новое состояние:

новое состояние := старое состояние + скорость · отрезок времени.

Формально эта запись выглядит так:

$$dx(t)/dt = f(x(t))$$

или (в дискретной форме)

$$[x(t + \Delta t) - x(t)]/\Delta t = f(x(t))$$

и, окончательно,

$$x(t + \Delta t) = x(t) + f(x(t)) \cdot \Delta t.$$

Зная новое состояние и считая его уже достигнутым, то есть старым состоянием, используем ту же формулу (подставляя, конечно, в нее уже новые значения). И так продолжаем, прибавляя все новый отрезок времени, двигаясь во времени.

Методом решения дифференциальных уравнений, в общем случае, является интегрирование их по независимой переменной времени t . Простейшим методом численного интегрирования дифференциальных уравнений является метод Эйлера (см. лекцию 10).

При вычислении по методу Эйлера необходимо вычислять значения всех параметров системы *параллельно*, так как производная в каждый момент времени отдельного параметра зависит как от значения самого параметра в данный момент времени, так и от значения другого параметра в этот момент времени.

Рассмотрим практически применение метода Эйлера для расчета процесса изменения температур тел системы. Зададим значения коэффициентов модели: $K_{12} = K_{21} = 0.2$, $K_{31} = 0.1$, $K_{32} = 0.05$, $K_{42} = 0.1$. Зададим начальные условия системы (в момент времени $t = 0$): $X_1(0) = 30^\circ \text{C}$, $X_2(0) = 70^\circ \text{C}$, $X_3(0) = 22^\circ \text{C}$, $X_4(0) = 15^\circ \text{C}$. Выбираем шаг моделирования Δt равный, например, 0.2 с. Примем конечное значение времени моделирования за $T_k = 4$ с.

Подставим значения коэффициентов:

$$X_1(t + \Delta t) = X_1(t) + [0.2 \cdot (X_2(t) - X_1(t)) + 0.1 \cdot (22 - X_1(t))] \cdot \Delta t$$

$$X_2(t + \Delta t) = X_2(t) + [0.2 \cdot (X_1(t) - X_2(t)) + 0.05 \cdot (22 - X_2(t)) + 0.1 \cdot (15 - X_2(t))] \cdot \Delta t$$

и приступаем к моделированию. Процесс моделирования и численные значения отражены в табл. 11.1.

Таблица 11.1.

Таблица расчета изменения значений переменных системы во времени

t	$dX_1(t)/dt$	$dX_2(t)/dt$	$X_1(t)$	$X_2(t)$
0.0	7.20	-15.90	30	70
0.2	6.13	-14.50	31.44	66.82
0.4	5.18	-13.24	32.67	63.92
0.6	4.34	-12.10	33.70	61.27
0.8	3.60	-11.08	34.57	58.85
1.0	2.94	-10.16	35.29	56.63
1.2	2.36	-9.33	35.88	54.60
1.4	1.84	-8.59	36.35	52.74
1.6	1.39	-7.91	36.72	51.02
1.8	0.99	-7.30	37.00	49.44
2.0	0.64	-6.75	37.19	47.97
2.2	0.33	-6.25	37.32	46.62
2.4	0.06	-5.80	37.39	45.37
2.6	-0.18	-5.39	37.40	44.21
2.8	-0.38	-5.02	37.36	43.13
3.0	-0.56	-4.69	37.29	42.13
3.2	-0.71	-4.38	37.18	41.19
3.4	-0.85	-4.10	37.03	40.31
3.6	-0.96	-3.85	36.86	39.49
3.8	-1.06	-3.62	36.67	38.72
4.0	-1.14	-3.41	36.46	38.00

Во времени поведение системы будет выглядеть так, как показано на рис. 11.8.

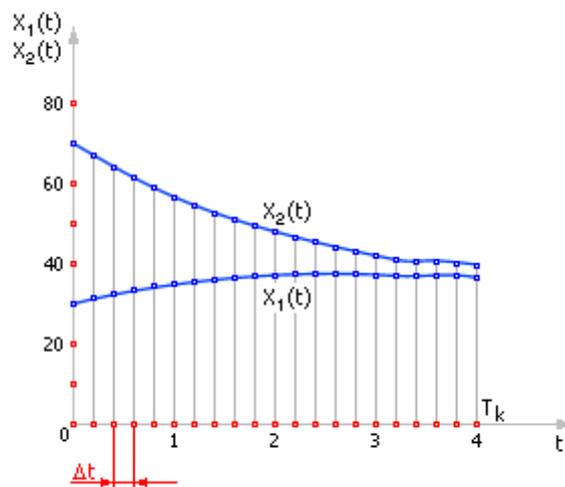


Рис. 11.8. Расчетные траектории поведения системы тел в задаче теплопроводности

На рис. 11.8 показана не вся траектория, а только ее часть.

Рассчитайте всю траекторию, подумайте и ответьте на следующие вопросы:

- почему графики в итоге при большом времени рассмотрения стремятся к числу 18.5;
- почему значение переменной X_1 сначала увеличивается, а потом падает;
- почему графики имеют переломы производных;
- что надо изменить в условиях задачи, чтобы график X_1 все время убывал?

В зависимости от реализации «машины», на которой автоматически будет имитироваться процесс, запись может выглядеть по-разному.

Формальная математическая запись

Приведем ниже вариант формальной математической записи для случая открытой системы тел со слабым нагревателем.

$$dX_1(t)/dt = K_{21} \cdot (X_2(t) - X_1(t)) + K_{31} \cdot (X_3(t) - X_1(t))$$

$$dX_2(t)/dt = K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t)) + K_{52} \cdot (X_5(t) - X_2(t))$$

$$dX_3(t)/dt = 0$$

$$dX_4(t)/dt = 0$$

$$dX_5(t)/dt = (K \cdot U(t) - X_5(t))/L + K_{25} \cdot (X_2(t) - X_5(t))$$

$$X_1(0) = a$$

$$X_2(0) = b$$

$$X_3(0) = c$$

$$X_4(0) = d$$

$$X_5(0) = h$$

$$U(t) = f(t)$$

$$K_{12} = 0.2; K_{21} = 0.2; K_{31} = 0.1; K_{32} = 0.05; K_{42} = 0.1; K_{52} = 0.4; K_{25} = 0.4; K = 1;$$

$$L = 1$$

$$T_k = 4$$

При реализации на математической машине среды Stratum эта запись будет дополнена элементами (функциями) ввода и вывода информации и командами управления (например, «стоп»), а часть элементов будет спрятана (например, при задании начальных условий) [9].

Иногда начальные условия могут быть помещены внутрь уравнений. Для этого в следующей записи мы воспользуемся дельта-функцией Дирака (см. лекцию 31).

$$dX_1(t)/dt = K_{21} \cdot (X_2(t) - X_1(t)) + K_{31} \cdot (X_3(t) - X_1(t)) + a \cdot \text{delta}(t)$$

$$dX_2(t)/dt = K_{12} \cdot (X_1(t) - X_2(t)) + K_{32} \cdot (X_3(t) - X_2(t)) + K_{42} \cdot (X_4(t) - X_2(t)) + K_{52} \cdot (X_5(t) - X_2(t)) + b \cdot \text{delta}(t)$$

$$dX_3(t)/dt = c \cdot \text{delta}(t)$$

$$dX_4(t)/dt = d \cdot \text{delta}(t)$$

$$dX_5(t)/dt = (K \cdot U(t) - X_5(t))/L + K_{25} \cdot (X_2(t) - X_5(t)) + h \cdot \text{delta}(t)$$

$$U(t) := f(t)$$

$K_{12} := 0.2; K_{21} := 0.2; K_{31} := 0.1; K_{32} := 0.05; K_{42} := 0.1; K_{52} := 0.4; K_{25} := 0.4;$
 $K := 1; L := 1;$
 $T_k := 4$
 $t := t + \Delta t$
 $\text{osc2d}(T, X_1)$
 $\text{stop}(T > T_k)$

$\text{Osc2d}(T, X_1)$ — функция, рисующая на экране точку с координатами T и X_1 для каждого момента времени. Выше данная функция приведена условно: в действительности надо использовать несколько функций для настройки окон, масштаба изображения, цвета, толщины, стиля линии и тому подобного. Подробно реализацию двухмерного осциллографа в среде «Stratum-2000» вы можете посмотреть в тексте имиджа OSCSpace2D.

Вариант формальной математической записи подразумевает, что весь код заключен в рамках одного элемента (имиджа, если использовать термины среды «Stratum-2000»). Будет более наглядно, если распределить код по отдельным элементам, связав их между собой связями (см. рис. 11.9), в этом случае структура проекта останется обозримой; кроме того, если потребуется изменить что-то в одном из блоков, то такой подход позволит не менять код в остальных блоках. Еще один плюс такого подхода — отдельные «кирпичики» схемы могут быть применены неоднократно в готовом виде в виде копии имиджа [9].

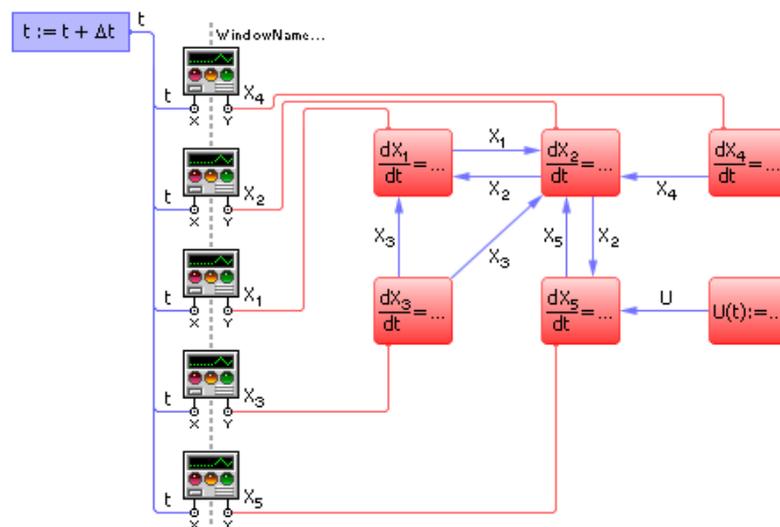


Рис. 11.9. Схема проекта в среде Stratum-2000, реализующего полную модель теплопроводности системы тел с элементами интерфейса

Алгоритмическая реализация

В случае реализации на алгоритмической машине следует воспользоваться алфавитом алгоритмов: блоками начала и конца алгоритма, ввода-вывода информации, задания начальных условий, вычислительными блоками, блоками условия, конструкциями цикла. На рис. 11.10 представлена блок-схема алгоритма, имитирующего теплообмен между телами.

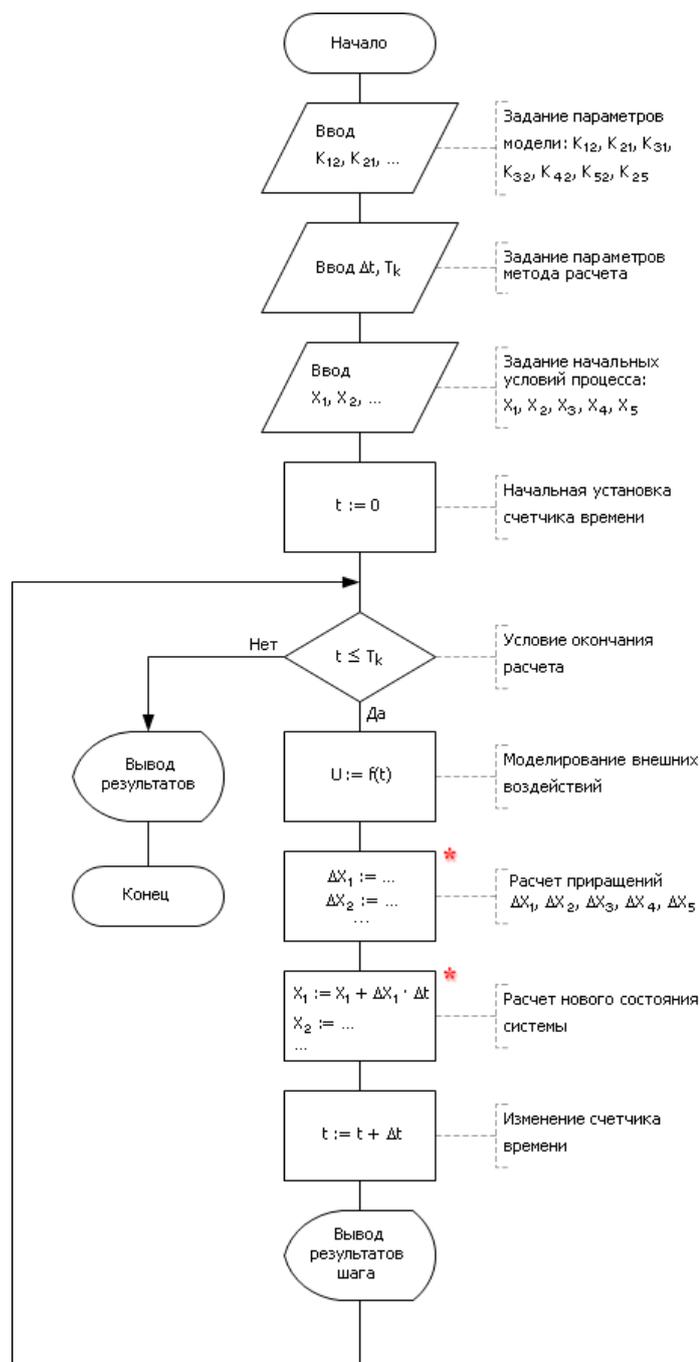


Рис. 11.10. Блок-схема алгоритма, имитирующего теплообмен между телами

Для реализации нам понадобились следующие инструменты:

- устройства ввода информации;
- устройство вывода информации;
- вычислительные блоки для реализации математических формул;
- цикл для имитации течения времени.

В теле цикла присутствуют:

- счетчик времени;
- условие выхода из цикла для контроля за временем моделирования;
- блок задания внешних воздействий;
- блок расчета приращений;
- блок расчета нового состояния системы;
- блок вывода текущих состояний.

12. УРАВНЕНИЯ ВЫСШИХ ПОРЯДКОВ

Если при построении модели динамической системы оказалось, что в его записи использованы производные второй и более степени, это значит, что мы имеем дело с уравнениями высших порядков. Уравнения высших порядков приводят к системе уравнений первого порядка. Рассмотрим это на примере [9].

Пример 1. Дано следующее дифференциальное уравнение второго порядка: $y'' + y'/x + y = 0$. Необходимо привести его к системе уравнений первого порядка.

Обозначим: $dy/dt = z$; $d^2y/dt^2 = z'$. Тогда: $z' + z/x + y = 0$. И далее приходим к системе дифференциальных уравнений первого порядка:

$$\begin{cases} \frac{dy}{dt} = z \\ \frac{dz}{dt} = -y - \frac{1}{x} \cdot z \end{cases}$$

Далее решаем эту систему уже известными нам способами, например, методом Эйлера (см. Лекцию 10. Численные методы интегрирования дифференциальных уравнений. Метод Эйлера).

13. СТРУКТУРНО ПЕРЕСТРАИВАЕМЫЕ МОДЕЛИ

В ряде случаев объект, который мы моделируем или которым мы управляем, может в процессе функционирования **качественно** меняться. До этого были случаи, когда изменялись **количественно** отдельные параметры системы, значения ее отдельных переменных, но не менялись сами уравнения, названия или количество переменных, их связи. В структурно изменяющихся системах, например, могут исчезать старые или появляться новые связи, элементы в составе системы. В процессе моделирования может измениться сама структура моделируемой системы. Модель также должна в процессе моделирования иметь возможность изменить свои свойства и уравнения [9].

Задача 1. В бак наливаются две разные жидкости. Имеется отверстие для слива в нижней части бака и отверстие для аварийного слива сбоку (см. рис. 13.1). Цель имитации — анализ поведения системы.

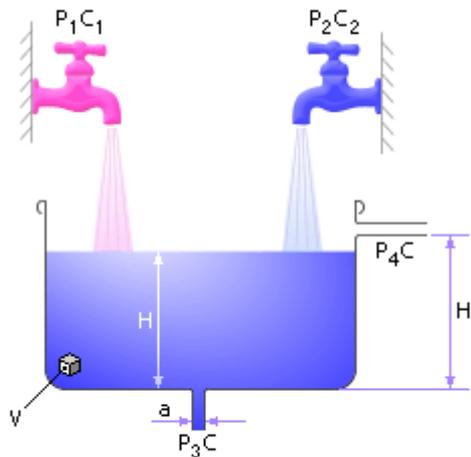


Рис. 13.1. Моделируемая система

Обозначим: P — расход жидкости [л/час], C — концентрация [кг/м³], S — площадь [м²], V — объем жидкости [м³]. Обозначим концентрации входных потоков как C_1 и C_2 . Оба потока смешиваются в баке, в результате чего образуется жидкость с некоторой средней концентрацией C . Через выходные отверстия будет сливаться жидкость именно с концентрацией C . Тогда: $dC/dt = (P_1 \cdot C_1 + P_2 \cdot C_2 - P_3 \cdot C - P_4 \cdot C)/V$.

Изменение концентрации зависит от соотношения расхода жидкости (поступающей и удаляемой из бака) и концентрации соответствующих потоков: $dV/dt = P_1 + P_2 - P_3 - P_4$.

Изменение объема жидкости в баке зависит от соотношения расхода жидкости (поступающей и удаляемой из бака) соответствующих потоков: $H = V/S$.

Высота жидкости в баке зависит от объема и площади бака: $P_3 = a \cdot H$.

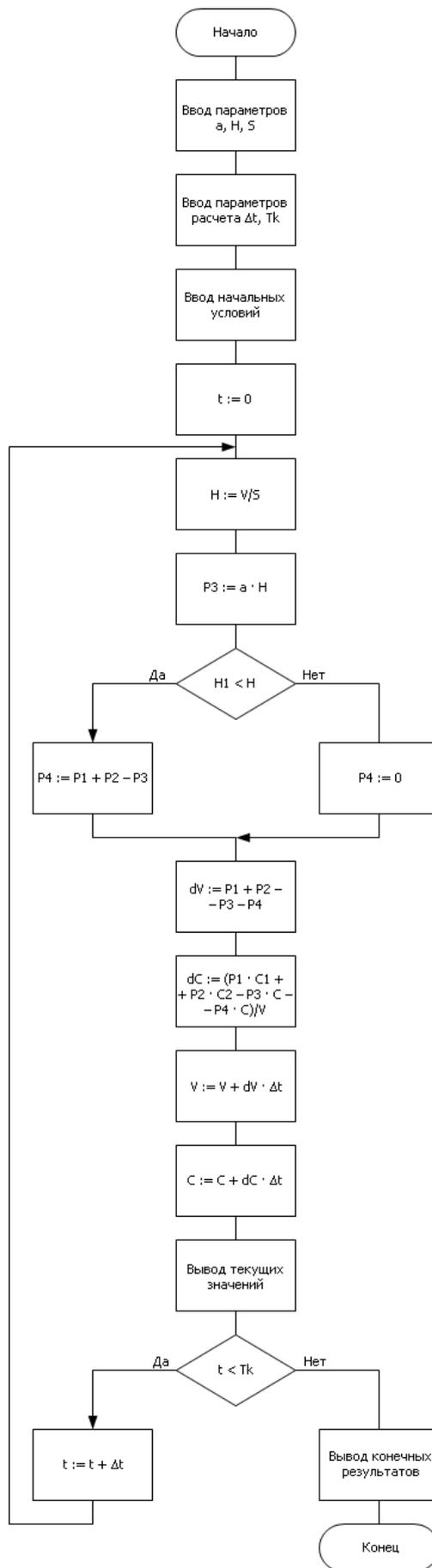
Расход жидкости через сливное отверстие в дне бака зависит от столба жидкости - чем больше столб жидкости, тем сильнее он давит на нижние ее

слои и тем быстрее жидкость выливается из бака (здесь a - коэффициент пропорциональности).

Чтобы учесть влияние на работу системы аварийного клапана, заметим, что если высота H жидкости в баке превысит уровень H_1 , то жидкость начнет переливаться через отверстие аварийного слива сбоку, иначе расхода жидкости через данное отверстие нет. Математически это можно выразить следующей системой:

$$P_4 = \begin{cases} 0, & H \leq H_1 \\ P_1 + P_2 - P_3, & H > H_1 \end{cases}$$

Алгоритм реализации такой модели показан на блок-схеме (см. рис. 13.2). По сравнению с прежними реализациями появился блок условия **IF** — с помощью него подключается один или другой вариант записи уравнения. **Изменение структуры имитируется условным блоком.**



14. УТОЧНЕННЫЙ МЕТОД ЭЙЛЕРА

Метод Эйлера для расчета дифференциальных уравнений имеет небольшую точность расчета. Как было показано ранее (см. Лекцию 10), точность расчета у него зависит от размера шага линейно, зависимость точности от шага — первой степени. Чтобы увеличить точность в 10 раз, надо уменьшить шаг в 10 раз. Можно ли увеличить точность на порядок, но при этом сэкономить на количестве вычислений? Да, такие методы есть. Модифицированный метод Эйлера имеет точность второго порядка. В методе Эйлера производная берется в начале шага и по ней прогнозируется движение системы на конец шага, считая, что во время шага производная неизменна. В течение всего шага производная считается той, какой она была в самом начале шага, это основной источник неточности [9].

Улучшение метода состоит в том, что берется производная не в начале шага, а как промежуточное или среднее на разных участках одного шага. В разных вариантах метода вычисляют несколько производных в разных частях шага и усредняют их. В этом случае число вычислений увеличивается, — но не в десятки раз, — а вот точность возрастает на порядок, в этом и состоит выигрыш.

Требуется решить уравнение $y' = f(x, y, t)$. Идея уточненного метода Эйлера состоит в том, что производную вычисляют не в i -ой точке, а между двумя соседними точками: i и $i + 1$. Данная процедура состоит из следующих шагов:

- в точке i вычисляют значение производной: $f(x_i, y_i, t)$;
- делают пол-шага и вычисляют значение функции на середине отрезка: $y_{i+1/2} = y_i + f(x_i, y_i, t) \cdot \Delta t/2$;
- в точке $i + 1/2$ вычисляют производную: $f(x_{i+1/2}, y_{i+1/2}, t + \Delta t/2)$;
- делается полный шаг из точки i в точку $i + 1$ по значению уточненной производной: $y_{i+1} = y_i + f(x_{i+1/2}, y_{i+1/2}, t + \Delta t/2) \cdot \Delta t$;
- значение t увеличивается: $t := t + \Delta t$. Вся процедура повторяется сначала.

Данный метод обладает точностью $O^2(h)$, то есть на порядок выше, чем метод Эйлера, при увеличении числа вычислений всего в 2 раза.

На рис. 14.1 показано, какой будет ошибка ε (расхождение между реальным и вычисленным теоретическим значением), если шаг делается по значению производной, вычисленной в точке i , как это делается в методе Эйлера. Эта ошибка может быть достаточно велика.

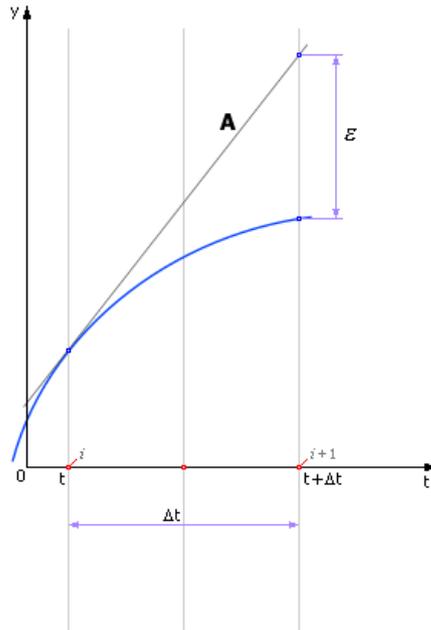


Рис. 14.1. Движение реальной и расчетной системы по методу Эйлера и расхождение между ними

На рис. 14.2 показано, как по значению производной, вычисленной в точке i , делается полшага до точки $t + \Delta t/2$ (направление производной показано линией **A**). И в точке $t + \Delta t/2$ вычисляют новую производную. Касательная в точке $t + \Delta t/2$ будет другой - линия **B**. Ее наклон равен производной в точке $t + \Delta t/2$.

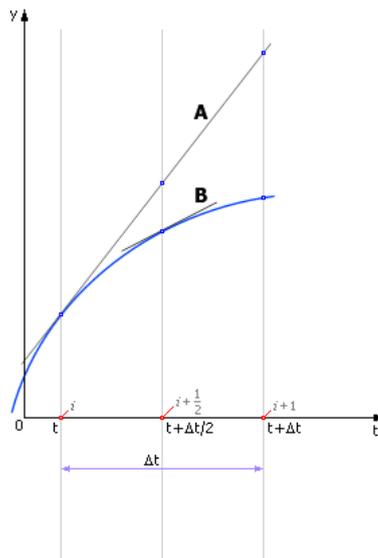


Рис. 14.2. Уточнение значения производной внутри шага расчета

Далее переносят линию **B** обратно в точку t . Это соответствует тому, что из точки t снова делается, — но уже полный, — шаг Δt до точки $t + \Delta t$ по направлению, соответствующему линии **C** (рис. 14.3). Линия **C** параллельна **B**. То есть значение производной в точке t берется искусственно равным про-

изводной в точке $t + \Delta t/2$. Ошибка расчета (см. ε_1) во многих случаях при этом уменьшается.

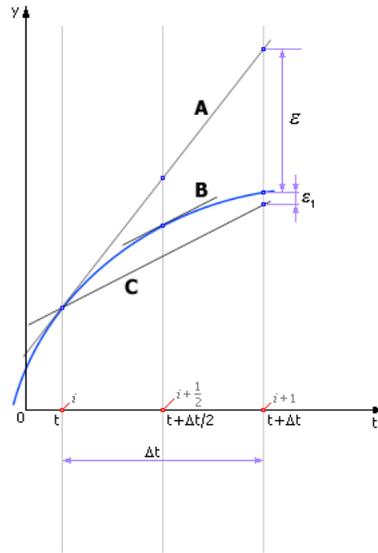


Рис. 14.3. Движение реальной системы и системы, рассчитанной модифицированным методом Эйлера, и расхождение между ними.

Существует другой вариант модифицированного метода Эйлера, когда производную для того, чтобы сделать шаг из точки i , берут не в i -ой точке и не в $i + 1/2$, а как среднее арифметическое двух производных: производной в точке i (направление производной показано на рис. 14.4 линией **A**) и производной в точке $i + 1$ (направление производной показано линией **B**). Направление «средней» производной показано линией **C**.

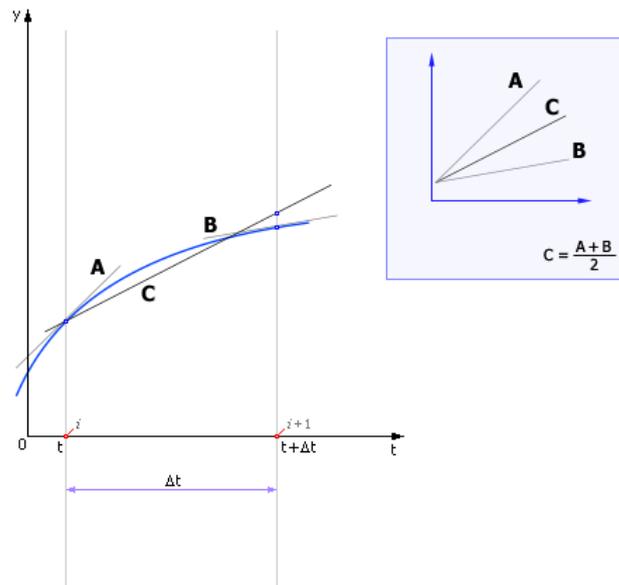


Рис. 14.4. Расчет движения системы по среднему значению производной на шаге

15. МЕТОД РУНГЕ-КУТТЫ

Метод Рунге-Кутты используют для расчета стандартных моделей достаточно часто, так как при небольшом объеме вычислений он обладает точностью метода $O^4(h)$.

Для построения разностной схемы интегрирования воспользуемся разложением функции [9]

$$\frac{dy}{dx} = f(x, y(x)), 0 < x \leq T, y(0) = y_0 \quad (1)$$

в ряд Тейлора:

$$y(x_{k+1}) = y(x_k) + y'(x_k)h + y''(x_k)\frac{h^2}{2} + \dots$$

Заменяем вторую производную в этом разложении выражением

$$y''(x_k) = (y'(x_k))' = f'(x_k, y(x_k)) \approx \frac{f(\tilde{x}, \tilde{y}) - f(x_k, y(x_k))}{\Delta x}$$

где

$$\tilde{x} = x_k + \Delta x, \tilde{y} = y(x_k + \Delta x)$$

Δx подбирается из условия достижения наибольшей точности записанного выражения. Для дальнейших выкладок произведем замену величины «у с тильдой» разложением в ряд Тейлора:

$$\tilde{y} = y(x_k + \Delta x) = y(x_k) + y'(x_k)\Delta x + \dots$$

Для исходного уравнения (1) построим вычислительную схему:

$$y_{k+1} = y_k + f(x_k, y_k) \cdot h + \frac{h^2}{2 \Delta x} \cdot (f(x_k + \Delta x, y_k + y'_k \Delta x) - f(x_k, y_k))$$

которую преобразуем к виду:

$$\begin{aligned} y_{k+1} &= y_k + h \cdot \left[\left(1 - \frac{h}{2 \Delta x}\right) \cdot f(x_k, y_k) + \frac{h}{2 \Delta x} f(x_k + \Delta x, y_k + y'_k \Delta x) \right] = \\ &= y_k + h \cdot \left[\left(1 - \frac{h}{2 \Delta x}\right) \cdot f(x_k, y_k) + \frac{h}{2 \Delta x} f\left(x_k + \frac{\Delta x}{h} h, y_k + f(x_k, y_k) \frac{\Delta x}{h} h\right) \right] \end{aligned}$$

Введем следующие обозначения:

$$\alpha = \frac{h}{2 \Delta x}, \beta = 1 - \frac{h}{2 \Delta x}, \gamma = \frac{\Delta x}{h}, \delta = f(x_k, y_k) \frac{\Delta x}{h}$$

Эти обозначения позволяют записать предыдущее выражение в форме:

$$y_{k+1} = y_k + h \cdot [\beta \cdot f(x_k, y_k) + \alpha \cdot f(x_k + \gamma \cdot h, y_k + \delta \cdot h)]$$

Все введенные коэффициенты зависят от величины Δx и могут быть определены через коэффициент α , который в этом случае играет роль параметра:

$$\beta = 1 - \alpha, \gamma = \frac{1}{2\alpha}, \delta = f(x_k, y_k) \frac{2}{\alpha}$$

Окончательно схема Рунге-Кутты принимает вид:

$$y_{k+1} = y_k + h \cdot \left[(1 - \alpha) \cdot f(x_k, y_k) + \alpha \cdot f\left(x_k + \frac{h}{2}, y_k + f(x_k, y_k) \frac{h}{2\alpha}\right) \right]$$

Та же схема в форме разностного аналога уравнения (1):

$$\frac{y_{k+1} - y_k}{h} = (1 - \alpha) \cdot f(x_k, y_k) + \alpha \cdot f\left(x_k + \frac{h}{2}, y_k + f(x_k, y_k) \frac{h}{2\alpha}\right)$$

При $\alpha = 0$ получаем как частный случай уже известную схему Эйлера:

$$y_{k+1} = y_k + h \cdot f(x_k, y_k)$$

При $\alpha = 1$:

$$y_{k+1} = y_k + h \cdot f\left(x_k + \frac{h}{2}, y_k + \frac{h}{2} \cdot f(x_k, y_k)\right)$$

При $\alpha = 1$ проведение расчетов на очередном шаге интегрирования можно рассматривать как последовательность нижеследующих операций.

1. Вычисляется выражение, представляющее собой полушаг интегрирования по схеме Эйлера, то есть определяется приближенное значение искомой функции в точке $x_k + h/2$:

$$y_{k+1/2} = y_k + \frac{h}{2} \cdot f(x_k, y_k)$$

2. Для той же промежуточной точки находится приближенное значение производной:

$$y'_{k+1/2} = f\left(x_k + \frac{h}{2}, y_{k+1/2}\right)$$

3. Определяется уточненное значение функции в конечной точке всего шага, причем по схеме Эйлера с вычисленным на предыдущем шаге значением производной:

$$y_{k+1} = y_k + h \cdot y'_{k+1/2}$$

Геометрические построения (см. рис. 15.1) показывают, что получаемое в такой последовательности решение лежит «ближе» к истинному, чем вычисляемое по схеме Эйлера, то есть следует ожидать более высокой точности решения, получаемого методом Рунге-Кутты. Ранее мы назвали эту схему «модифицированным методом Эйлера».

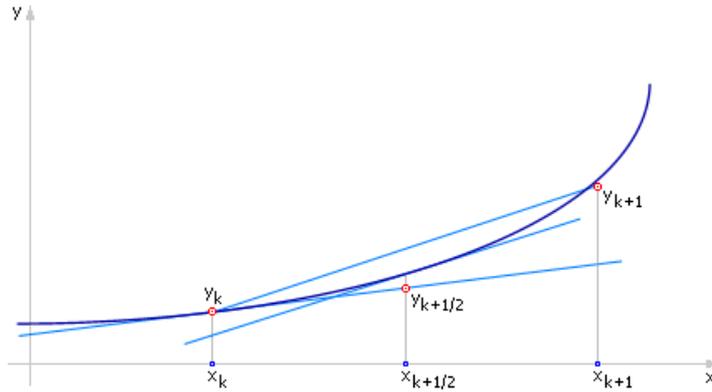


Рис. 15.1. Расчет на шаге методом Рунге-Кутты при значении параметра $\alpha = 1$

Рассмотрим схему при $\alpha = 0.5$ (геометрическая интерпретация результата приведена на рис. 15.2).

$$y_{k+1} = y_k + \frac{h}{2} \cdot \left[f(x_k, y_k) + f(x_k + h, y_k + h \cdot f(x_k, y_k)) \right]$$

1. Выполняется полный шаг метода Эйлера с целью определения приближенного значения искомой функции на конце отрезка интегрирования:

$$\hat{y}_{k+1} = y_k + h \cdot f(x_k, y_k)$$

2. Для этой же точки вычисляется приближенное значение производной:

$$y'_{k+1} = f(x_k + h, \hat{y}_{k+1})$$

3. Находится среднее значение двух производных, определенных на концах отрезка:

$$y'_{k+1/2} = \frac{1}{2} \cdot [y'_k + y'_{k+1}]$$

4. Вычисляется значение искомой функции в конечной точке всего шага по схеме Эйлера с усредненным значением производной:

$$y_{k+1} = y_k + h \cdot y'_{k+1/2}$$

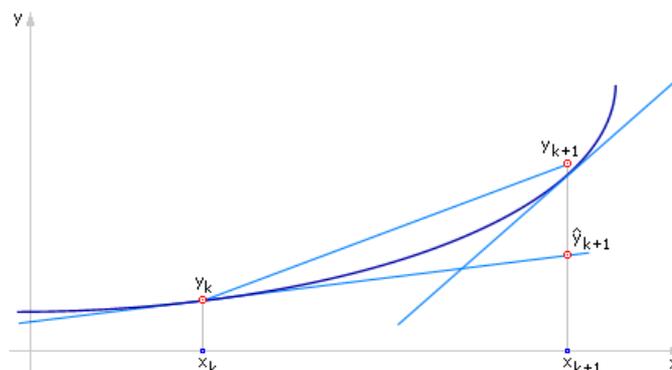


Рис. 15.2. Расчет на шаге методом Рунге-Кутты при значении параметра $\alpha = 0.5$

Иногда получающееся выражение называют **схемой (методом) Эйлера-Коши**. Геометрически понятно, что получаемый указанным способом результат также должен быть «ближе» к истинному решению, чем получаемый по схеме Эйлера.

Решить уравнение $dy/dx = -y$, $y(0) = 1$ методом Рунге-Кутты.

Поскольку правая часть дифференциального уравнения имеет вид: $f(x, y) = -y$, схема метода при $\alpha = 0.5$ представляется следующим образом:

$$\begin{aligned} y'_k &= f(x_k, y_k) = -y_k \\ \hat{y}_{k+1} &= y_k - h \cdot y_k = y_k \cdot (1 - h) \\ y'_{k+1} &= f(x_k + h, \hat{y}_{k+1}) = -y_k \cdot (1 - h) \\ y'_{k+1/2} &= \frac{y'_k + y'_{k+1}}{2} = -\frac{y'_k}{2} \cdot (2 - h) \\ y_{k+1} &= y_k + h \cdot y'_{k+1/2} = y_k - y_k \cdot \frac{h}{2} \cdot (2 - h) = y_k \frac{(h - 1)^2 + 1}{2} \end{aligned}$$

Построим последовательность значений искомой функции:

$$\begin{aligned} y_0 &= y(0) = 1 \\ y_1 &= y_0 \frac{(h - 1)^2 + 1}{2} = \frac{(h - 1)^2 + 1}{2} \\ y_2 &= y_1 \frac{(h - 1)^2 + 1}{2} = \left[\frac{(h - 1)^2 + 1}{2} \right]^2 \\ y_3 &= y_2 \frac{(h - 1)^2 + 1}{2} = \left[\frac{(h - 1)^2 + 1}{2} \right]^3 \\ &\dots \\ y_n &= \left[\frac{(h - 1)^2 + 1}{2} \right]^n \end{aligned}$$

Результаты получаемого численного решения для значения аргумента $x = 10$ при различных шагах интегрирования приведены в табл. 15.1. Три верные значащие цифры получены для шага $h = 0.01$.

Таблица 15.1.

Результаты численного решения y_n методом Рунге-Кутты второго порядка дифференциального уравнения $y' = -y$ с начальным условием $y(0) = 1$

Величина шага h	0.5	0.25	0.1	0.01	0.001	0.0001
Число шагов n	20	40	100	1000	10 000	100 000
$y_n \cdot 10^4$	0.827181	0.514756	0.462229	0.454076	0.454000	0.453999

Оценим погрешность аппроксимации уравнения (1) разностной схемой метода Рунге-Кутты, подставляя точное решение в разностный аналог исходного дифференциального уравнения и вычисляем невязку:

$$\psi_k = \frac{y(x_{k+1}) - y(x_k)}{h} - (1 - \alpha) \cdot f(x_k, y(x_k)) - \alpha \cdot f\left(x_k + \frac{h}{2\alpha}, y(x_k) + \frac{h}{2\alpha} f(x_k, y(x_k))\right)$$

Подставим разложения функций

$$\begin{aligned} y(x_{k+1}) &= y(x_k) + y'(x_k) \cdot h + y''(x_k) \cdot \frac{h^2}{2} + \dots \\ f\left(x_k + \frac{h}{2\alpha}, y(x_k) + \frac{h}{2\alpha} f(x_k, y(x_k))\right) &= \\ &= f(x_k, y(x_k)) + \frac{h}{2\alpha} \cdot \left[\frac{\partial f(x_k, y(x_k))}{\partial x} + \frac{\partial f(x_k, y(x_k))}{\partial y} f(x_k, y(x_k)) \right] + \dots \end{aligned}$$

в полученное выражение:

$$\begin{aligned}
\psi_k &= \frac{y(x_k) + y'(x_k) \cdot h + y''(x_k) \cdot \frac{h^2}{2} + K - y(x_k)}{h} - (1 - \alpha)f(x_k, y(x_k)) - \\
&\quad - \alpha \cdot \left\{ f(x_k, y(x_k)) + \frac{h}{2\alpha} \cdot \left[\frac{\partial f(x_k, y(x_k))}{\partial x} + \frac{\partial f(x_k, y(x_k))}{\partial y} f(x_k, y(x_k)) \right] + \dots \right\} = \\
&= y'(x_k) + y''(x_k) \cdot \frac{h}{2} - f(x_k, y(x_k)) + \alpha \cdot f(x_k, y(x_k)) - \alpha \cdot f(x_k, y(x_k)) - \\
&\quad - \frac{h}{2} \cdot \left[\frac{\partial f(x_k, y(x_k))}{\partial x} + \frac{\partial f(x_k, y(x_k))}{\partial y} f(x_k, y(x_k)) \right] + O(h^2) = \\
&= y'(x_k) - f(x_k, y(x_k)) + \frac{h}{2} \cdot \left[y''(x_k) - \frac{\partial f(x_k, y(x_k))}{\partial x} - \frac{\partial f(x_k, y(x_k))}{\partial y} f(x_k, y(x_k)) \right] + O(h^2)
\end{aligned}$$

Учитывая уравнение (1), а также выражение для производной

$$y''(x_k) = (y'(x_k))' = (f(x_k, y(x_k)))' = \frac{\partial f(x_k, y(x_k))}{\partial x} + \frac{\partial f(x_k, y(x_k))}{\partial y} f(x_k, y(x_k))$$

окончательно получаем, что $\psi_k = O(h^2)$, метод Рунге-Кутты, независимо от значения параметра α , имеет второй порядок аппроксимации.

Методы Рунге-Кутты третьего и четвертого порядков

Рассмотрим две различные схемы Рунге-Кутты, предназначенные для численного решения обыкновенных дифференциальных уравнений первого порядка и имеющие *третий* порядок аппроксимации:

$$\left\{ \begin{array}{l} K_1 = f(x_k, y_k) \\ K_2 = f(x_k + \frac{h}{2}, y_k + \frac{h}{2} K_1) \\ K_3 = f(x_k + h, y_k - h K_1 + 2h K_2) \\ y_{k+1} = y_k + \frac{h}{6} \cdot (K_1 + 4K_2 + K_3) \end{array} \right. \quad \left\{ \begin{array}{l} K_1 = f(x_k, y_k) \\ K_2 = f(x_k + \frac{h}{3}, y_k + \frac{h}{3} K_1) \\ K_3 = f(x_k + \frac{2h}{3}, y_k + \frac{2h}{3} K_2) \\ y_{k+1} = y_k + \frac{h}{4} \cdot (K_1 + 3K_3) \end{array} \right.$$

И две схемы Рунге-Кутты, имеющие *четвертый* порядок аппроксимации:

$$\left\{ \begin{array}{l} K_1 = f(x_k, y_k) \\ K_2 = f(x_k + \frac{h}{2}, y_k + \frac{h}{2} K_1) \\ K_3 = f(x_k + \frac{h}{2}, y_k + \frac{h}{2} K_2) \\ K_4 = f(x_k + h, y_k + h K_3) \\ y_{k+1} = y_k + \frac{h}{6} \cdot (K_1 + 2K_2 + 2K_3 + K_4) \end{array} \right. \quad \left\{ \begin{array}{l} K_1 = f(x_k, y_k) \\ K_2 = f(x_k + \frac{h}{4}, y_k + \frac{h}{4} K_1) \\ K_3 = f(x_k + \frac{h}{2}, y_k + \frac{h}{2} K_2) \\ K_4 = f(x_k + h, y_k + h K_1 - 2h K_2 + 2h K_3) \\ y_{k+1} = y_k + \frac{h}{6} \cdot (K_1 + 4K_3 + K_4) \end{array} \right.$$

Решить методом Рунге-Кутты четвертого порядка уравнение $dy/dx = -y$, $y(0) = 1$.

В соответствии с приведенными выше соотношениями определяем коэффициенты:

$$\begin{aligned}
 K_1 &= -y_k \\
 K_2 &= -\left(y_k + \frac{h}{2} K_1\right) = -y_k \cdot \left(1 - \frac{h}{2}\right) \\
 K_3 &= -\left(y_k + \frac{h}{2} K_2\right) = -y_k \cdot \left(1 - \frac{h}{2} \cdot \left(1 - \frac{h}{2}\right)\right) \\
 K_4 &= -\left(y_k + h K_3\right) = -y_k \cdot \left[1 - h \cdot \left(1 - \frac{h}{2} \cdot \left(1 - \frac{h}{2}\right)\right)\right] \\
 y_{k+1} &= y_k + \frac{h}{6} \cdot \left(K_1 + 2K_2 + 2K_3 + K_4\right) = y_k \cdot \left(1 + \frac{h^4 - 4h^3 + 12h^2 - 24h}{24}\right)
 \end{aligned}$$

Построим последовательность значений искомой функции:

$$\begin{aligned}
 y_0 &= y(0) = 1 \\
 y_1 &= y_0 \cdot \left(1 + \frac{h^4 - 4h^3 + 12h^2 - 24h}{24}\right) = 1 \cdot \left(1 + \frac{h^4 - 4h^3 + 12h^2 - 24h}{24}\right) \\
 y_2 &= y_1 \cdot \left(1 + \frac{h^4 - 4h^3 + 12h^2 - 24h}{24}\right) = 1 \cdot \left(1 + \frac{h^4 - 4h^3 + 12h^2 - 24h}{24}\right)^2 \\
 y_3 &= y_2 \cdot \left(1 + \frac{h^4 - 4h^3 + 12h^2 - 24h}{24}\right) = 1 \cdot \left(1 + \frac{h^4 - 4h^3 + 12h^2 - 24h}{24}\right)^3 \\
 y_n &= \left(1 + \frac{h^4 - 4h^3 + 12h^2 - 24h}{24}\right)^n
 \end{aligned}$$

Результаты получаемого численного решения для значения аргумента $x = 10$ при различных шагах интегрирования приведены в табл. 15.2. Три верные цифры получены для шага $h = 0.25$.

Таблица 15.2.
Результаты численного решения y_n методом Рунге-Кутты четвертого порядка дифференциального уравнения $y' = -y$ с начальным условием $y(0) = 1$

Величина шага h	0.5	0.25	0.1	0.01	0.001	0.0001
Число шагов n	20	40	100	1000	10 000	100 000
$y_n \cdot 10^4$	0.457608	0.454181	0.454003	0.453999	0.453999	0.453999

Сравнение таблиц 15.1 и 15.2 с решениями одной и той же задачи позволяет сделать вывод, что более высокая степень аппроксимации дифференциального уравнения разностным аналогом позволяет получать более точное решение при более крупном шаге и, следовательно, меньшем числе шагов, то есть приводит к снижению требуемых ресурсов ЭВМ.

На сегодняшний день для грубого расчета вычисления производятся методом Эйлера, для точного расчета — методом Рунге-Кутты.

16. МЕТОДЫ ПРОГНОЗА И КОРРЕКЦИИ (ИТЕРАЦИОННЫЕ МЕТОДЫ)

Изученные ранее методы обладали одной важной особенностью — каждому методу соответствует обычно определенный класс точности, который мы обозначали как O^i . Например, метод Эйлера обладал первым классом точности O^1 . Это означало, что с уменьшением шага в 10 раз (на порядок) точность результата повышается тоже в 10 раз (на один порядок). Метод Рунге-Кутты обладает 4 порядком точности — O^4 , при уменьшении шага в 10 раз, результат улучшается в 10 000 раз. Поскольку этот метод по сравнению с методом Эйлера использует всего в 4 раза больше вычислений, то использование его более выгодно. На сегодняшний день известны методы до 8 порядка точности (например, метод Prince Dortmund), хотя одновременно стоит иметь в виду, что написание алгоритмов для них — задача достаточно трудная. Достоинством всех этих алгоритмов является то, что объем вычислений для них заранее известен [9].

Если требуется достичь любой точности на шаге, то следует использовать методы прогноза и коррекции. Этот подход состоит в том, что расчет траектории, задаваемой уравнением, на каждом шаге происходит многократно. А именно, сначала происходит расчет приближенного значения функции на конце шага какой-либо простой формулой (например, методом Эйлера), далее в этой точке вычисляется производная, и расчет происходит снова из начальной точки на шаге, но с уточненным значением производной. Последняя операция — уточнения производной и значения функции на конце шага — происходит многократно на каждом шаге, то есть до тех пор, пока вычисленные значения (функции и производной в конце шага) не перестанут меняться или будут меняться уже незначительно, меньше чем задаваемая заранее величина ε . Только тогда можно сказать, что точность ε достигнута.

За счет итерационной процедуры на каждом отдельном шаге можно достичь любой, наперед заданной точности ε . За такое достоинство метода приходится платить: к сожалению, невозможно сказать заранее, сколько итераций потребуется для достижения на шаге заданной точности ε . Поэтому такие методы нельзя, например, использовать в системах реального времени.

Рассмотрим для примера два метода из этого класса. Как и ранее задача состоит в нахождении функции $y(t)$ из дифференциального уравнения $dy/dt = f(y, x, t)$ или множества функций из системы таких уравнений.

Метод Эйлера с итерациями

1) Предсказывающая формула вычисляет (прогнозирует) значение функции на правом конце шага: $y_{k+1} = y_k + f_k \cdot \Delta t$.

2) Рассчитывается производная в точке $k + 1$ подстановкой y в исходное уравнение в $k + 1$ точке: $f_{k+1} = f(t + \Delta t, y_{k+1})$.

3) Уточняющая формула, используя старое значение производной (с шага 1) и уточненное с шага 2, дает уточненное значение y_{k+1} : $y_{k+1} = y_k + (f_k + f_{k+1}) \cdot \Delta t / 2$. Здесь же производится подсчет итераций счетчиком i : $i := i + 1$.

4) Проверка точности: $|y_{k+1}^{i\text{-я итерация}} - y_{k+1}^{(i+1)\text{-я итерация}}| \leq \varepsilon$. Если условие выполнено и точность ε достигнута, то переходим на следующий шаг 5), иначе осуществляется переход на шаг 2) и процесс уточнения повторяется с новыми значениями y и f , причем их старое значение берется с предыдущей итерации.

5) Подготовка к новому шагу: изменение счетчика времени t на величину Δt и изменение номера шага k :
 $t := t + \Delta t$
 $k := k + 1$.

6) Проверка окончания расчета: $t \leq T$. Если условие выполняется, расчет продолжается для следующей точки, переход на 1), иначе — конец.

Метод Милна

1) По предсказывающей формуле вычисляется грубое значение y на правом конце интервала: y_{k+1} : $y_{k+1} = y_{k-3} + 4/3 \cdot (2 \cdot f_k - f_{k-1} + 2 \cdot f_{k-2}) \cdot \Delta t$.

2) Рассчитывается производная в $k + 1$ точке: $f_{k+1} = f(t + \Delta t, y_{k+1})$.

3) Снова рассчитывается y_{k+1} по уточненной формуле, используя уже новое значение производной в точке $k + 1$: $y_{k+1} = y_{k-1} + 1/3 \cdot (f_{k+1} + 4 \cdot f_k + f_{k-1}) \cdot \Delta t$.

4) Рассчитывается производная в $k + 1$ точке с учетом вновь вычисленного более точного значения y_{k+1} : $f_{k+1} = f(t + \Delta t, y_{k+1})$. Здесь же производится подсчет итераций счетчиком i : $i := i + 1$.

5) Проверка точности: $|y_{k+1}^{i\text{-я итерация}} - y_{k+1}^{(i+1)\text{-я итерация}}| \leq \varepsilon$. Если условие выполнено, и точность ε достигнута, то переходим на следующий шаг 6), иначе осуществляется переход на шаг 3) и процесс уточнения повторяется с новыми значениями y и f , причем их старое значение берется с предыдущей итерации.

6) Подготовка к новому шагу: изменение счетчика времени t , изменение номера шага k :
 $t := t + \Delta t$
 $k := k + 1$.

7) Проверка окончания расчета: $t \leq T$. Если условие выполняется, то расчет продолжается для следующей точки, и осуществляется переход на шаг 1), иначе — конец.

17.МОДЕЛИРОВАНИЕ СИСТЕМ С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ

Если объект характеризуется некоторым параметром, различным по своему значению в разных точках объекта, то можно сказать, что значения такого параметра распределены (по объекту). Если таких параметров несколько, то объект рассматривается как система с распределенными параметрами. Для осуществления расчетов систему в таком случае удобно разбить на элементарные объемы (слои). Покажем это на примере.

Пример 1. Рассмотрим процесс сушки материала (см. рис. 17.1). Допустим, что материал представляет собой кучу сырья, над которой установлен нагреватель. В процессе сушки тепло неравномерно проникает в глубь кучи, и в различных слоях материал имеет различную температуру. То есть в данном примере можно сказать, что параметр температуры распределен по объекту «куча». Опишем каждый слой своим уравнением, используя процедуру построения модели для отдельного слоя из лекции 11.

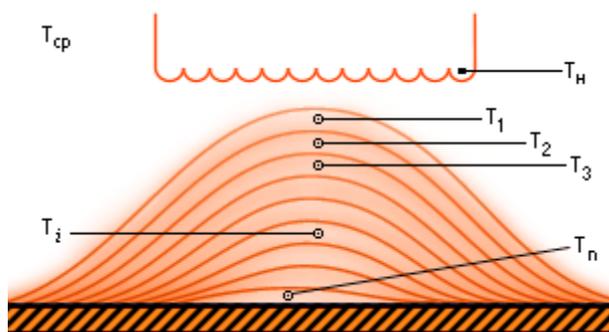


Рис. 17.1. Схема технологического процесса сушки материала, разбитого на слои

Изменение температуры в первом слое (T_n — температура нагревателя, $T_{сп}$ — температура среды):

$$\frac{dT_1}{dt} = k_1 \cdot (T_n - T_1) + k_2 \cdot (T_2 - T_1) + k_3 \cdot (T_{сп.} - T_1) \quad (1)$$

Изменение температуры в i -ом слое:

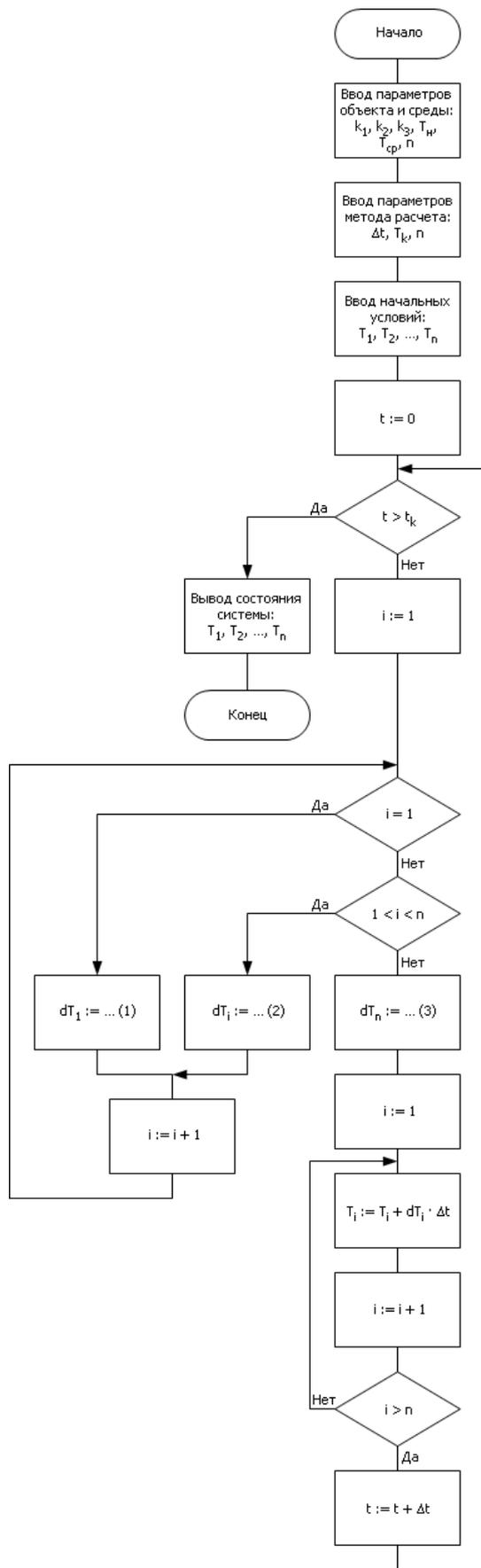
$$\frac{dT_i}{dt} = k_2 \cdot (T_{i+1} - T_i) + k_2 \cdot (T_{i-1} - T_i) \quad (2)$$

Изменение температуры в n -ом слое:

$$\frac{dT_n}{dt} = k_2 \cdot (T_{n-1} - T_n) \quad (3)$$

Тогда поведение системы «кучи сырья» опишется системой дифференциальных уравнений, каждое из которых опишет отдельный слой «кучи».

Алгоритм расчета такой системы показан на рис. 17.2. Особенностью алгоритма является то, что он содержит в дополнение к циклу по времени (см. алгоритм на рис. 10.5) вложенный цикл по номеру слоя. На каждом такте необходимо отдельно просчитать изменения в каждом из слоев «кучи». Распределенность объекта (введение дополнительной координаты) имитируется дополнительным циклическим блоком.



18.МОДЕЛИРОВАНИЕ СИСТЕМ С РАСПРЕДЕЛЕННЫМИ ПАРАМЕТРАМИ ПРИ ПЕРЕМЕЩЕНИИ МАТЕРИАЛЬНЫХ МАСС

Еще раз усложним задачу из предыдущей лекции.

Рассмотрим процессы нагрева, сушки и одновременно перемещения неоднородной массы. Теперь (см. рис. 18.1) будем массу не только сушить, но и в процессе сушки передвигать ее по конвейеру к потребителю. Пусть длина конвейера — L , а линейная скорость его перемещения — V . Для реалистичности усовершенствуем модель, учитывая два параметра «кучи» — температуру T и влажность W каждого ее участка. Исходное состояние сырья, находящегося в бункере 1, обозначим как T_0 и W_0 .

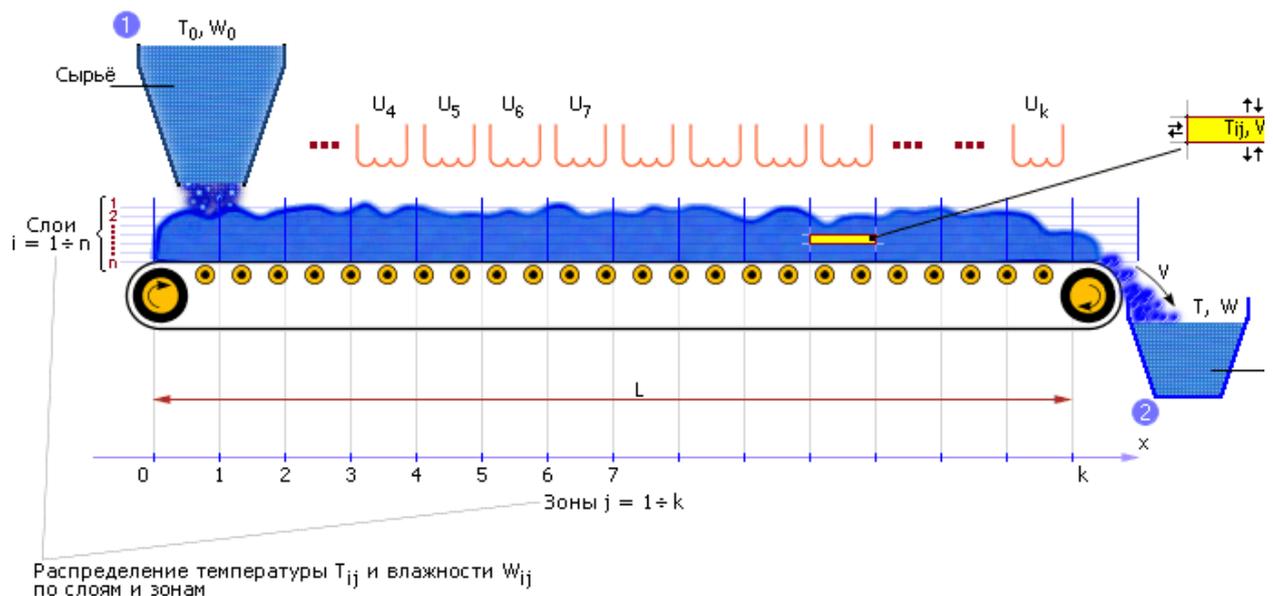


Рис. 18.1. Схема процесса сушки и перемещения сырья

Для отличия различных фрагментов массы не только по слоям, но и на различных участках конвейера введем деление массы на зоны. Пусть всего будет n слоев и k зон [9].

Время h , в течение которого лента передвигается на одну зону, составляет:

$$h = \frac{L}{k} \cdot \frac{1}{V}$$

Идея моделирования состоит в том, что мы будем осуществлять при временно неподвижном конвейере нагрев и сушку (расчет температуры и влажности) по зонам и слоям, а затем перемещать сырье из зоны в зону одномоментно, имитируя перемещение конвейера.

То есть для компьютерной имитации на цифровых машинах полезно разделить процесс сушки и процесс перемещения. Так на одном такте времени Δt имеет смысл осуществить несколько микротактов сушки a . Именно столько микротактов лента стоит на месте. После этого происходит резкий сдвиг конвейера, и сырье из одной зоны попадает в следующую зону рывком. После этого снова осуществляется процесс сушки в новом положении конвейера.

Связь величин Δt , h и a задается формулой: $\Delta t = h \cdot a$ или так:

$$\Delta t = \frac{L \cdot a}{k \cdot V}$$

Далее: $T_{\text{ср.}}$ — температура среды; T_{H} — температура нагревателя ($T_{\text{H}j} = g \cdot U_j$); $T_{\text{тп.}}$ — температура транспортной ленты.

$$\frac{dT_{1j}}{dt} = k_1 \cdot (T_{\text{H}j} - T_{1j}) + k_2 \cdot (T_{\text{ср.}} - T_{1j}) + k_3 \cdot (T_{2j} - T_{1j}) \quad (1)$$

$$\dots$$

$$\frac{dT_{ij}}{dt} = k_3 \cdot (T_{i+1,j} - T_{ij}) + k_3 \cdot (T_{i-1,j} - T_{ij}) + k_3 \cdot (T_{2j} - T_{1j}) \quad (2)$$

$$\dots$$

$$\frac{dT_{nj}}{dt} = k_4 \cdot (T_{\text{тп.}} - T_{nj}) + k_3 \cdot (T_{n-1,j} - T_{nj}) \quad (3)$$

$$\frac{dW_{1j}}{dt} = k_5 \cdot (W_{2j} - W_{1j}) + k_6 \cdot (T_{\text{ср.}} - T_{1j}) \cdot W_{1j} \quad (4)$$

$$\dots$$

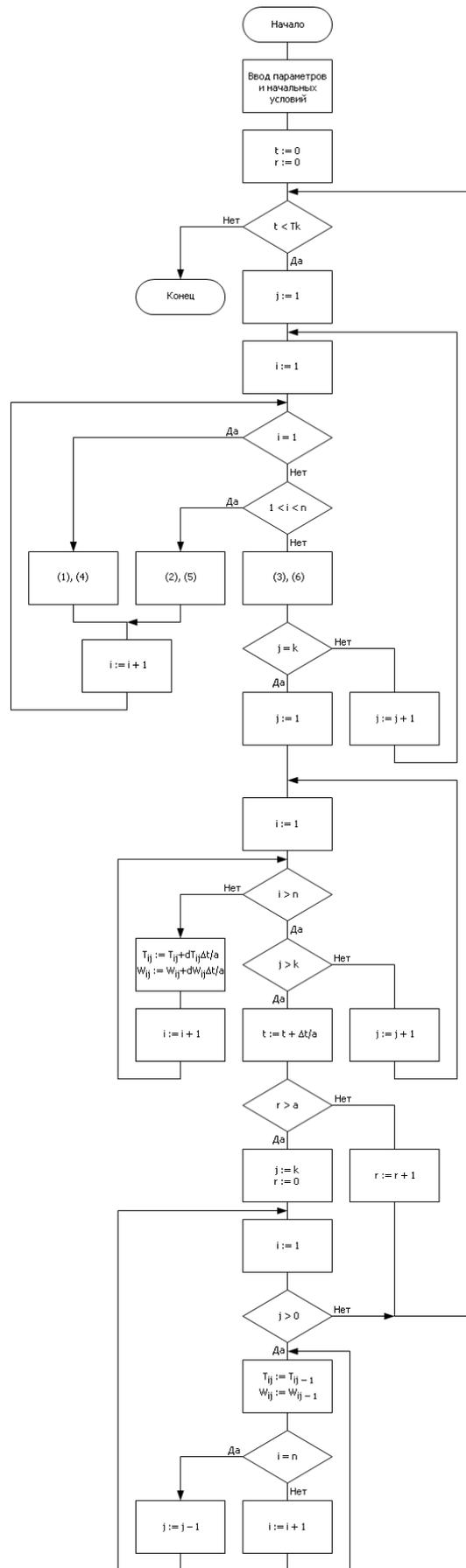
$$\frac{dW_{ij}}{dt} = k_6 \cdot (W_{i+1,j} - W_{ij}) + k_6 \cdot (W_{i-1,j} - W_{ij}) \quad \text{или} \quad [k_6 \cdot (T_{i+1,j} - T_{ij}) \cdot W_{ij}] \quad (5)$$

$$\dots$$

$$\frac{dW_{nj}}{dt} = k_6 \cdot (W_{n-1,j} - W_{nj}) \quad (6)$$

Тогда поведение системы «кучи сырья» опишется системой дифференциальных уравнений, каждое из которых опишет отдельный слой и зону «кучи» как по температуре, так и по влажности.

Алгоритм расчета такой системы показан на рис. 18.2. Особенностью алгоритма является то, что он содержит, в дополнение к циклу по времени и по слою (см. алгоритм на рис. 17.2), еще один вложенный цикл по номеру зоны. На каждом такте необходимо отдельно просчитать изменения не только в каждом из слоев «кучи», но в каждой зоне.



19. УРАВНЕНИЯ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Из математики и физики известны случаи аналитического описания систем с распределенными параметрами уравнениями в частных производных. Ими являются уравнения диффузии, тепломассопереноса и другие. Рассмотрим варианты их имитации на компьютере [9].

Системы с распределенными параметрами - это системы, в которых значение некоторого изучаемого параметра изменяется не только во времени, но и в пространстве (одно-, двух- или многомерном), меняясь от точки к точке пространства по какому-то закону. Примером может служить энергия, концентрация вещества, напряженность поля и другие физические величины. Неизменным изучаемый параметр может считаться только в бесконечно малой области пространства. Поэтому систему с распределенными параметрами часто представляют как систему из упорядоченного в пространстве множества элементов, внутри которых (в каждой точке отдельного взятого элемента) описываемый параметр одинаков, а у разных элементов различен. Элементы системы образуют пространственную структуру.

В качестве примера приведем проект для изучения теплопроводности стены дома. На рис. 19.1 представлена реализация такой системы в среде «Stratum-2000». Схема собрана из элементов. Элементы двухмерной системы имитируют кирпичи, из которых сложена стена. На рис. 19.1 видны кирпичи, сделанные из разного материала (что обозначено желтым, голубым, красным цветами) и уложенные в определенном порядке, что, по мнению автора схемы, улучшает теплоизоляцию помещения. Серые и зеленые элементы на краях стены задают краевые условия, имитируя окружающую среду. Рядом показано для примера, как элементы в такой системе связаны друг с другом. По связям соседние элементы в процессе расчета обмениваются информацией между собой о температуре. Заметим, что в данном проекте, который имеет вид конструктора, пользователь схемы может менять параметры элементов системы, условия, конструкцию стены, порядок укладки (топологию), что при решении задачи аналитическими методами крайне затруднительно и требует применения методов моделирования.

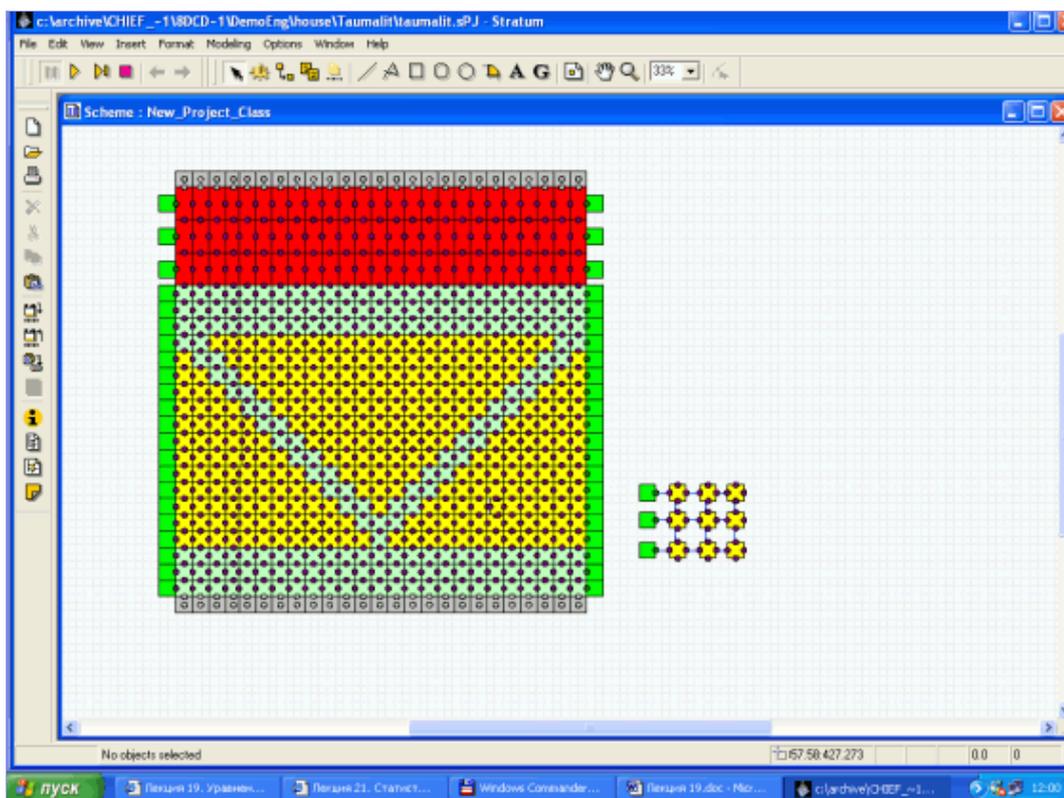


Рис. 19.1. Двухмерная система с распределенными параметрами для исследования свойств теплопроводности стены дома

Существуют типовые уравнения, описывающие отдельные свойства систем с распределенными параметрами. Рассмотрим их.

Уравнение диффузии

Уравнение диффузии описывает распространение (растекание) со временем по протяженному телу некоторой субстанции, например, тепла или концентрации. В одномерном случае тело представляется протяженным вдоль оси x .

На рис. 19.2 показан пример распределения вдоль оси x такого параметра как температура T . Из обычного опыта хорошо известно, что в каждый момент времени t температура T на разных участках тела x имеет разные значения, то есть меняется в зависимости от участка и времени. Должен существовать закон, по которому изменяется величина этого параметра T как функции от (x, t) . Для температуры этот закон чаще всего задается уравнением диффузии.

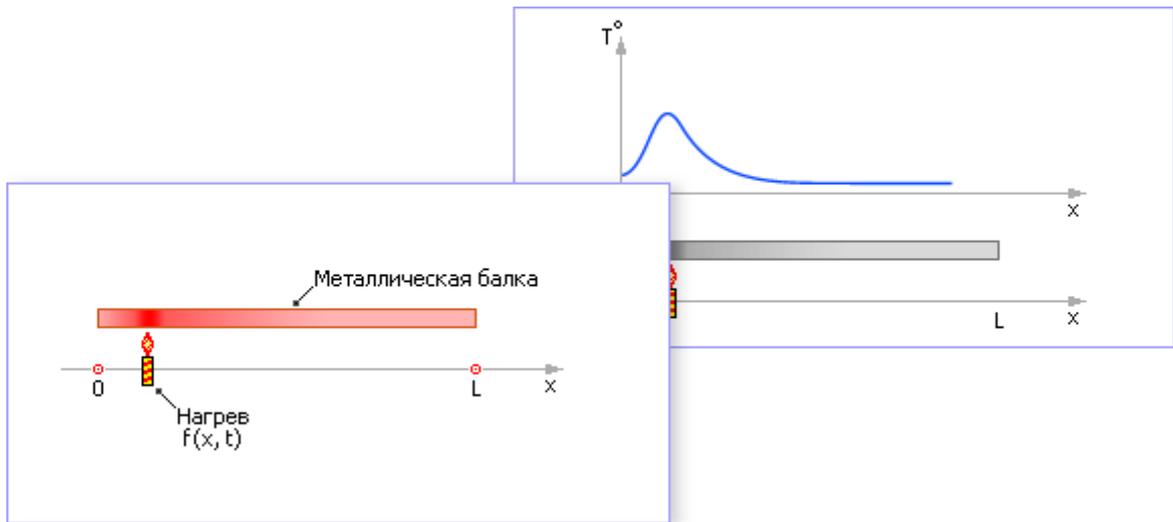


Рис. 19.2. Пример применения уравнения диффузии к описанию процесса нагрева протяженного одномерного тела

Если изменяемый параметр (в общем случае) обозначить как y , время, в течение которого отслеживаются изменения параметра, обозначить как t , а ось, вдоль которой происходят изменения параметра, как x , то уравнение диффузии имеет вид:

$$\frac{\partial y}{\partial t} = \alpha \cdot \frac{\partial^2 y}{\partial x^2} + f(x, t)$$

и обычно дополняется условиями — значениями переменной y на краях и границах: на левом краю $x = 0$, на правом краю $x = L$, на границе — начальные условия ($t = 0$):

$$y(x, 0) = f_1(x),$$

$$y(0, t) = f_2(t),$$

$$y(L, t) = f_3(t),$$

где $f_1(x)$, $f_2(t)$ и $f_3(t)$ — заданные функции.

На рис. 19.3 представлен схематически вид области, для которой определены граничные и начальные условия. Функции $f_1(x)$, $f_2(t)$, $f_3(t)$ и само уравнение диффузии предопределяют поведение функции $y(x, t)$ внутри этой области, чей полный вид обычно надо определить. Если на схеме дополнительно построить ось y (см. рис. 19.4), то визуально на рисунке можно отобразить и сам вид функций. На рисунке четко видно, что в углах схемы значения задаваемых функций должно совпадать.

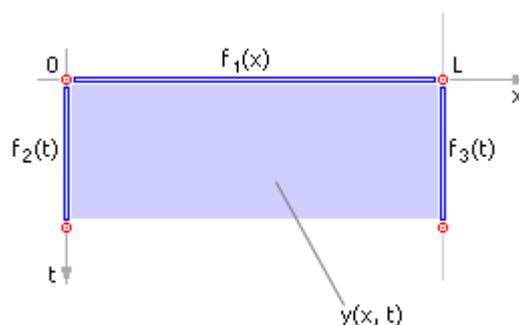


Рис. 19.3. Схема задания начальных и краевых условий для систем с распределенными параметрами в осях x , t (пример)

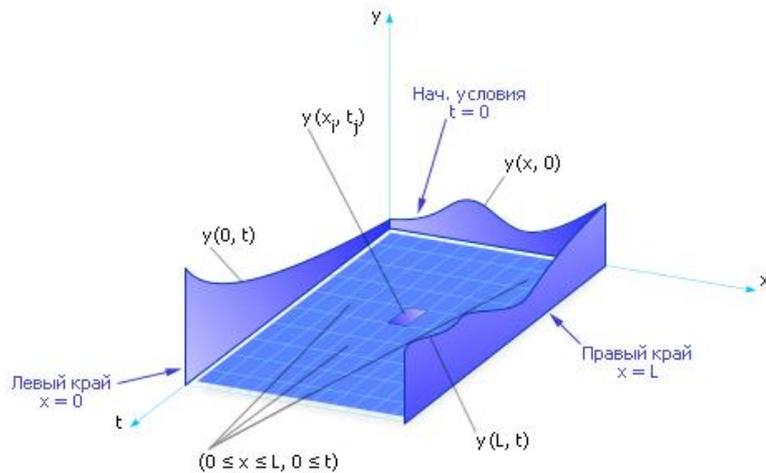


Рис. 19.4. Схема задания начальных и краевых условий для систем с распределенными параметрами в осях x , t , y (пример)

Коэффициент α имеет смысл коэффициента теплопроводности; $f(x, t)$ имеет смысл функции, описывающей работу источников и стоков тепла.

Величина y , описывающая распределение температуры, является функцией двух переменных — протяженности тела x и времени t : $y(x, t)$. Графически функция представляется поверхностью (см. рис. 19.5) или набором изолиний (см. рис. 19.6), вид которых обычно требуется определить.

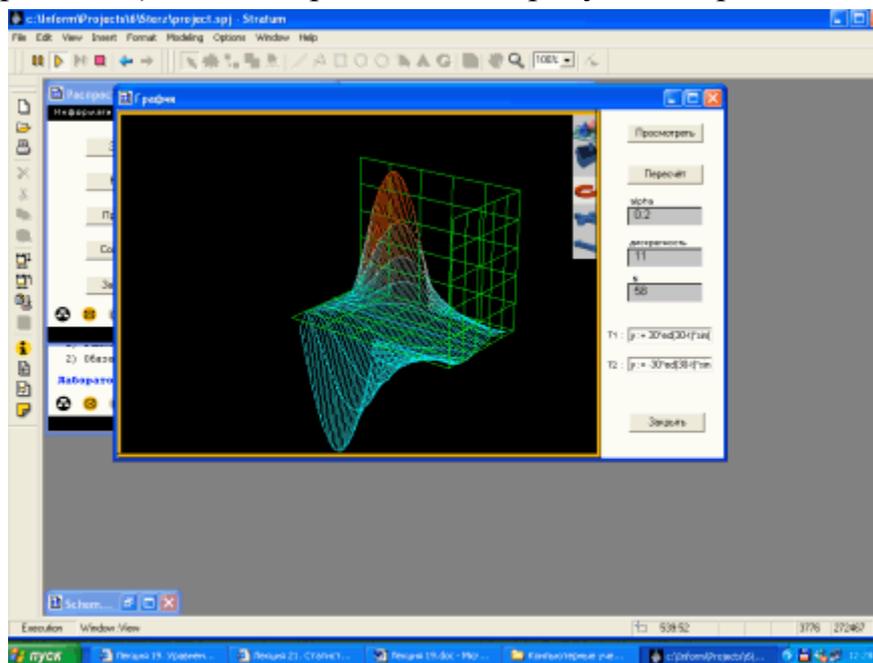


Рис. 19.5. Расчет распространения тепла в одномерном стержне со временем

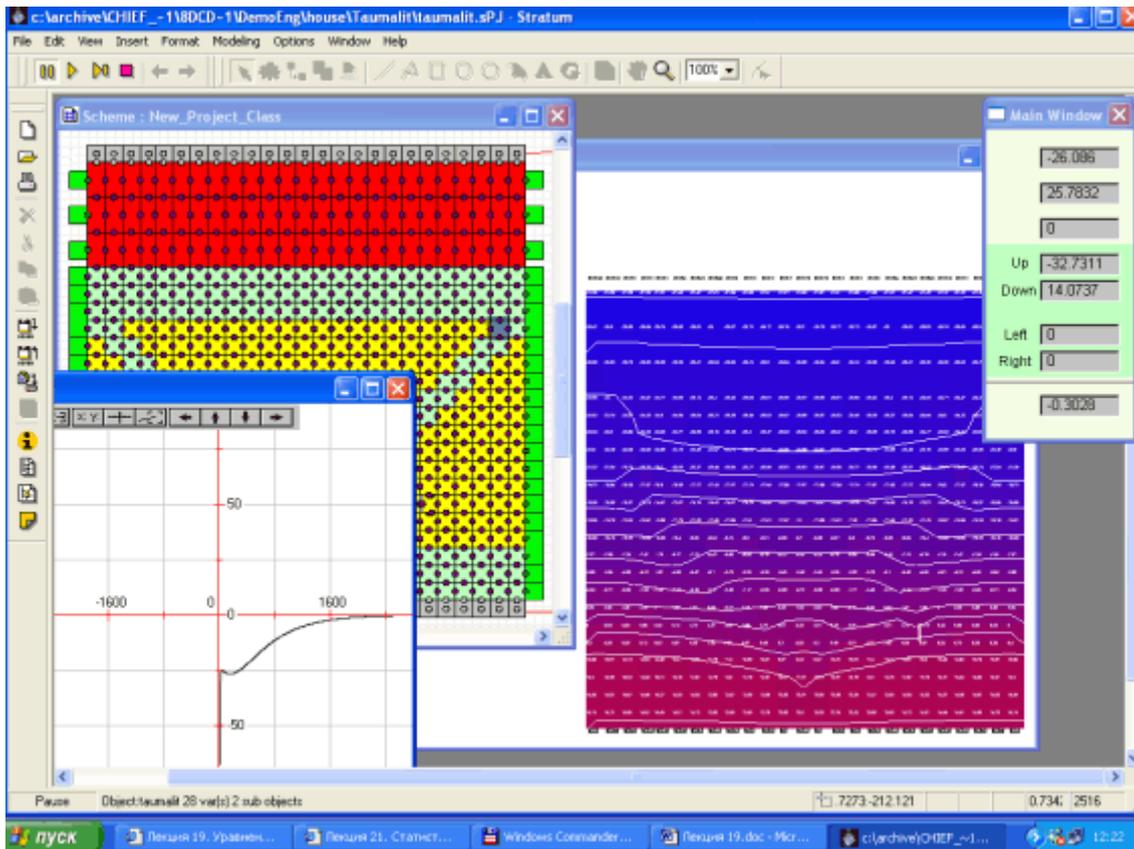


Рис. 19.6. Распространение тепла по элементам двумерной системы (расчет статической задачи, представленной на рис. 19.1). В отдельном окне видна оценка точности расчета в зависимости от количества итераций

Если заменить выражения производных их дискретным аналогом, то в разностном виде уравнение будет выглядеть так:

$$\frac{y(x, t) - y(x, t - \Delta t)}{\Delta t} = \alpha \cdot \frac{y(x + \Delta x, t - \Delta t) - 2y(x, t - \Delta t) + y(x - \Delta x, t - \Delta t)}{\Delta x^2} + f(x, t - \Delta t)$$

или, выражая неизвестное через известные величины:

$$y(x, t) = \frac{\Delta t \cdot \alpha}{\Delta x^2} y(x - \Delta x, t - \Delta t) + \left(1 - \frac{2\Delta t \cdot \alpha}{\Delta x^2}\right) \cdot y(x, t - \Delta t) + \frac{\Delta t \cdot \alpha}{\Delta x^2} y(x + \Delta x, t - \Delta t) + \Delta t \cdot f(x, t - \Delta t)$$

или

$$y(x, t) = a \cdot y(x - \Delta x, t - \Delta t) + b \cdot y(x, t - \Delta t) + c \cdot y(x + \Delta x, t - \Delta t) + f(x, t - \Delta t) \cdot \Delta t$$

В результате получена расчетная формула, реализуемая на цифровой вычислительной машине. Благодаря этой формуле можно рассчитать значение параметра y в любой точке (x, t) .

Назовем значение $y(x, t)$ узлом расчета. Тогда схематично расчет выглядит как сетка узлов на поле, составленном из частей тела и отрезков времени (см. рис. 19.7). Сама формула расчета одного узла зависит от состояния трех узлов (левого $y(x - \Delta x, t - \Delta t)$, правого $y(x + \Delta x, t - \Delta t)$, собственного $y(x, t - \Delta t)$) в предыдущий $(t - \Delta t)$ момент времени и напоминает треугольный шаблон. До начала расчета известны состояния всех узлов для $t = 0$. Применяя формулу последовательно ко всем узлам для следующего момента времени, можно определить температуру во всех узлах следующего временного слоя $(t + \Delta t)$. Кроме самого левого и самого правого узлов — их состояние вычислено быть не может, но оно задано крайевыми условиями.

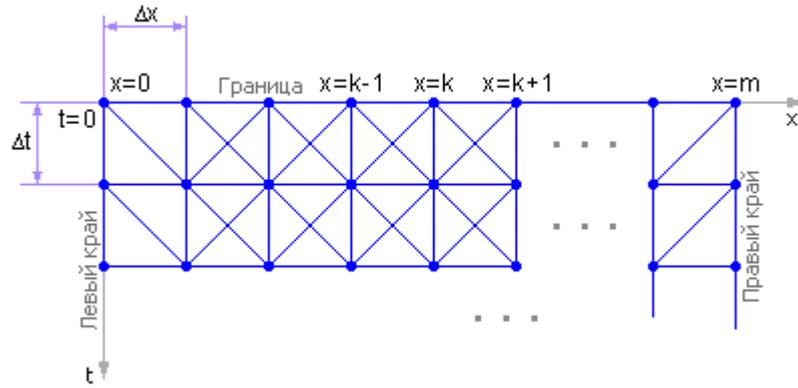


Рис. 19.7. Схема расчета одномерной динамической системы с распределенными параметрами методом конечных разностей с использованием явного шаблона

Если процедуру повторять, переходя от одной точки тела x к другой, и далее от одного временного слоя к другому, то по данной формуле можно вычислить значение температуры в любой части тела в любой момент времени. Таким образом, расчетом покрывается все поле ($L \times T$) (см. рис. 19.7). Последовательное определение неизвестных значений в данном случае возможно, потому что шаблон имеет вид явного выражения — единственное неизвестное в формуле выражено через ранее вычисленные значения.

Заметим, что при больших значениях производных и больших значениях шагов расчет может дать неверные решения. Решения могут оказаться неточными или даже неустойчивыми (качественно неверными) (см. лекцию 10. «Численные методы интегрирования дифференциальных уравнений. Метод Эйлера»).

Условие устойчивости для треугольного шаблона при решении уравнения диффузии: $\Delta x / \Delta t > \alpha$ (см. подробнее рис. 19.12).

При моделировании возможно применение других разностных формул (шаблонов) (см. рис. 19.8). При выборе шаблона необходимо принимать во внимание: явный шаблон или нет, какую он обеспечивает точность и при каких значениях шагов он обеспечивает устойчивость расчета. Так, например, шаблон в виде прямоугольника — неявный: в одной расчетной формуле содержится сразу две неизвестные величины. Поэтому при использовании такого шаблона необходимо решать систему алгебраических уравнений размером ($L \cdot T$).

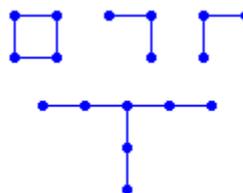


Рис. 19.8. Схемы типовых шаблонов для решения уравнений с распределенными параметрами

На практике устойчивости, а далее — точности добиваются получением решений с использованием разных шаблонов и разных значений шага. Если

значения искомой переменной, вычисленные с шагом h и с шагом $h/2$, отличаются в узлах с одинаковыми индексами не более чем на 1—5%, то вычисленное значение принимают за приближенное решение задачи. Иначе уменьшают шаг еще в два раза, и процедуру оценки повторяют. Свойства уравнения диффузии отражены на рис. 19.9 и заключаются в том, что при возникновении неоднородности в какой-то из частей тела со временем тепло за счет процессов теплообмена перетекает в соседние области. Температуры соседних областей выравниваются, усредняются. Темп процесса зависит от величины коэффициента теплопроводности.

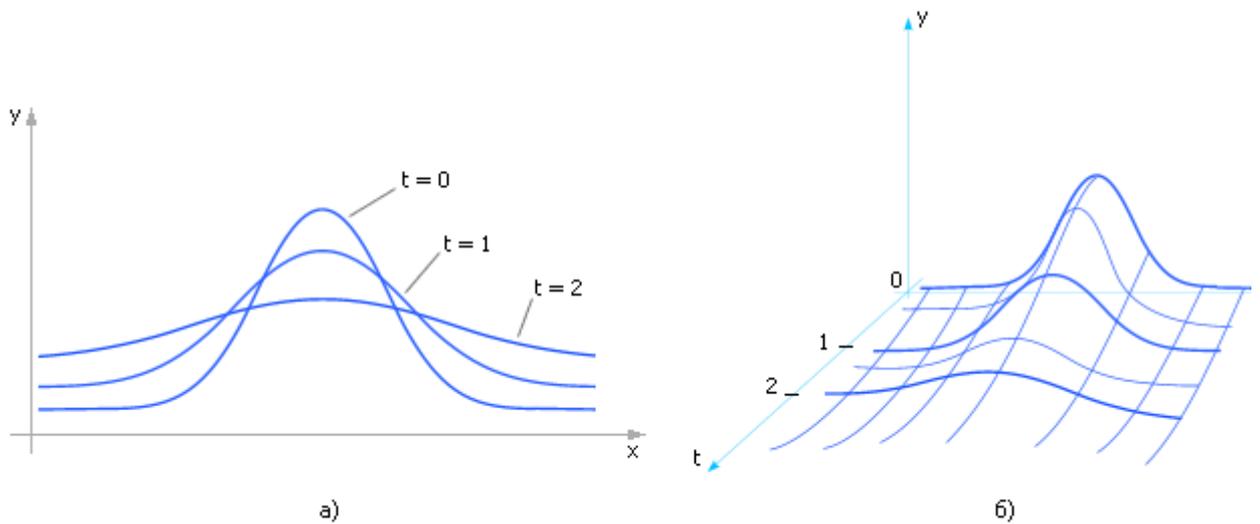


Рис. 19.9. Характерный вид решения уравнения диффузии. На рисунке отражено изменение распределения параметра y вдоль оси x во времени. а) — в двух осях; б) — в трех осях

Если принять условие, что задача стационарная, то есть процессы протекают так долго, что все переходные процессы успели закончиться (производная по времени равна 0), то уравнение диффузии приобретает следующий вид (для случая двухмерного пространства — оси x и z) без источников и стоков:

$$\partial^2 y / \partial x^2 + \partial^2 y / \partial z^2 = 0.$$

В разностном виде уравнение имеет вид:

$$(Y_{i+1,j} - 2 \cdot Y_{i,j} + Y_{i-1,j}) / \Delta x^2 + (Y_{i,j-1} - 2 \cdot Y_{i,j} + Y_{i,j+1}) / \Delta z^2 = 0.$$

Если принять $\Delta x = \Delta z$, то уравнение примет вид:

$$4 \cdot Y_{i,j} - Y_{i+1,j} - Y_{i-1,j} - Y_{i,j-1} - Y_{i,j+1} = 0.$$

Легко понять, что шаблон расчета уравнения неявный и имеет вид креста (чтобы рассчитать значение температуры в узле сетки, надо знать температуры его соседей слева, справа, сверху и снизу). Если стена дома имеет

размеры 2 метра на 2 метра, а шаг $\Delta x = \Delta z = 20$ мм, то всего для расчета температурного режима стены придется решать систему из 10 000 линейных уравнений с 10 000 неизвестных $Y_{i,j}$:

$$4 \cdot Y_{i,j} - Y_{i+1,j} - Y_{i-1,j} - Y_{i,j-1} - Y_{i,j+1} = 0, \text{ для } i = 1 \div 100 \text{ и } j = 1 \div 100,$$

к которым следует присоединить 400 штук краевых условий:

$$Y_{0,j} = f_1(j);$$

$$Y_{101,j} = f_2(j);$$

$$Y_{i,0} = f_3(i);$$

$$Y_{i,101} = f_4(i).$$

Вид решения уравнения показан на рис. 19.6.

Уравнение тепломассопереноса

В ряде физических процессов масса вещества или энергия движутся внутри воображаемой системы, перемещаясь из одних ее частей в другие. Такие процессы описываются уравнением тепломассопереноса. Для одномерного случая это уравнение имеет следующий вид: $\partial \rho / \partial t + \partial q / \partial x = B$, где $\rho(x, t)$ — плотность вещества или энергии; $q(x, t)$ — поток вещества или энергии; $B(x, t)$ — функция источников и стоков. $\rho(x, t)$, $q(x, t)$, $B(x, t)$ — функции распределения соответствующих переменных в пространстве x и времени t .

Разобьем поток вдоль оси x на элементарные ячейки, взаимодействующие друг с другом. Каждая ячейка характеризуется плотностью ρ находящегося в ней вещества и потоком вещества q через ее границы (левую и правую). Ячейки пронумеруем индексом i . Ячейку называют элементарным объемом.

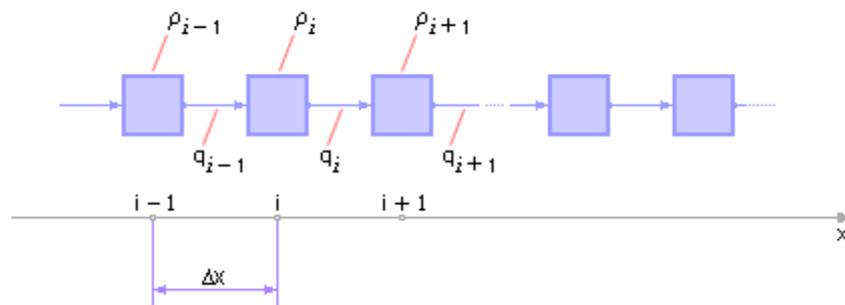


Рис. 19.10. Схема уравнения теплопроводности (пример)

В разностном виде (после подстановки выражений численного вычисления производных) уравнение теплопроводности будет иметь вид:

$$\rho(x, t + \Delta t) = \rho(x, t) + \frac{\Delta t}{\Delta x} \cdot (q(x + \Delta x, t) - q(x, t) + b(x, t))$$

Уравнение и соответствующая ему схема на рис. 19.10 показывает, сколько вещества добавится в элементарном объеме $\partial \rho / \partial t$ за время Δt , если через левую границу этого участка x за это время перетекает $q(x) \cdot \Delta t$ частиц вещества, и если через правую границу этого участка $(x + \Delta x)$ за это же время перетекает $q(x + \Delta x) \cdot \Delta t$ частиц вещества, и на участке есть источник или сток вещества величиной b .

В зависимости от решаемой задачи уравнение может быть дополнено. Функция $b(x, t)$ обычно задана. Если требуется определить закон движения

вещества $q(x, t)$, чтобы обеспечить заданное пользователем распределение вещества по оси x со временем $\rho(x, t)$, то для такого расчета достаточно одного уравнения.

Если задан закон движения вещества $q(x, t)$ и требуется определить, где в результате этого движения, когда и сколько будет вещества $\rho(x, t)$, то также достаточно одного уравнения.

Но часто требуется определить одновременно и $\rho(x, t)$ и $q(x, t)$ при заданных начальных условиях, то есть вопрос состоит в том, как быстро $q(x, t)$ и сколько $\rho(x, t)$ вещества окажется в определенных точках x . Тогда одного уравнения для расчета двух неизвестных недостаточно. И основное уравнение должно быть дополнено вспомогательными выражениями, указывающими, как связаны между собой $\rho(x, t)$, $q(x, t)$. Такое выражение должно указывать, как быстро движется вещество, если задано его количество. В производстве и технике мы называем такое выражение регулятором: как связано количество деталей ρ , хранящихся на участке x , с желаемой скоростью обработки таких деталей q .

В зависимости от того, с каким веществом мы имеем дело, могут существовать различные выражения дополнительной связи (дополнительное уравнение) между потоком q и плотностью ρ . Например, для жидкости, где все частицы движутся *одновременно*: $q = \rho \cdot v$, где v — скорость перемещения частиц в потоке. Или более сложное уравнение движения, связывающее скорость движения потока $v(x, t)$ с причиной этого движения, силой F :

$$\partial v / \partial t + v \cdot \partial v / \partial x = F / \rho.$$

В свою очередь, сила F может быть вызвана к действию перепадом давления P : $F = \partial P / \partial x$.

Легко увидеть в данном уравнении в качестве его основы второй закон Ньютона, связывающий ускорение $\partial v / \partial t$ и силу F через массу ρ . Дополнительное слагаемое $v \cdot \partial v / \partial x$ появляется постольку, поскольку при движении появляется приток вещества через границы x и $(x + \Delta x)$ элементарного участка, который несет за собой импульс $v \cdot \partial v / \partial x \cdot \rho$.

В разностном виде при $q = \rho \cdot v$ уравнение переноса будет иметь вид:

$$\rho(x, t + \Delta t) = \rho(x, t) + \frac{v \cdot \Delta t}{\Delta x} \cdot (\rho(x + \Delta x, t) - \rho(x, t) + b(x, t))$$

Расчет уравнения производится аналогично уравнению диффузии.

Свойства уравнения переноса заключаются в перемещении неоднородностей в потоке со временем вдоль пространственной оси. Скорость перемещения связана с величинами v или q . На рис. 19.11 показан вид одного из возможных вариантов решения уравнения тепломассопереноса. Распределение плотности (неоднородность) может перемещаться вдоль оси со временем.

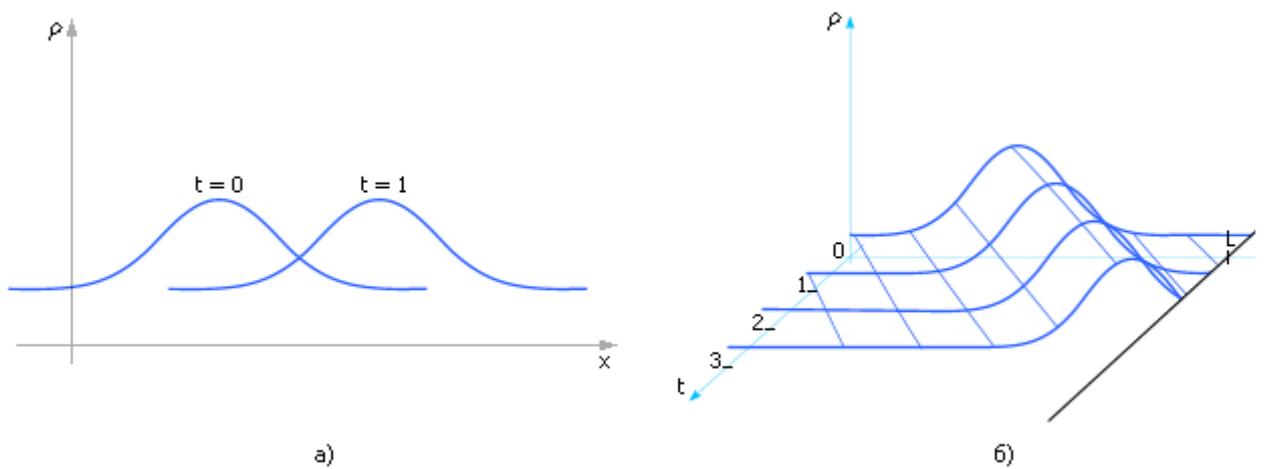


Рис. 19.11. Свойства уравнения тепломассопереноса. а) — в двух осях; б) — в трех осях

Следует обращать внимание на устойчивость уравнения. Устойчивость связана с видом выбранного разностного шаблона и его параметрами: шагом по t , шагом по x , а также скоростью (поток) перемещения массы (см. рис. 19.12). Чем больше шаг, тем больше риск получения неустойчивых (качественно неверных) решений. Но чем меньше шаги, тем медленнее происходит компьютерный расчет всего поля, так как объем вычислений возрастает значительно.

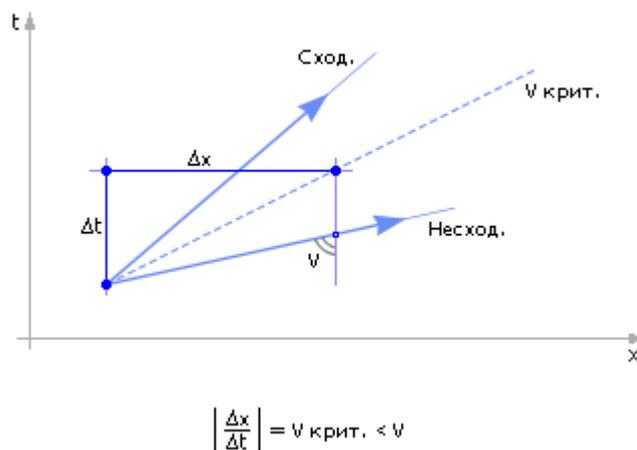


Рис. 19.12. Соотношение параметров расчета для получения устойчивого решения

В реальных системах обычно присутствуют как диффузионные свойства, так и свойства уравнения переноса, движения. Поэтому эффекты переноса и диффузии смешиваются, и картины в результате расчета получаются достаточно сложные, зависящие от краевых и начальных условий и соотношения коэффициентов уравнений, как показано, например, на рис. 19.13.

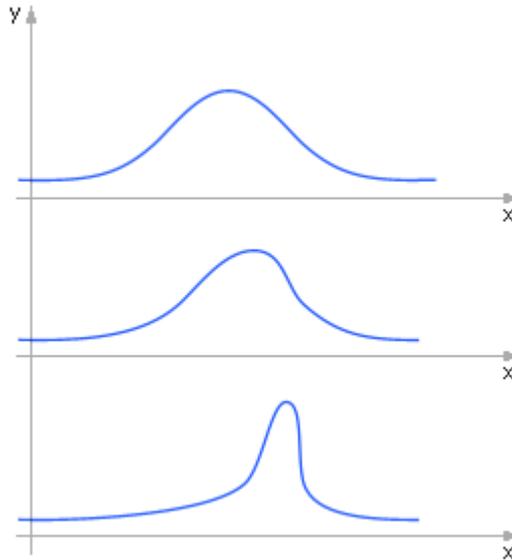


Рис. 19.13. Вид возможных решений уравнения тепломассопереноса с элементами диффузии. На рисунке видно проявление нескольких свойств одновременно

К распределенным системам относятся производственные системы (см. лекцию 31. «Моделирование производственных процессов и систем»).

Одной из самых сложных задач, рассматриваемых в классе систем с распределенными параметрами, является прогноз погоды. При этом учитываются температура, давление, скорость ветра, влажность, плотность воздуха, рельеф местности. Все эти величины являются функцией трех пространственных координат и времени. Ежедневно анализу в центрах прогнозирования погоды подвергается 10^6 исходных данных, не считая фотографий. Расчет производится два раза в сутки: в 00 часов и в 12 часов по Гринвичу. Обмен измерительной информацией между метеостанциями разных стран идет в рамках Всемирного метеорологического общества, куда входит и Российская Федерация. В среднем для расчета одного прогноза учитывается информация от 4 000 приземных источников (включая 800 морских) и 650 аэростанций. Сеть измеряющих приборов покрывает земной шар неравномерно и это осложняет расчеты.

Нетрудно оценить в среднем шаг расчетной сетки. При радиусе Земли $R = 6\,400$ км площадь Северного полушария равна: $S = 260\,000\,000$ км². Таким образом, на одну аэростанцию падает $S/650 = 400\,000$ км² или квадрат со стороной, равной расстоянию от Москвы до Петербурга. На одну наземную станцию падает квадрат со стороной 300 км. Такой шаг сетки расчета ведет к огромным трудностям при борьбе за устойчивость расчета.

Свои проблемы вносит и недостаточная точность исходных данных, поступающая от источников — погрешность измерений на море составляет обычно 10%, на суше — 15%.

При расчетах прогноза погоды кроме описанных уравнений дополнительно учитывают законы сохранения (массы, импульса, энергии), свойства веще-

ства (Клаузиса-Клапейрона, Ван-дер Ваальса, фазовых переходов, поглощения солнечной энергии, излучения и другие). Учитывают также силы тяжести, вязкости, Кориолиса и другие. Расчет погоды ведут обычно на сверхмощных компьютерах типа Cray (см. [видео Cray](#)).

20. ТЕХНОЛОГИЯ ИСПОЛЬЗОВАНИЯ КОМПЬЮТЕРНЫХ МОДЕЛЕЙ

Как мы уже отмечали в лекции 01 «Понятие моделирования. Способы представления моделей», модели строятся для решения определенных задач, поэтому здесь мы рассмотрим типы таких задач и то, как в них используются модели, которые мы уже научились строить.

Модель — закономерность, преобразующая входные значения в выходные: $Y = M(X)$. Под этим можно понимать таблицу, график, выражение из формул, закон (уравнение) и т. д. Это вопрос способа записи закономерности.

Под Y в системотехнике понимают некоторый интересующий исследователя или владельца системы показатель. Каждая система существует или создается, чтобы реализовать определенную цель. Нет систем без целей. Вот цель-то и является выходным, последним параметром в цепи преобразований от входа к выходу, который может нас интересовать, так как ради него, собственно, и прodelываются все преобразования. Те переменные, которые как-то не связаны по цепям с выходным показателем, не относятся к рассматриваемой системе и должны быть отброшены [9].

Представим нашу систему как граф. Это возможно, так как система есть элементы и связи между ними, что соответствует вершинам и дугам графа. Дальнейшее изложение материала будем вести на примере графа, изображенного на рис. 20.1.

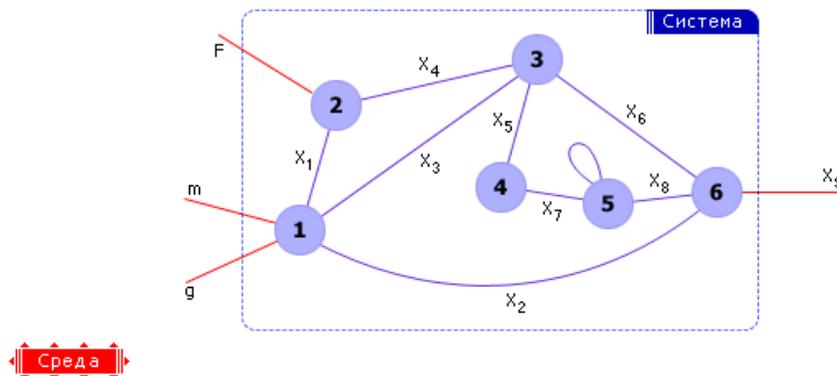


Рис. 20.1. Представление системы в виде графа

Элементы системы описываются законами, то есть уравнениями (переменные, операции между ними и знаки уравнивания) или системами уравнений в общем случае, что соответствует вершинам графа. Если уравнений в вершине несколько, то эту вершину всегда, при желании, снова можно будет разбить на подграф, где каждой вершине уже будет соответствовать только одно уравнение. Связи графа указывают на связи элементов системы между собой, то есть связь соответствует общей для двух вершин переменной. Итак, каждая вершина ассоциируется с формулой (например, присвоением выра-

жения или уравнением), связывающим переменную вершины с остальными переменными, доступными ей по ее связям.

Если граф достаточно подробный, таков, что каждой вершине соответствует только одно уравнение, то можно ассоциировать вершины с переменными. Одна вершина — одна переменная. Преобразованный таким образом граф мы уже видели в лекции 11. Переход от одной формы представления графа к другой возможен всегда.

Обратите внимание: часть связей находится внутри графа — это внутренние связи системы. Часть связей графа связывает **переменные системы X с внешними переменными**, которые не являются частью системы, а являются частью среды. Эти связи пересекают границы графа, границы системы.

Для моделирования очень важно определить, где проходит эта граница, что будет подлежать моделированию, а что нет. Что будет описано причинно-следственными связями, а что нет и так и останется бесконечно большим и непознанным. Еще раз обратимся к тексту лекции 01: «...«Модель — поиск конечного в бесконечном» — эта мысль принадлежит Д. И. Менделееву. Что отбрасывается, чтобы превратить бесконечное в конечное? В модель включаются только существенные аспекты, представляющие объект, и отбрасываются *все остальные* (бесконечное большинство)...»

Итак, граница отделяет конечное (система) от бесконечного (среда). Граница эта может проходить в нескольких разных местах. Это говорит о том, что модель может граничить с несколькими областями, не имеющими описания, которые задаются по отношению к системе только как воздействующие на нее сигналы, как данные, но не как законы (см. рис. 1.12). Бесконечность может оказаться и внутри системы, тогда такая система называется открытой (см. лекцию 11 «Построение модели динамической системы в виде дифференциальных уравнений и расчет ее методом Эйлера»).

Заметьте, граф является сильно связанным образованием, количество связей в таком графе больше чем количество вершин. В сложных системах связей намного больше, чем вершин. Для нормального графа каждая его вершина должна иметь связь с любой другой вершиной графа через цепочку связей. Если в составе графа вы обнаружили несвязанные между собой куски, то модель системы некорректна или вы имеете дело с двумя или несколькими независимыми системами. Если граф большой, то имеет смысл разрезать его на небольшие подграфы и изучать их по отдельности (см. рис. 20.2). Логично разрезать граф по таким линиям, чтобы при этом разрывалось как можно меньше связей, и получалось как можно больше отдельных кусков (подграфов). Заметим, что этими действиями мы, фактически, привели граф к иерархической форме. Иерархия — это способ борьбы со сложностью изучаемой системы. В этом случае между линиями разреза оставляют одну вершину, структуру которой расшифровывают отдельно.

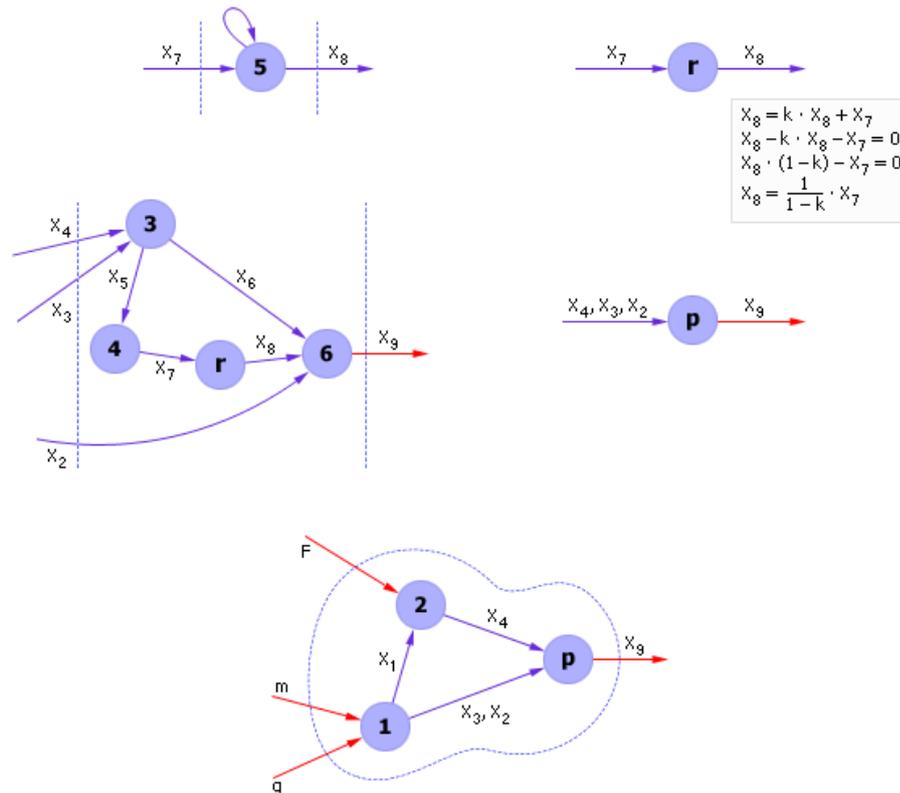


Рис. 20.2. Разбиение графа (системы) на подграфы (подсистемы)

Применение этого приема очень эффективно, если в графе встречаются несколько одинаковых подграфов. В этом случае они изучаются отдельно и один только раз, а результат используется многократно и обобщается на все остальные случаи. Далее в курсе «Модели и методы искусственного интеллекта» мы отметим этот прием как основной в мышлении человека, которое только и занимается тем, что строит в голове модели объектов окружающего мира, свертывает сложные конструкции в новые понятия, ежеминутно решает задачи на иерархических моделях и борется, таким образом, со сложностью окружающего мира.

Итак, граф задает своей структурой модель системы, которая выражена как система взаимосвязанных уравнений (см. рис. 20.3).

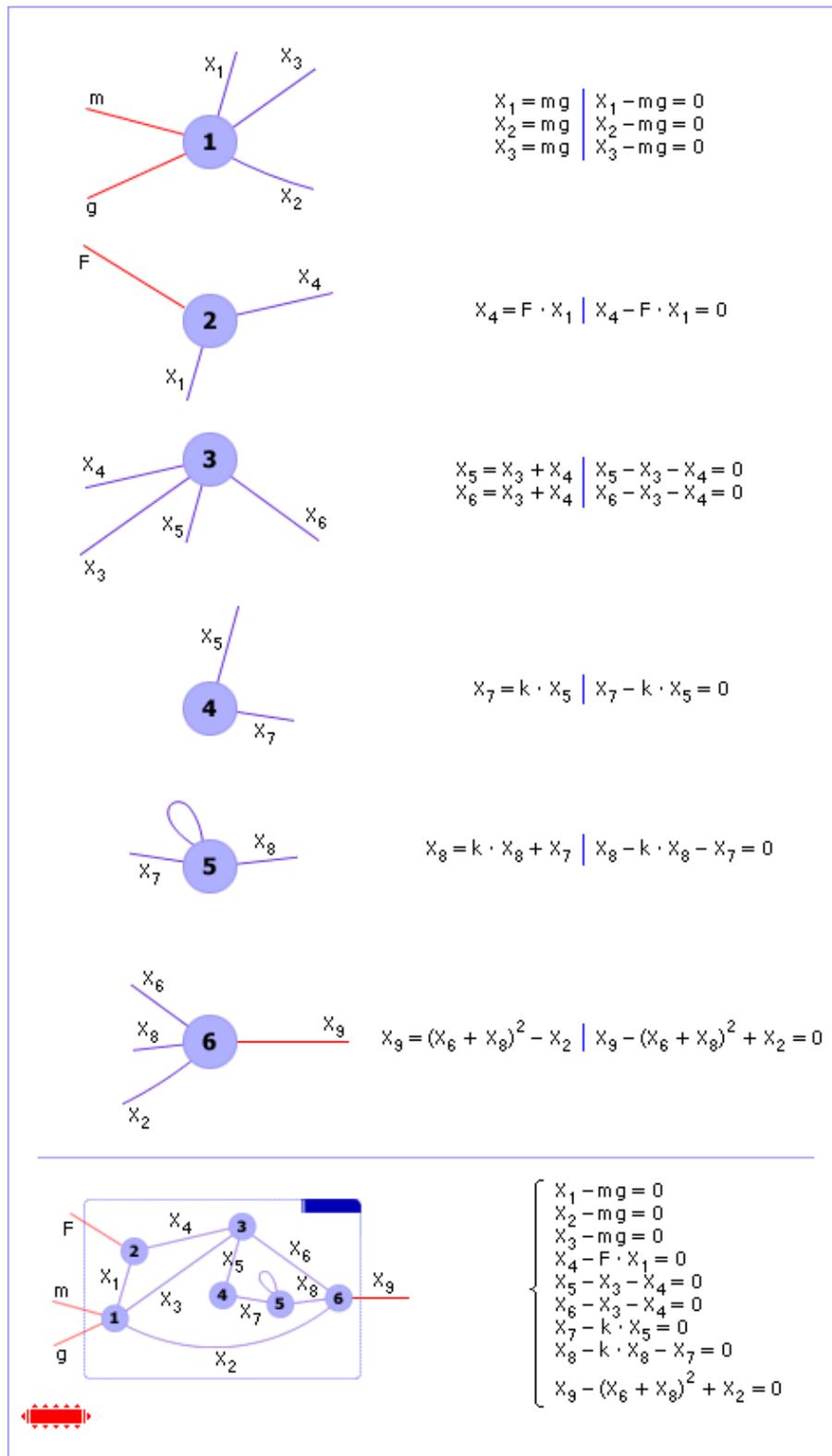


Рис. 20.3. Иллюстрация соответствия вершин графа описанию подсистем большой системы

Или в самом общем виде пишут:

$$\bar{M}(\bar{X}) = 0, \text{ где } \bar{M}, \bar{X} \text{ — векторы}$$

Если среди внешних переменных определить цель и определить управление (за счет чего достигается цель), то часть внешних переменных будет называться **выходными переменными (цель)**, а другая — **входными переменными**. Если определен вход (управление) и выход (цель) на графе, то

связи графа становятся направленными, от входа к выходу (см. рис. 20.4). Эти связи выражают причинно-следственные отношения — изменения на входе ведут к изменению значений на выходе.

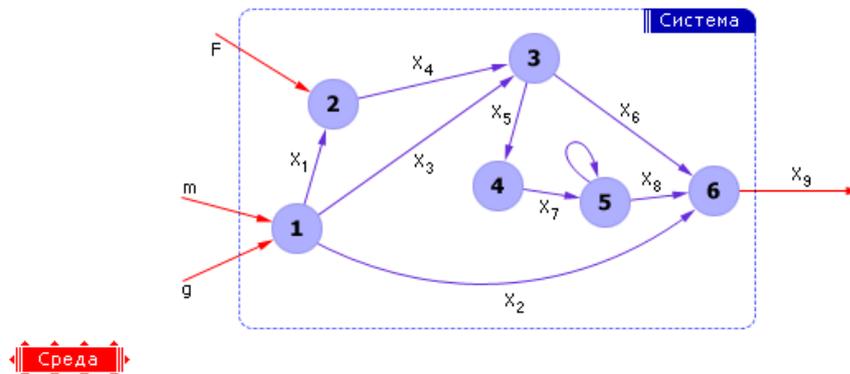


Рис. 20.4. Представление системы в виде ориентированного графа, граф соответствует определенной задаче, решаемой на системе

В этом виде граф соответствует задаче, решаемой на графе. Задача упорядочивает порядок вычислений.

Если теперь применить последовательно уравнения системы от назначенного пользователем входа к выходу, то с математической точки зрения образуется цепочка выражений (см. рис. 20.5). Искомые переменные будут выражены в итоге по цепочке через входные переменные (см. рис. 20.6). Система уравнений подстановкой свертывается в формулу.

$$x_9 = M_6(M_5(M_4(M_3(M_2(M_1(m, g)), F))))$$

$$x_9 \stackrel{6}{=} (x_6 + x_8)^2 - x_2 \stackrel{5}{=} (x_6 + \frac{1}{1-k} \cdot x_7)^2 - x_2 \stackrel{4}{=} (x_6 + \frac{1}{1-k} \cdot k \cdot x_5)^2 - x_2 \stackrel{3}{=} ((x_3 + x_4) \cdot (1 + \frac{k}{1-k}))^2 - x_2 \stackrel{2}{=} ((x_3 + F \cdot x_1) \cdot \frac{1}{1-k})^2 - x_2 \stackrel{1}{=} ((mg + F \cdot mg) \cdot \frac{1}{1-k})^2 - mg$$

Рис. 20.5. Явное решение задачи «Управление выходом x_9 системы через вход (m, g, F) » путем подстановки

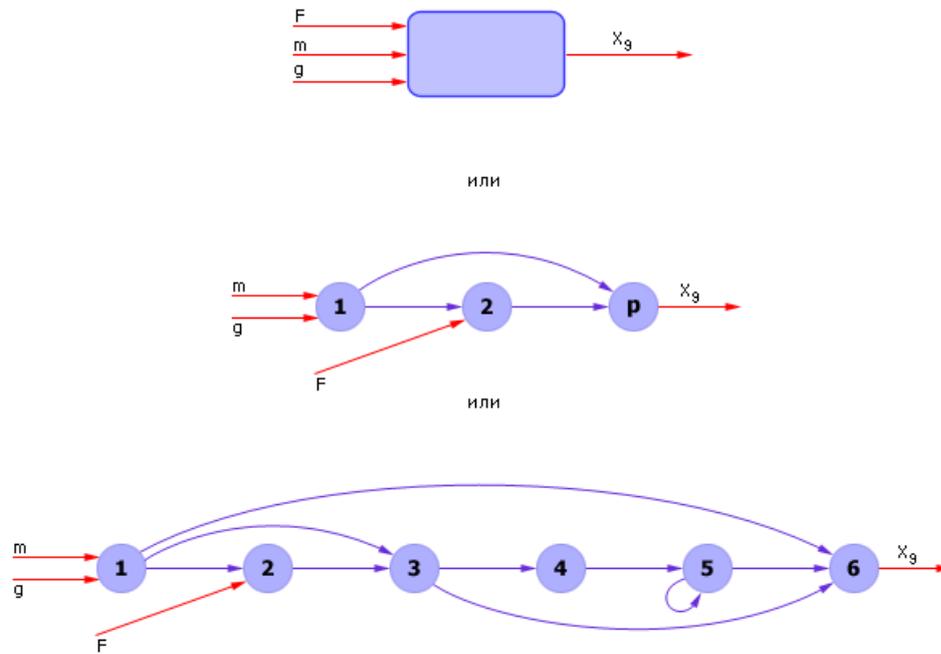


Рис. 20.6. Процедура последовательной детализации графа (операция композиции и декомпозиции)

В общем виде это выглядит так: $Y = M(M \dots (M(X)) \dots)$. Такая математическая структура называется **композицией** и задает цепочку (последовательность) вычислений, а значит алгоритм вычисления ответа задачи, что в свою очередь определяет решение системы. Решение может быть как численным, так и аналитическим. Если задача будет другой, то модель всей системы развернется в другую цепочку, от других входных переменных к другому выходу. Композиция, соответствующая задаче, изменится, но модель всей системы останется неизменной.

Конечно, не всегда цепочка может выразить явно зависимость выхода от входа, еще чаще это происходит, когда выражают вход через выход (вход как функция выхода). Выражая искомое через известное, требуется применение обратных к каждой из примененных в модели операции. Например, к $x_1 = \sin(x_2)$ применимо обратное преобразование $x_2 = \arcsin(x_1)$, к $x_1 = x_2^2$ требуется применить обратную операцию $x_2 = \sqrt{x_1}$ и так далее. А это не всегда возможно. Это зависит от того, насколько развита алгебра (правила преобразований) данного вида выражений. Если алгебра не может определить некоторые обратные преобразования к ряду выражений, операций или функций, то тогда модель остается неявной, и приходится применять специальные методы расчета неявных уравнений. Решения в этом случае получают численными методами.

К таким осложнениям приходят также модели, содержащие петли в графе (см. рис. 20.7).

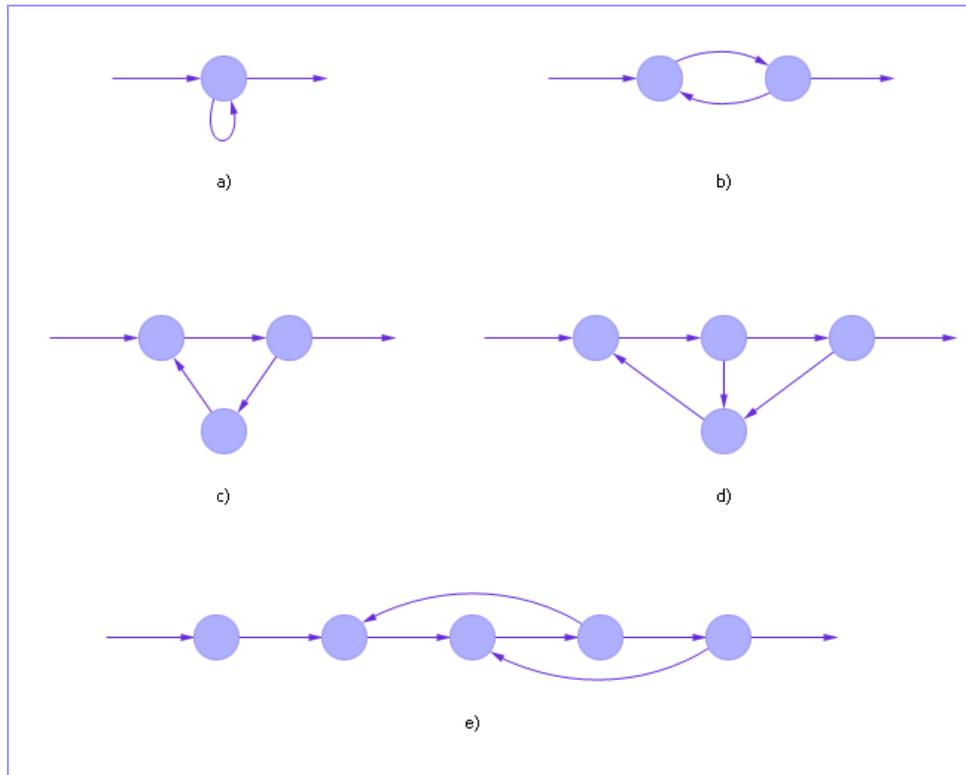


Рис. 20.7. Графы различной сложности, содержащих петли

Если определен вход и выход и на модели определена задача, граф становится ориентированным. Задача определяет композицию модели, способ вычисления ответа. Если общая формула системы решается, то формула явная, и алгоритм ее реализации на цифровых машинах будет линейным, если аналитического решения нет, то формула неявная, и алгоритм будет циклическим.

Теперь настало время уточнить понятие входных переменных, поскольку их много и список их весьма неоднороден. Надо иметь в виду, что входные переменные, которые ранее мы обозначали как X , могут быть обозначены в целях детализации как X_i, U_i, P_i, Q_i .

Во-первых, X может быть не одной переменной, а целым **вектором переменных** $\{X_1, X_2, \dots, X_n\}$, так как сложные системы, которые мы моделируем, обычно связаны со средой множеством факторов $\{X_1, X_2, \dots, X_n\}$. Их значения обычно мало интересны или недоступны напрямую для изменения владельцем системы, но они существуют. Иногда это часть внутренних переменных системы, переменные состояния, фазовые переменные, память системы и так далее.

Во-вторых, логически удобно разделить вектор X на входные переменные (собственно X) и **переменные управления** U . Тогда под X обычно понимают не зависящие от воли владельца системы факторы, а под U — факторы, которыми владелец системы может непосредственно распоряжаться по собственной воле. Такие факторы принято называть **управляемыми переменными** или просто **управлением**. Заметим, что обычно значения переменных U чем-то ограничены. В самом деле, нельзя ведь открыть водопроводный кран больше чем на 1 (кран открыт полностью) или меньше чем на 0

(кран полностью закрыт). Поэтому если понимать под U степень открытия крана, то $0 \leq U \leq 1$. В других случаях пишут более общий вариант $U_{\min} \leq U \leq U_{\max}$. В этом смысле здесь и далее мы будем считать, что управление, поскольку оно ограничено, это некоторый ресурс.

В-третьих, P — *мало меняющиеся переменные*, которые в этом случае называют **параметрами системы**; по своей сути, конечно, они мало отличаются от X . В прикладных задачах их часто выносят отдельно, так как динамически они (на отрезке времени рассмотрения или существования задачи) не меняются и не меняют свойств системы.

В-четвертых, помехи Q . Это переменные, которые действуют на систему помимо воли ее владельца и ухудшают значение желаемого показателя Y . Помехи всегда действуют во вред владельцу системы, занижая желаемые показатели системы. Управление U — фактор, который призван компенсировать негативное действие помех Q на выходной показатель цели Y . То есть при одном и том же значении U , при действии помех, в отличие от случая их отсутствия, показатель Y будет ниже. Ликвидировать вообще все действующие на объект помехи часто не удается по трем перечисленным ниже причинам.

1. Помеха действует опережающе, быстрее, чем мы ее можем компенсировать, так как предугадать помехи и подвести под это заранее соответствующее управление сложно (хотя иногда это сделать можно специальными методами предсказания и опережения).
2. Помеха обычно действует редко, но в большом количестве (большой амплитуды), и ресурса управления для ее компенсации сразу обычно не хватает. Кстати, держать большой запас управления под возможные большие помехи везде, где они могут возникнуть, накладно, поскольку такой ресурс является омертвленным капиталом (затратами, не приносящими прибыль). Именно поэтому в больших системах возникают «black-out», каскадные сбои и отключения, катастрофы. Управляющий ресурс обычно хранят в одном месте и стараются доставить его как можно скорее к месту помехи или местам, где уже успели сформироваться отклонения, вызванные ими. К тому же эффекту приводят редкие совпадения нескольких помех, которые сами по себе в отдельности некатастрофичны.
3. Случай, когда помеха действует часто (всегда) и понемногу тоже существует и хорошо компенсируется управлением (а статические системы). Случай, когда помеха действует помногу и часто, невероятен, так как следует признать, что в этом случае система плохо спроектирована, и ее надо попросту перепроектировать, а не управлять ею. Случай, когда помеха действует редко и помалу, тривиален и прост для управления или сводится к предыдущим вариантам как частный.
4. Помеха возникает часто не там, где ее ждали и где находится ресурс, достаточный для ее компенсации.

Поэтому часто борются не с самими помехами, а с **отклонениями** переменных X и Y от идеальных плановых их значений, борются с последствиями аварий, а не с их причинами.

Отойдем на минуту от серьезного разговора и поясним важную мысль на шуточном примере. «Чтобы корова давала больше молока и меньше ела, ее надо больше доить и меньше кормить». Автор этой фразы — известный шоумен Николай Фоменко («Русское радио»). Пример демонстрирует достаточно распространенную ошибку, когда инженер путает и считает что $X = -U$ (чтобы убрать помеху, надо подать компенсирующий сигнал такой же величины и на ту же переменную), чего практически в сложных системах не бывает. Задача управления в этой шутке решается чересчур тривиально или, точнее сказать, просто неверно поставлена, она попросту отсутствует. Причина этого заключается в отсутствии модели M . Здесь просто не описана корова как система, как сено превращается в молоко. Не учтено, что управление (сено) не может быть использовано (доставлено на выход) в качестве цели (молока).

Посредством управления U удается часто снизить негативное действие помех на целевой показатель Y . Действие управления на помеху есть, но оно неявное, точнее сказать и помеха Q , и управление U действуют на показатель Y , при этом управление выбирается таким, чтобы свести на нет негативное действие помех на Y .

Усилия по компенсации помех всегда чего-то стоят владельцу системы, так как используют тот самый ресурс, который мы обозначили ранее как U_{\max} . Итак, заметьте: с управлением всегда связано понятие ресурса. Управление черпает свои силы в ресурсе. Если ресурс мал, то управление связано и не может справиться с сильной помехой.

Если ресурс мгновенно возобновляем, то $U_{\min} \leq U \leq U_{\max}$. Если ресурс обладает свойством аддитивности, накапливается и тратится, не может мгновенно возобновиться, то

$$\int_0^x U_{\text{ип}}(t) dt \leq \int_0^x U_{\text{п}}(t) dt$$

где $U_{\text{ип}}(t)$ — темп использования ресурса, $U_{\text{п}}(t)$ — темп поставки ресурса.

Как известно из математики и было уже рассмотрено в [лекции 01](#), с выражением $Y = M(X)$ можно решить три вида задач, которые приведены в табл. 20.1.

Таблица 20.1.
Формы записи модели и типы решаемых задач

	Известно	Неизвестно	Решение
Прямая задача	X, M	Y	$Y = M(X)$
Обратная задача	Y, M	X	$X = M^{-1}(Y)$
Задача настройки модели	X, Y	M	$M = f(X, Y)$

Задача анализа (прямая задача)

Изучается влияние некоторого входного параметра U на конечный результат или показатель Y . 1, 2, ..., N — эксперименты, проводимые с моделью. Если подавать неоднократно различные значения U на модель M (алгоритм показан на рис. 20.10), то, измеряя Y , в результате моделирования на выходе модели можно построить зависимость $Y = M(U)$, см. рис. 20.8. Обыч-

но реально ограничиваются некоторым набором входных воздействий $U_{\min} \leq U \leq U_{\max}$, проходя значения U точку за точкой с определенным шагом ΔU . При этом, во время такого эксперимента часть входных параметров X замораживают, оставляя их значения неизменными. При необходимости можно повторить эксперимент по перебору U из интервала $U_{\min} \leq U \leq U_{\max}$ при другом значении X . В этом случае получается семейство кривых $Y = M(U, X)$, см. рис. 20.9.

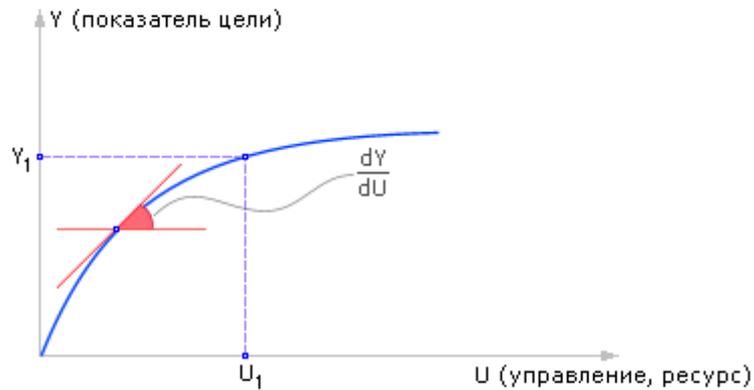


Рис. 20.8. Вид зависимости цели Y от управления U , полученный экспериментально на модели системы

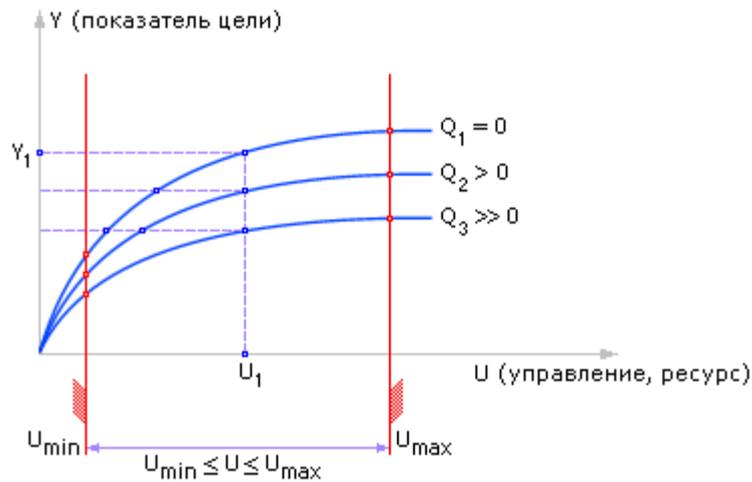


Рис. 20.9. Вид зависимости цели Y от управления U при различных значениях действующих помех Q

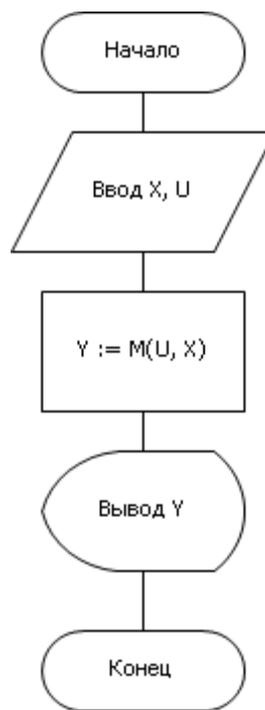


Рис. 20.10. Алгоритм, применяемый при решении прямой задачи исследования модели

Испытывая неоднократно модель при различных входных сигналах, мы можем получить зависимость выхода от входа. Такая задача называется прямой (см. лекцию 01). Результатом задачи является кривая, семейство кривых, таблица, а когда это возможно, то формула, закон и т. д. Основной вопрос анализа — познание свойств объекта. «Воздействуем на объект и смотрим, что получится, как он реагирует, делаем вывод о его свойствах, возможностях».

Важными понятиями в системотехнике являются «управляемость» и «наблюдаемость». По виду кривых $Y = M(U)$ (рис. 20.9) можно определить, для всех ли значений Y возможно некоторое значение входного сигнала (U, X) ? Любое ли значение Y можно достигнуть, используя переменные (U, X) из выбранного диапазона. То есть характер кривых указывает, в какой области Y объект является управляемым. Понятие «управляемость» касается выходной переменной.

Наблюдаемость — возможность измерения, анализа той или иной характеристики объекта. Иногда из-за того, что некоторая величина не может быть непосредственно измерена в результате эксперимента, приходится, чтобы получить о ней хоть какое-то представление, довольствоваться косвенными показателями. Понятие «наблюдаемость» касается выходной переменной. Проектировать системы надо так, чтобы качество наблюдаемости и управляемости были обеспечены [9].

Отношение изменения Y к изменению U (при фиксированном X) называется чувствительностью Y по U . Обычно, так как кривая $Y = M(U)$ для сложных систем нелинейна, то изменение U принимают небольшой величиной, в идеале $\Delta U \rightarrow 0$. В математическом смысле, чувствительность — это производная dY/dU . Понятие чувствительности касается отношения выхода ко входу (рис. 20.8).

Чтобы сократить количество испытаний, входные воздействия выбирают по определенному правилу. Естественно желание получить необходимый объем информации о системе при минимальном количестве испытаний. Таковую систему испытаний планируют факторным экспериментом.

Задача синтеза (обратная задача)

Цель задачи синтеза — нахождение экстремума функции результата. Когда анализ закончен и построены функции, графики, таблицы, когда объект (его свойства и поведение) исследован во всех вариантах возможных входных воздействий, имеет смысл найти среди всего этого многообразия откликов наилучший. Обычно выход — цель функционирования системы, и логично принять, что цель должна принимать лучшие из всех возможных значений, потому имеет смысл найти такие значения входных параметров U , при которых выходной показатель Y примет свое наилучшее значение (экстремум). При этом под экстремумом может подразумеваться как минимум, так и максимум зависимости $Y(U)$. Чтобы найти экстремум, модель включают в контур (см. рис. 20.11) с некоторым алгоритмом A , осуществляющим автоматическое управление входом U и построенным так, что в результате его работы производится поиск такого входного воздействия U на модель M , при котором она выдает наилучший выходной результат.

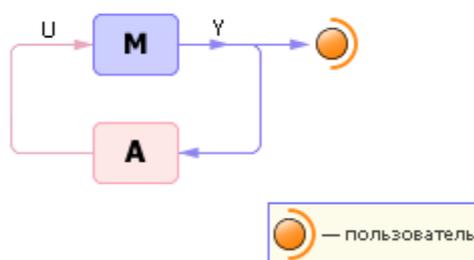


Рис. 20.11. Схема решения обратных задач (синтез)

Существуют различные алгоритмы поиска оптимума функции $Y = M(U)$. Упомянем три из них (подробно эти и другие методы вы будете изучать в дисциплине «Системный анализ и исследование операций»).

1. Метод перебора. Алгоритм метода перебора представлен на рис. 20.12. Этот метод обеспечивает поиск глобального экстремума, но расточителен к вычислительным ресурсам, так как просматривает все возможные входные значения U с определенным шагом H и выбирает наилучший среди всех выходных результатов Y (см. рис. 20.13). Наилучшее из встреченных Y сохраняется и уточняется в ячейке R , значение U при этом значении Y сохраняется по ходу алгоритма в ячейке Z . Кроме этого имеется риск пропустить нужную точку, «перешагнув» через нее из-за слишком большого размера шага H .

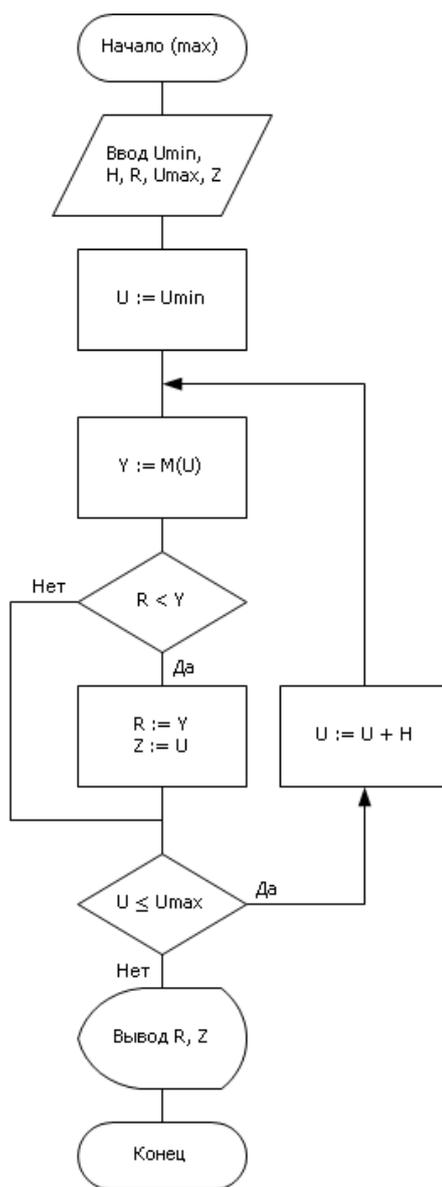


Рис. 20.12. Алгоритм перебора, примененный к решению задачи синтеза — поиск наилучшего U для максимизации Y

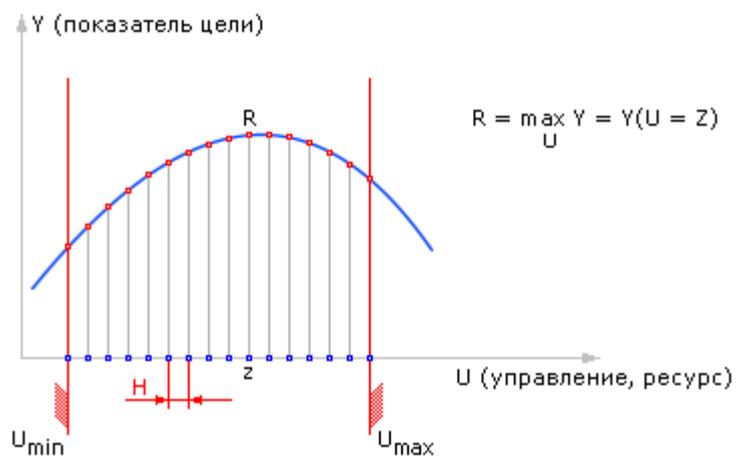


Рис. 20.13. Рисунок поиска экстремума функции $Y = M(U)$ методом перебора

3. Метод деления шага пополам. Этот метод более экономичен по отношению к методу перебора, так как анализирует по ходу движения, в каком направлении происходит улучшение (увеличение или уменьшение) функции $Y = M(U)$ и старается двигаться именно в этом направлении. Если при этом попутно обнаруживается, что тенденция во время движения изменилась (уменьшение Y сменилось на его увеличение или наоборот), то алгоритм (рис. 20.14) разворачивается обратно (то есть меняет знак приращения входного сигнала на обратный) и снова идет в нужную сторону, но шаг при этом уменьшается вдвое. Это позволяет подойти к точке экстремума с малым значением шага поиска, что лучше локализует результат (см. рис. 20.15).

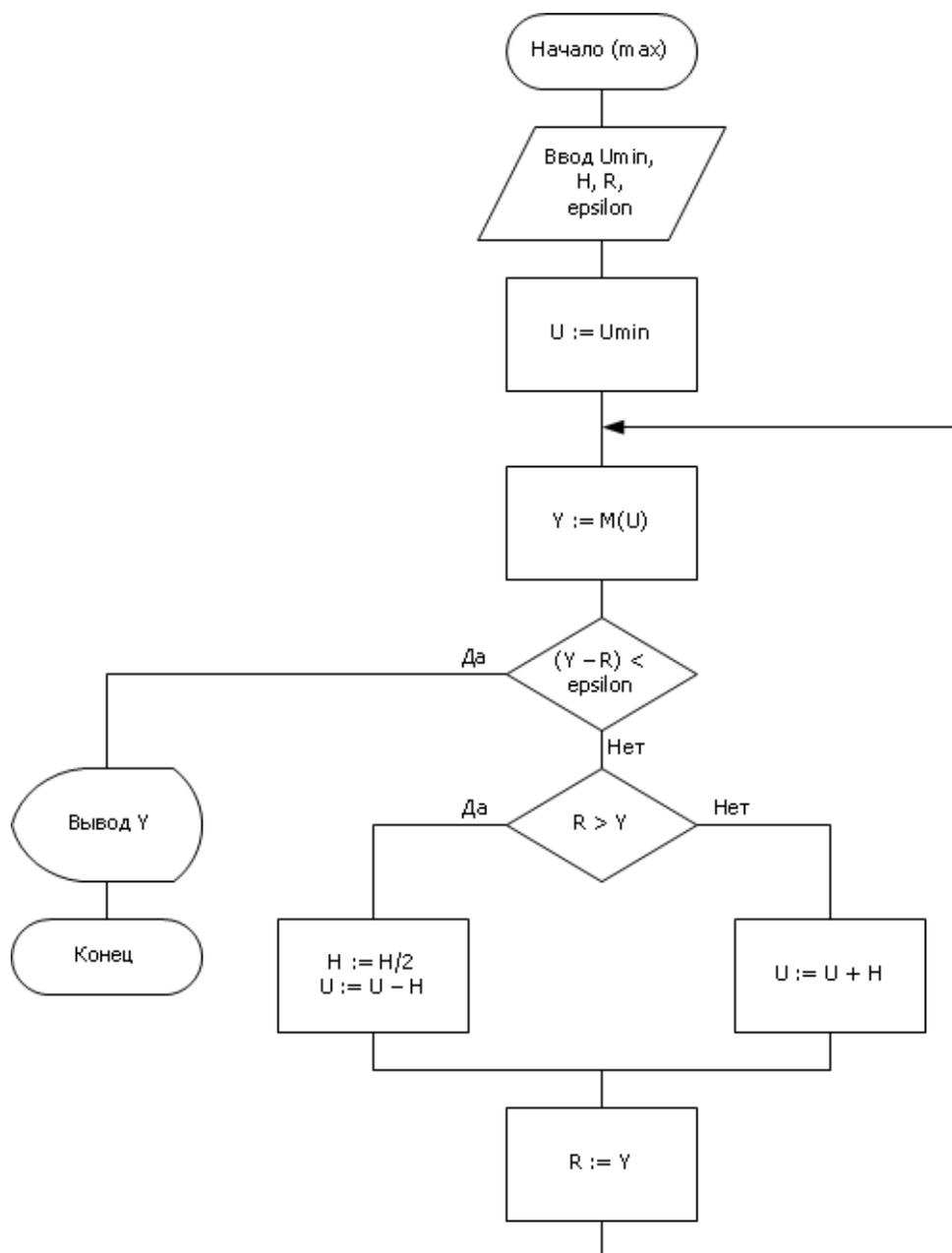


Рис. 20.14. Алгоритм деления шага пополам, примененный к решению задачи синтеза — поиск наилучшего U для максимизации Y

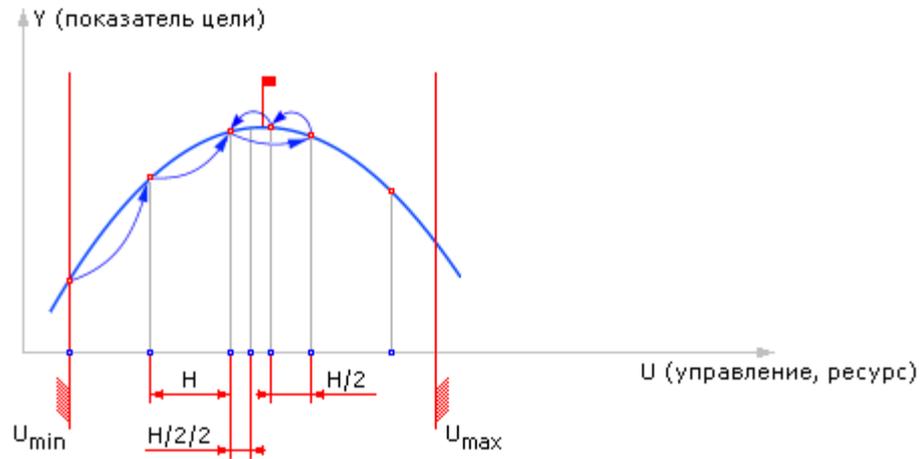


Рис. 20.15. Рисунок поиска экстремума функции $Y = M(U)$ методом деления шага пополам

5. Метод градиента. Метод, алгоритмическая реализация которого представлена на рис. 20.16, использует свойства гипотетической кривой $Y = M(U)$, а именно тот факт, что по направлению производной p , вычисленной на основе двух рядом стоящих точек, можно определить, в какую сторону уменьшается (увеличивается) значение цели Y и двигаться сразу в нужную сторону (см. рис. 20.17). Такая стратегия существенно сокращает время поиска. Недостатком алгоритма является то, что производная может перестать меняться в области локального экстремума, и глобальный экстремум такой алгоритм не найдет.

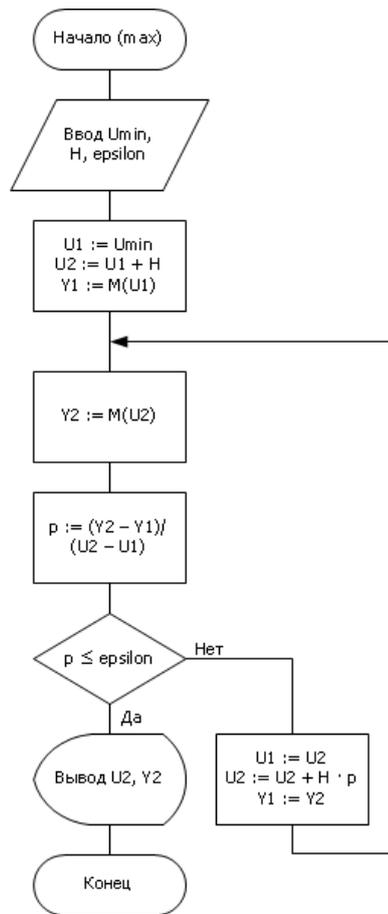


Рис. 20.16. Алгоритм поиска экстремума методом градиента, примененный к решению задачи синтеза

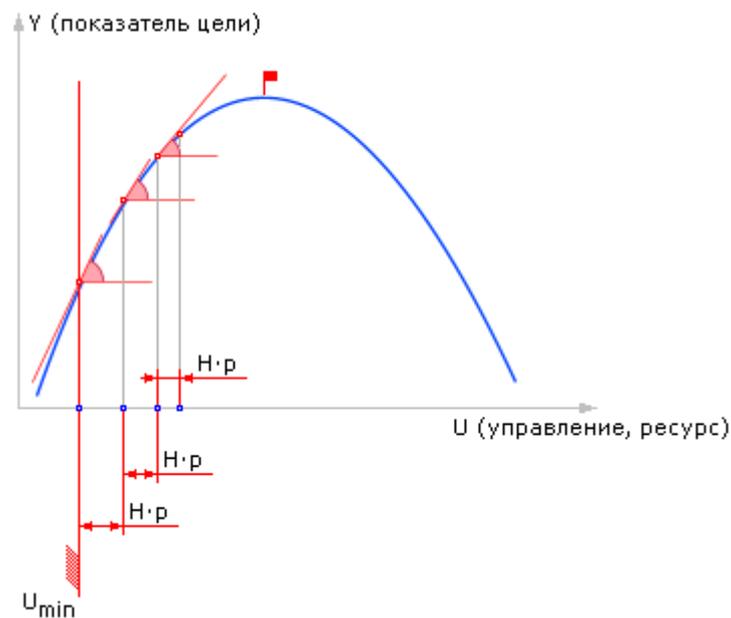


Рис. 20.17. Рисунок поиска экстремума функции $Y = M(U)$ методом градиента

Задачу настройки модели мы уже подробно изложена в лекциях 2-8 (см. лекцию 2), и останавливаться на ней мы уже не будем. Это способы построения собственно самой модели.

Тренажеры

Этот класс задач, использующих модели, применяют для выработки навыков обучения у управляющего персонала. К тренажерам близки компьютерные игры. Управление моделью в данном случае осуществляет человек-оператор, который наблюдает за выходом модели (см. рис. 20.18). Воздействуя на вход модели, оператор старается добиться нужного выходного результата, и в процессе этих действий получает необходимые навыки по управлению, которые затем может перенести на реальный объект. На рис. 20.19 показан примерный вид алгоритма реализации тренажера на базе модели.

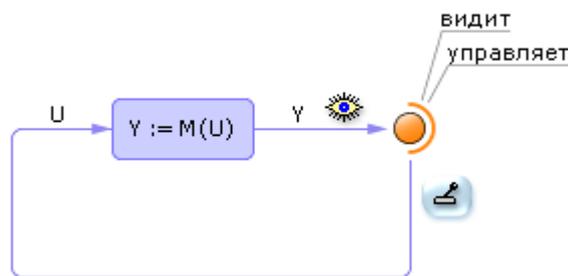


Рис. 20.18. Схема использования модели в тренажерах

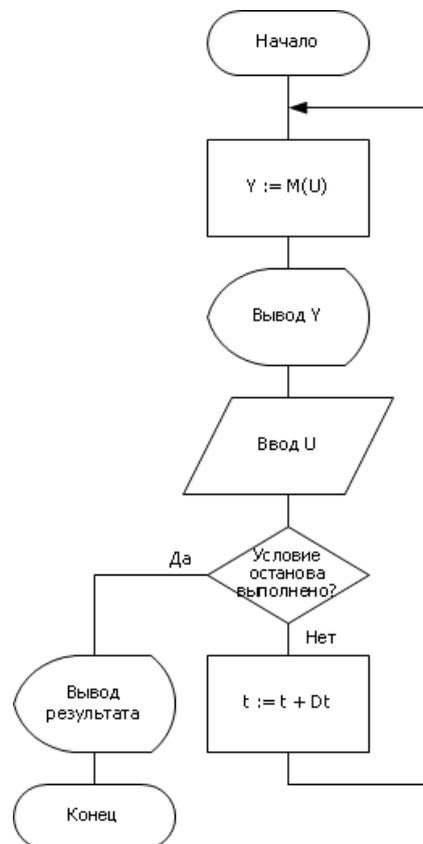


Рис. 20.19. Вид алгоритма реализации тренажера

Тренажер должен обладать качествами наблюдаемости и управляемости. Оператор может и должен судить о качестве своих действий, только наблю-

дая какие-то важные для себя результаты на выходе. И модель должна быть построена таким образом, чтобы можно было достичь хотя бы в принципе искомых результатов какими-то входными воздействиями на нее (управляемость).

Параллельно с моделью может функционировать система оценки деятельности оператора, а также блок автоматического определения наилучших решений, которые могут в определенных режимах (например, режим обучения или подсказки) помогать оператору (см. рис. 20.20). Для этого к модели следует подключить экспертную систему, дающую рекомендации оператору в затруднительных для него случаях.

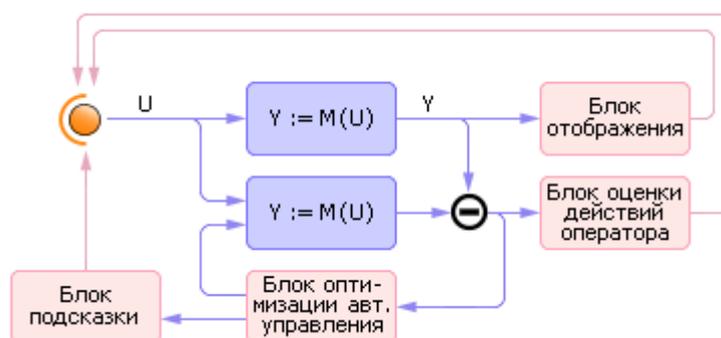


Рис. 20.20. Схема построения тренажера с функциями экспертной системы

Для выработки устойчивых навыков у персонала в процессе тренажа в информацию вносят дополнительные помехи, имитирующие реальные сложности, возникающие на объекте. Можно вносить помехи на входе (нарушается управляемость), на выходе (нарушается наблюдаемость) или в переменные состояния модели (см. рис. 20.21). Следует различать равнодушно действующие помехи и целенаправленное противодействие. В первом случае речь идет о случайном процессе, мешающем оператору достичь цель. Случайная помеха может, как увеличить свое значение, так и равновероятно уменьшить его, то есть чаще всего среднее значение помехи на большом интервале времени равно нулю. Во втором случае речь идет о целенаправленной дезинформации оператора (среднее ее действия не равно нулю).

Для тренажа играет большую роль среднее значение и дисперсия величины помех, которые постепенно наращивают с ростом опыта оператора.

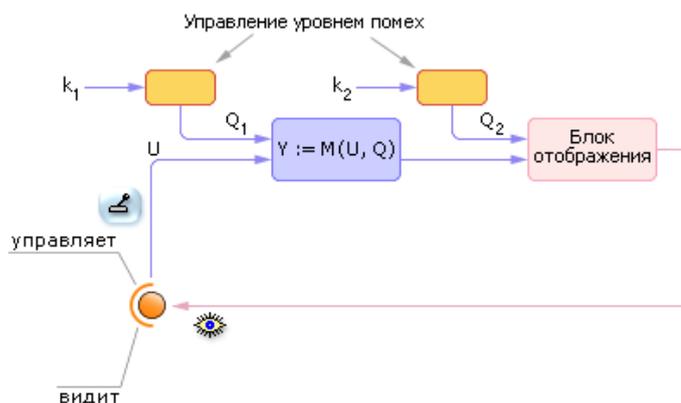


Рис. 20.21. Схема тренажера, дополненного генератором помех

Важнейшими понятиями для системы являются управление, помехи, цель. Системная характеристика должна связать эти понятия вместе. Методика снятия характеристик такова [9].

Выделяем переменные U (вход, управление) и Y (выход, цель) для исследования.

Закрепляем остальные X в виде некоторого фиксированного значения. Для каждого U из диапазона допустимых значений $U_{\min} \leq U \leq U_{\max}$ наблюдаем и фиксируем в табл. 20.2 результат Y .

Таблица 20.2.
Зависимость результата от управления при отсутствии действия помехи.
Точки зависимости сняты как результат работы имитационной модели

Q (помехи)	U (управление)	Y (результат)
0	0	0
0	1	10
0	2	15
0	3	17
0	4	18

На графике (см. рис. 20.22) строим точку с координатами (U, Y) . В результате ряда экспериментов получается кривая $Y(U)$, которая показывает зависимость выхода от входа, цели от управления.

Обычно, если мы имеем дело моделью, отражающей сложную систему достаточно реально, с большой степенью адекватности, зависимость $Y(U)$ должна иметь примерно такую зависимость, как показано на рис. 20.22.

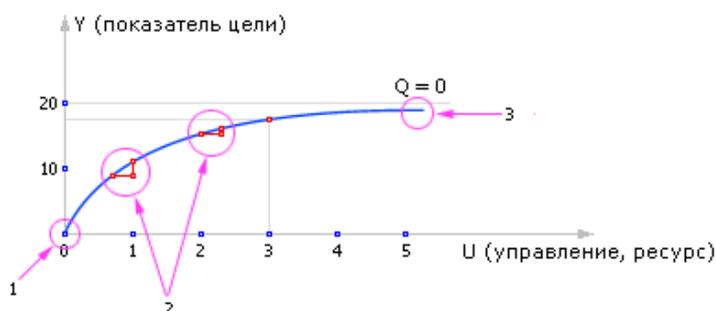


Рис. 20.22. Примерный вид зависимости показателя цели от управления (ресурса), характерный для сложных систем

На рис. 20.22 четко видны следующие закономерности.

1. Кривая исходит из нуля. Если не управлять, не вкладывать ресурс, не стараться, сама по себе система вряд ли выдаст тот результат, который нужен вам, который вам полезен.
2. Кривая нелинейна. Сначала (при малых значениях U) приложение даже небольших усилий ведет к хорошему приросту показателя цели. Далее (при больших значениях U) каждый очередной успех ΔY дается все большим усилием ΔU . Говоря математическим языком, сначала при малых U , производная $\Delta Y / \Delta U$ имеет большое значение, с ростом U ее значение уменьшается.

3. Кривая имеет затухающий характер, стремится к насыщению. Действительно, представьте, если вы вложите огромные значения ресурса (например, затратите на рекламу своей автозаправки в Перми миллиард рублей), то вряд ли получите соответствующую прибыль. Прибыль Y на этом участке U будет лимитирована уже другими факторами (наличие желающих заправиться будет не больше, чем число автовладельцев в городе, например). Реальная система рано или поздно, но выходит на определенный «предел», на «упор», насыщение.

Выделите на объекте переменную помеха Q . По смыслу эта переменная должна мешать достигать цель и не зависеть от воли владельца, управлять ею он не может.

Далее сменить значение Q (ранее мы считали, что оно равно 0) и провести все описанные выше действия снова (см. табл. 20.3).

Таблица 20.3.
Зависимость результата от управления при повышенном уровне действия помехи.
Точки зависимости сняты как результат работы имитационной модели

Q (помехи)	U (управление)	Y (результат)
1	0	0
1	1	8
1	2	13
1	3	15
1	4	16

Построить по таблице экспериментов график (см. рис. 20.23). Очевидно, что при действии возмущений Q график 2 пройдет ниже, чем график 1, поскольку наличие помехи означает, что для достижения того же эффекта Y следует приложить больше управляющих усилий U . Заметим, что менять Q во время изменения U не следует, чтобы четко видеть связь Y именно от U .

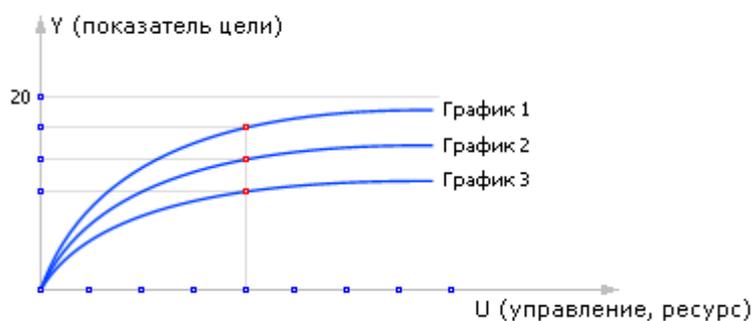


Рис. 20.23. Примерный вид зависимости показателя цели от управления (ресурса) и возмущения, характерный для сложных систем

Если помеха все-таки во время снятия кривой $Y(U)$ меняется самопроизвольно, а это бывает в том случае, когда помеха носит случайный характер, то следует для нанесения одной точки на график сначала провести несколько экспериментов при одном и том же U , а потом усреднить результат Y . Сред-

няя величина достоверна, чем одна из случайных реализаций. Сколько надо провести экспериментов для усреднения, чтобы обеспечить заданную точность ответа, мы обсудим с вами позднее в лекции 21 и 34.

Снова увеличьте Q и снова проведите эксперименты, и снова получите новую таблицу (см. табл. 20.4) и новый график (см. рис. 20.23) — $Y(U)$. В результате вы получите семейство кривых 1-2-3, отражающих зависимость цели, как от управления, так и от помехи.

Таблица 20.4.
Зависимость результата от управления при высоком уровне действия помехи.
Точки зависимости сняты как результат работы имитационной модели

Q (помехи)	U (управление)	Y (результат)
2	0	0
2	1	5
2	2	10
2	3	12
2	4	13

Снятие экспериментальных данных закончено. Теперь в любой момент при заданных Q и Y , используя графики 1, 2, 3, вы можете предсказать результат — необходимый для достижения цели Y уровень управления U . Такая задача, напомним, называется **обратной**.

Теперь, используя снятые зависимости, полезно найти наилучшие решения среди множества возможных. Для этого на каждом графике 1, 2, 3 дополнительно построим линию затрат S (spending), так как управление всегда чего-то стоит, и чем больше вы используете этот ресурс, тем больше придется за это платить. Наклон этой линии указывает на цену ресурса (см. рис. 20.24).

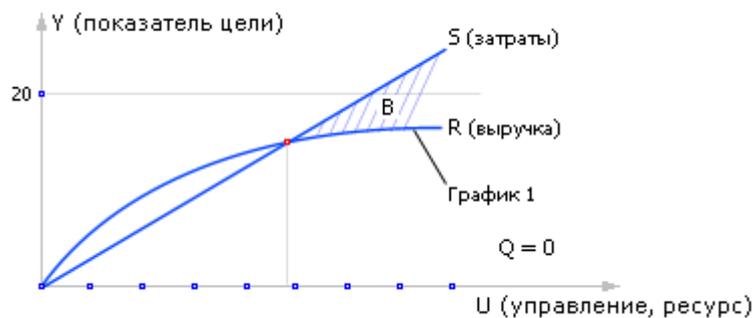


Рис. 20.24. Совмещенные графики выручки от реализации цели и затрат на ее достижение

Цена на ресурс неизменна и не зависит от того, сколько вы его используете (хотя, заметим, что бывают оптовые скидки).

Мы должны максимизировать цель Y . Тогда кривая R (receipts), выраженная в стоимостных единицах, символизирует выручку, а линия S (spending), выраженная в тех же стоимостных единицах символизирует затраты. Если вычесть из выручки затраты, то есть вычесть по точкам один график из другого ($R - S$), то получим в итоге прибыль P (profit): $P = R - S$. А именно, то, как прибыль зависит от управления (см. рис. 20.25).

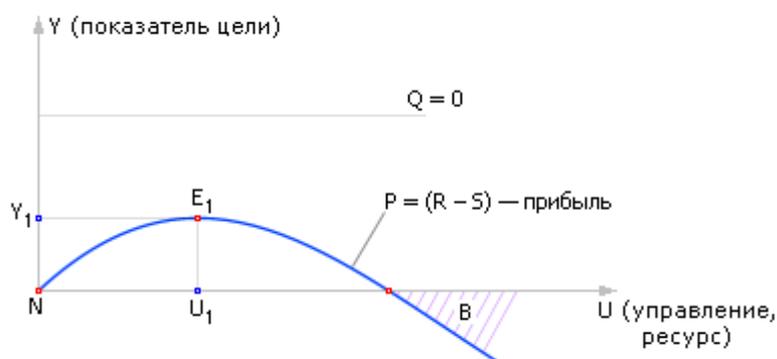


Рис. 20.25. Суммарный график прибыли, полученной (выручка минус затраты) в зависимости от величины управления (ресурса) U . Наилучшее решение — максимум прибыли — точка E_1 , наилучшее управление — U_1

Зона B (bankrupt) — зона банкротства, точка N (null) — точка «ничего не делай и нечего не имей» и точка E_1 (extremum) — зона наибольшей прибыли. Получить прибыли больше, чем Y_1 при этом уровне, помех Q не удастся. Эта точка символизирует тот простой факт, что результат, достигнутый любой ценой, не окупает чрезмерных усилий по его достижению, «все хорошо в меру». Любые управляющие воздействия, даже большие, чем U_1 , дают худший результат [9].

Аналогично найдем точку E на остальных графиках 2, 3 — E_2, E_3 .

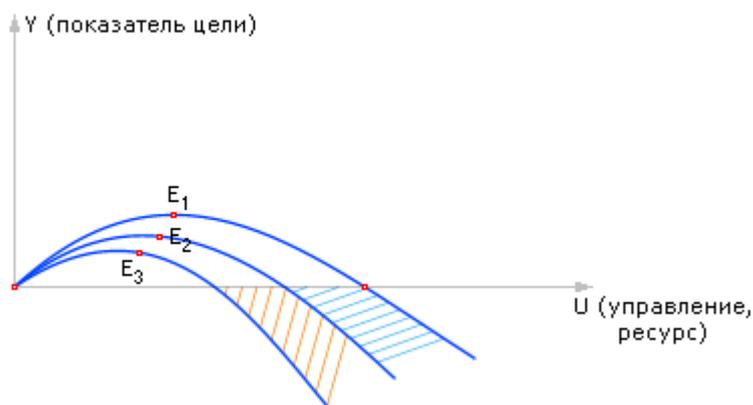


Рис. 20.26. Графики прибыли, в зависимости от величины управления (ресурса) U и возмущения Q . Точки наилучших решений E_i — максимум прибыли. Соответствующие им наилучшие управления — U_i

Сведем все точки E со всех трех графиков на один новый график (см. рис. 20.27).

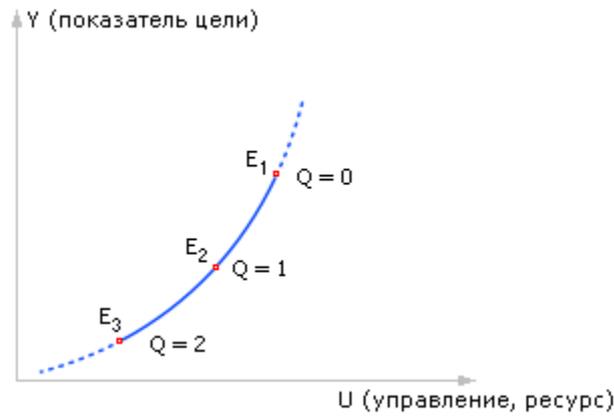


Рис. 20.27. Итоговый график наилучших решений E по критерию прибыли Y в зависимости от величины управления (ресурса) U и возмущений Q

Мы получили замечательную зависимость «кривая оптимальных значений цели Y в зависимости от наилучших решений U при заданном уровне помех Q », по которой можно узнать оптимальные прилагаемые управляющие усилия, необходимые для того, чтобы достичь наилучшего в этих условиях результата. Назовем эту кривую «взаимозависимость цели, управления и помех».

На графике видно, что наибольшая достижимая возможная прибыль уменьшается от точки к точке с увеличением величины помехи — точка E смещается. Например, возможен случай, если помехи очень сильны, а ресурс имеет фиксированную цену, то, возможно, что лучше ничего не делать.

Учтя вышесказанное и возвращаясь к лекции 01 (рис. 1.9—1.10), еще раз обратим внимание, что построение и использование моделей в составе программных продуктов — перспективное новое направление в проектировании программного обеспечения. Изучение оптимальных вариантов действий по управлению предприятием должно быть обеспечено инструментами моделирования.

21. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Статистическое моделирование — базовый метод моделирования, заключающийся в том, что модель испытывается множеством случайных сигналов с заданной плотностью вероятности. Целью является статистическое определение выходных результатов. В основе статистического моделирования лежит *метод Монте-Карло*. Имитацию используют тогда, когда другие методы применить невозможно [9].

Метод Монте-Карло

Рассмотрим метод Монте-Карло на примере вычисления интеграла, значение которого аналитическим способом найти не удастся.

Задача 1. Найти значение интеграла:

$$y = \int_{x_1}^{x_2} f(x) dx$$

На рис. 21.1 представлен график функции $f(x)$. Вычислить значение интеграла этой функции — значит, найти площадь под этим графиком.

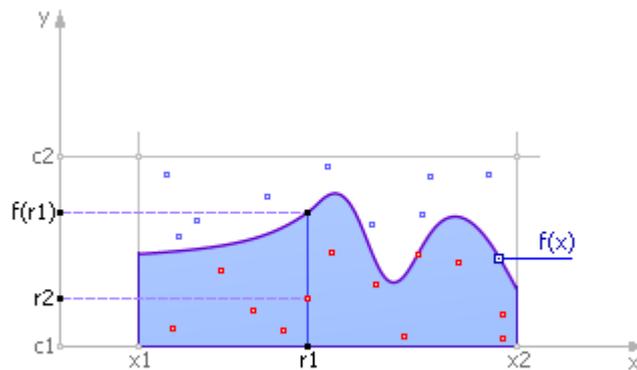


Рис. 21.1. Определение значения интеграла методом Монте-Карло

Ограничиваем кривую сверху, справа и слева. Случайным образом определяем точки в прямоугольнике поиска. Обозначим через N_1 количество точек, принятых для испытаний (то есть попавших в прямоугольник, эти точки изображены на рис. 21.1 красным и синим цветом), и через N_2 — количество точек под кривой, то есть попавших в закрашенную площадь под функцией (эти точки изображены на рис. 21.1 красным цветом). Тогда естественно предположить, что количество точек, попавших под кривую по отношению к общему числу точек пропорционально площади под кривой (величине интеграла) по отношению к площади испытываемого прямоугольника. Математически это можно выразить так:

$$\frac{N_2}{N_1} = \frac{y}{(x_2 - x_1)(c_2 - c_1)}$$

Рассуждения эти, конечно, статистические и тем более верны, чем большее число испытываемых точек мы возьмем.

Фрагмент алгоритма метода Монте-Карло в виде блок-схемы выглядит так, как показано на рис. 21.2.

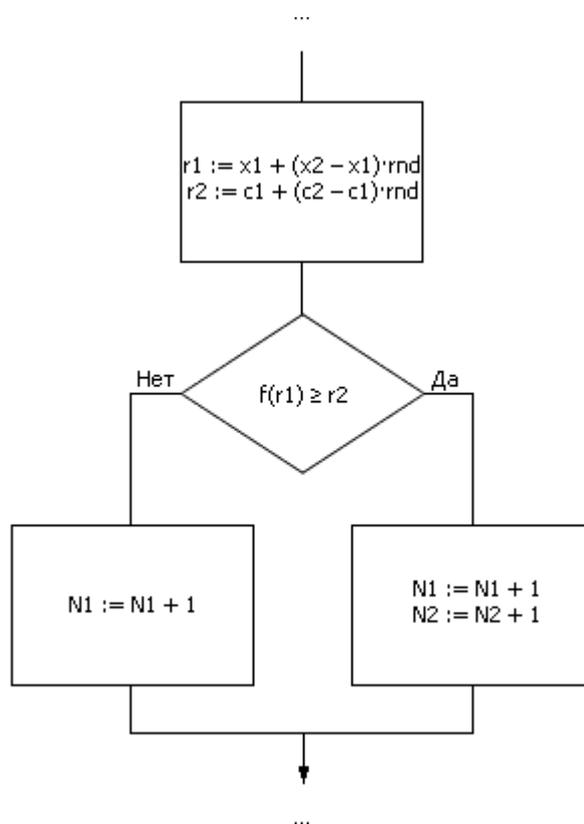


Рис. 21.2. Фрагмент алгоритма реализации метода Монте-Карло
Значения r_1 и r_2 на рис. 21.2 являются равномерно распределенными случайными числами из интервалов $(x_1; x_2)$ и $(c_1; c_2)$ соответственно.

Метод Монте-Карло чрезвычайно эффективен, прост, но необходим «хороший» генератор случайных чисел. Вторая проблема применения метода заключается в определении объема выборки, то есть количества точек, необходимых для обеспечения решения с заданной точностью. Эксперименты показывают: чтобы увеличить точность в 10 раз, объем выборки нужно увеличить в 100 раз; то есть точность примерно пропорциональна корню квадратному из объема выборки:

$$\text{точность} \cong \sqrt{\text{объем выборки}}$$

Схема использования метода Монте-Карло при исследовании систем со случайными параметрами

Построив модель системы со случайными параметрами, на ее вход подают входные сигналы от генератора случайных чисел (ГСЧ), как показано на рис. 21.3. ГСЧ устроен так, что он выдает *равномерно распределенные* случайные числа r_{pp} из интервала $[0; 1]$. Так как одни события могут быть более вероятными, другие — менее вероятными, то равномерно распределенные случайные числа от генератора подают на преобразователь закона случайных чисел (ПЗСЧ), который преобразует их в *заданный* пользователем закон распределения вероятности, например, в нормальный или экспоненциальный закон. Эти числа x подают на вход модели. Модель обрабатывает

входной сигнал x по некоторому закону $y = \varphi(x)$ и получает выходной сигнал y , который также является случайным.

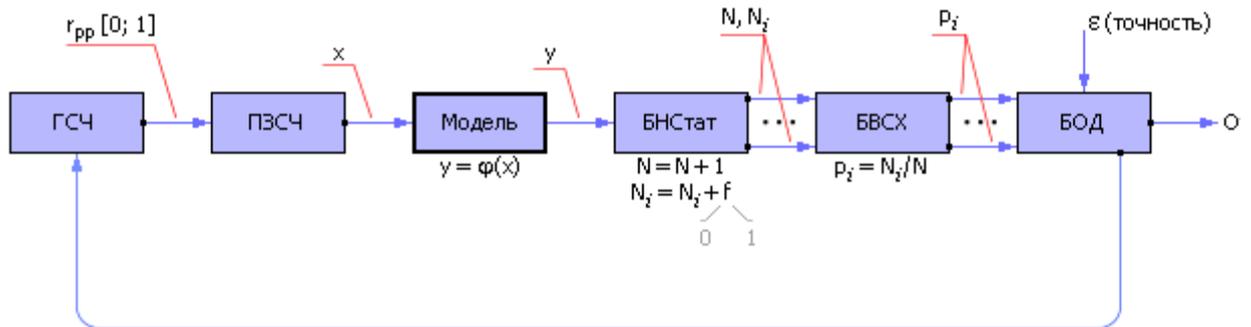


Рис. 21.3. Общая схема метода статистического моделирования

В блоке накопления статистики (БНСтат) установлены фильтры и счетчики. Фильтр (некоторое логическое условие) определяет по значению y , реализовалось ли в конкретном опыте некоторое событие (выполнилось условие, $f = 1$) или нет (условие не выполнилось, $f = 0$). Если событие реализовалось, то счетчик события увеличивается на единицу. Если событие не реализовалось, то значение счетчика не меняется. Если требуется следить за несколькими разными типами событий, то для статистического моделирования понадобится несколько фильтров и счетчиков N_i . Всегда ведется счетчик количества экспериментов — N .

Отношение N_i к N , рассчитываемое в блоке вычисления статистических характеристик (БВСХ) по методу Монте-Карло, дает оценку вероятности p_i появления события i , то есть указывает на частоту его выпадения в серии из N опытов. Это позволяет сделать выводы о статистических свойствах моделируемого объекта.

Событие А совершилось в результате проведенных 200 экспериментов 50 раз. Это означает, согласно методу Монте-Карло, что вероятность совершения события равна: $p_A = 50/200 = 0.25$. Вероятность того, что событие не совершится, равна, соответственно, $1 - 0.25 = 0.75$.

Когда говорят о вероятности, полученной экспериментально, то ее называют **частотью**; слово **вероятность** употребляют, когда хотят подчеркнуть, что речь идет о теоретическом понятии.

При большом количестве опытов N частота появления события, полученная экспериментальным путем, стремится к значению теоретической вероятности появления события.

В блоке оценки достоверности (БОД) анализируют степень достоверности статистических экспериментальных данных, снятых с модели (принимая во внимание точность результата ε , заданную пользователем) и определяют необходимое для этого количество статистических испытаний. Если колебания значений частоты появления событий относительно теоретической вероятности меньше заданной точности, то экспериментальную частоту принимают в качестве ответа, иначе генерацию случайных входных воздействий

продолжают, и процесс моделирования повторяется. При малом числе испытаний результат может оказаться недостоверным. Но чем более испытаний, тем точнее ответ, согласно центральной предельной теореме.

Оценивание ведут по худшей из частот, это обеспечивает достоверный результат сразу по всем снимаемым характеристикам модели.

Пример 1. Какова вероятность выпадения монеты орлом кверху при падении ее с высоты случайным образом?

Начнем подбрасывать монетку и фиксировать результаты каждого броска (см. табл. 21.1).

Таблица 21.1.
Результаты испытаний бросания монеты

Количество опытов N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Значение счетчика выпадения орла N_o	0	0	1	1	2	3	4
Значение счетчика выпадения решки N_p	1	2	2	3	3	3	3
Частость выпадения орла $P_o = N_o/N$	0	0	0.33	0.25	0.4	0.5	0.57
Частость выпадения решки $P_p = N_p/N$	1	1	0.66	0.75	0.6	0.5	0.43

Будем подсчитывать частость выпадения орла как отношение количества случаев выпадения орла к общему числу наблюдений. Посмотрите в табл. 21.1. случаи для $N = 1$, $N = 2$, $N = 3$ — сначала значения частости нельзя назвать достоверными. Попробуем построить график зависимости P_o от N — и посмотрим, как меняется частость выпадения орла в зависимости от количества проведенных опытов. Разумеется, при различных экспериментах будут получаться разные таблицы и, следовательно, разные графики. На рис. 21.4 показан один из вариантов.

Статистическое моделирование

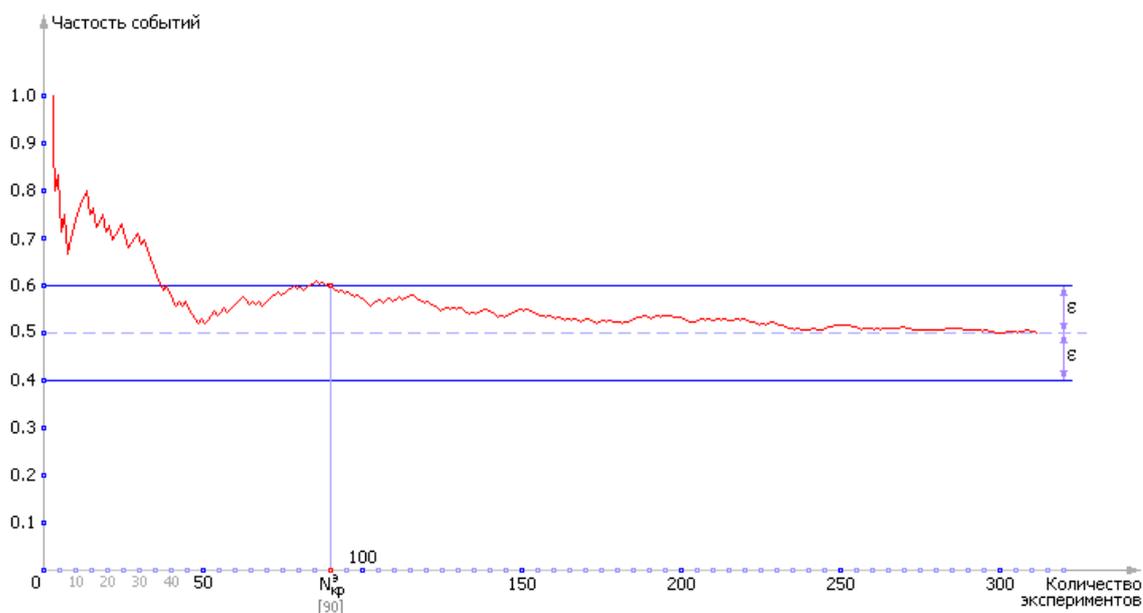


Рис. 21.4. Экспериментальная зависимость частоты появления случайного события от количества наблюдений и ее стремление к теоретической вероятности

1. При малых значениях N , например, $N = 1$, $N = 2$, $N = 3$ ответу вообще доверять нельзя. Например, $P_o = 0$ при $N = 1$, то есть вероятность выпадения орла при одном броске равна нулю! Хотя всем хорошо известно, что это не так. Пока мы получили очень грубый ответ. Посмотрите на график: в процессе **накопления** информации ответ медленно, но верно приближается к правильному (он выделен пунктирной линией). В данном случае правильный ответ нам известен: в идеале, вероятность выпадения орла равна 0.5 (в других, более сложных задачах, ответ нам, конечно, будет неизвестен). Допустим, что ответ нам надо знать с точностью $\varepsilon = 0.1$. Проведем две параллельные линии, отстоящие от правильного ответа 0.5 на расстояние 0.1 (см. рис. 21.4). Ширина образовавшегося коридора будет равна 0.2. Как только кривая $P_o(N)$ войдет в этот коридор так, что уже никогда не покинет, можно остановиться и посмотреть, для какого значения N это произошло. Это и есть **экспериментально вычисленное критическое значение** необходимого количества опытов $N_{кр}^{\varepsilon}$ для определения ответа с точностью $\varepsilon = 0.1$; ε -окрестность в наших рассуждениях играет роль своеобразной трубки точности. Ответы $P_o(91)$, $P_o(92)$ и так далее уже не меняют сильно своих значений (см. рис. 21.4); по крайней мере, у них не изменяется первая цифра после запятой, которой мы обязаны доверять по условиям задачи.
2. Причиной поведения кривой является действие **центральной предельной теоремы** (см. лекцию 25 и лекцию 34). Пока здесь мы формулируем ее в самом простом варианте «Сумма случайных величин

есть величина неслучайная». Мы использовали среднюю величину P_0 , которая несет в себе информацию о сумме опытов, и поэтому постепенно эта величина становится все более достоверной.

Если проделать еще раз этот опыт сначала, то, конечно, его результатом будет другой вид случайной кривой. И ответ будет другим, хотя примерно таким же. Проведем серию экспериментов (см. рис. 21.5). Такая серия называется **ансамблем реализаций**. Какому же ответу в итоге следует верить? Ведь они, хоть и являются близкими, все же разнятся. На практике поступают по-разному. Первый вариант — вычислить среднее значение ответов за несколько реализаций (см. табл. 21.2).

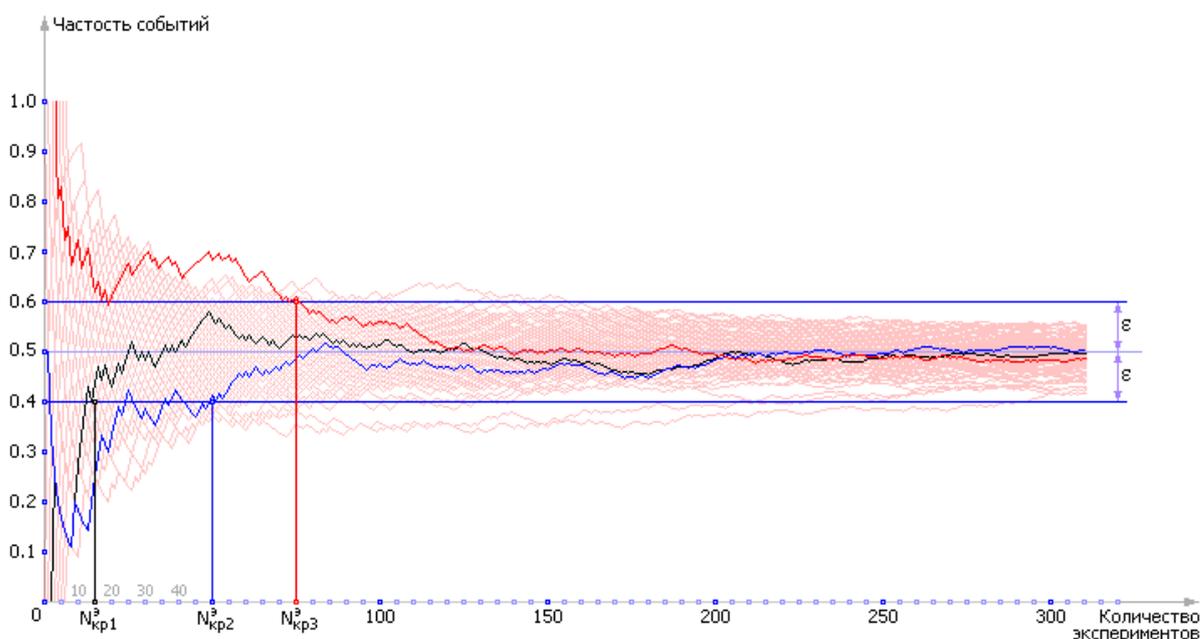


Рис. 21.5. Экспериментально снятый ансамбль случайных зависимостей частоты появления случайного события от количества наблюдений

Мы поставили несколько экспериментов и определяли каждый раз, сколько необходимо было сделать опытов, то есть $N_{кр}^3$. Было проделано 10 экспериментов, результаты которых были сведены в табл. 21.2. По результатам 10-ти экспериментов было вычислено среднее значение $N_{кр}^3$.

Таблица 21.2.

Экспериментальные данные необходимого количества бросков монеты для достижения точности $\varepsilon = 0.1$ при вычислении вероятности выпадения орла

Опыт	$N_{кр}^3$
1	288
2	95
3	50
4	29
5	113
6	210
7	30
8	42

9	39
10	48
Среднее $N_{кр.}^3$	94

Проведя 10 реализаций разной длины, мы определили, что достаточно в среднем было сделать 1 реализацию длиной в 94 броска монеты.

Рассмотрим график на рис. 21.5. На нем нарисовано 100 реализаций — 100 красных линий. Отметьте на нем абсциссу $N = 94$ вертикальной чертой. Есть какой-то процент красных линий, которые не успели пересечь ε -окрестность, то есть $(P^{эксп} - \varepsilon \leq P^{теор} \leq P^{эксп} + \varepsilon)$, и войти в коридор точности до момента $N = 94$. Обратите внимание, таких линий 5. Это значит, что 95 из 100, то есть 95%, линий достоверно вошли в обозначенный интервал.

Проведя 100 реализаций, мы добились примерно 95%-ного доверия к полученной экспериментально величине вероятности выпадения орла, определив ее с точностью 0.1. Для сравнения полученного результата вычислим теоретическое значение $N_{кр.}^T$ теоретически. Для этого придется ввести понятие доверительной вероятности Q_F , которая показывает, насколько мы готовы верить ответу. Например, при $Q_F = 0.95$ мы готовы верить ответу в 95% случаев из 100. Формула теоретического расчета числа экспериментов, которая будет подробно изучаться в лекции 34, имеет вид: $N_{кр.}^T = k(Q_F) \cdot p \cdot (1 - p) / \varepsilon^2$, где $k(Q_F)$ — коэффициент Лапласа, p — вероятность выпадения орла, ε — точность (доверительный интервал). В табл. 21.3 показаны значения теоретической величины количества необходимых опытов при разных Q_F (для точности $\varepsilon = 0.1$ и вероятности $p = 0.5$).

Таблица 21.3.

Теоретический расчет необходимого количества бросков монеты для достижения точности $\varepsilon = 0.1$ при вычислении вероятности выпадения орла

Доверительная вероятность Q_F	Коэффициент Лапласа $k(Q_F)$	Требуемое число опытов $N_{кр.}^T = k(Q_F) \cdot p \cdot (1 - p) / \varepsilon^2$
0.90	2.72	68
0.95	3.84	96
0.99	6.66	167

Полученная нами оценка длины реализации, равная 94 опытам очень близка к теоретической, равной 96. Некоторое несовпадение объясняется тем, что, видимо, 10 реализаций недостаточно для точного вычисления $N_{кр.}^3$. Если вы решите, что вам нужен результат, которому следует доверять больше, то измените значение доверительной вероятности. Теория говорит нам, что если опытов будет 167, то всего 1-2 линии из ансамбля не войдут в предложенную трубку точности. Количество экспериментов с ростом точности и достоверности растет очень быстро.

Второй вариант, используемый на практике — провести одну реализацию и **увеличить полученное для нее $N_{кр.}^3$ в 2 раза**. Это считают хорошей гарантией точности ответа (см. рис. 21.6).

0.01	9600
0.001	960000

Построим по табл. 21.4 график зависимости $N_{кр}^T(\varepsilon)$ (см. рис. 21.8).

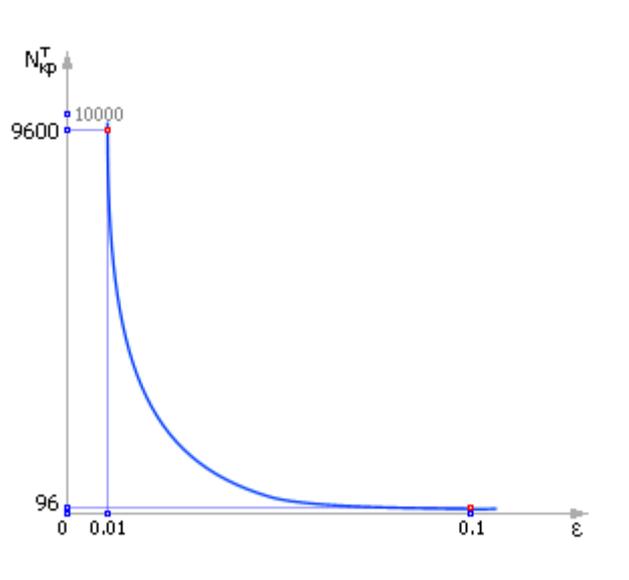


Рис. 21.8. Зависимость числа экспериментов, требуемых для достижения заданной точности ε при фиксированном $Q_F = 0.95$

Рассмотренные графики подтверждают приведенную выше оценку:

$$\text{точность} \cong \sqrt{\text{объем выборки}}$$

Заметим, что оценок точности может быть несколько. Некоторые из них будут еще обсуждаться в лекции 34.

Пример 2. Нахождение площади фигуры методом Монте-Карло. Определим методом Монте-Карло площадь пятиугольника с координатами углов $(0, 0)$, $(0, 10)$, $(5, 20)$, $(10, 10)$, $(7, 0)$.

Нарисуем в двухмерных координатах заданный пятиугольник, вписав его в прямоугольник, чья площадь, как нетрудно догадаться, составляет $(10 - 0) \cdot (20 - 0) = 200$ (см. рис. 21.9).

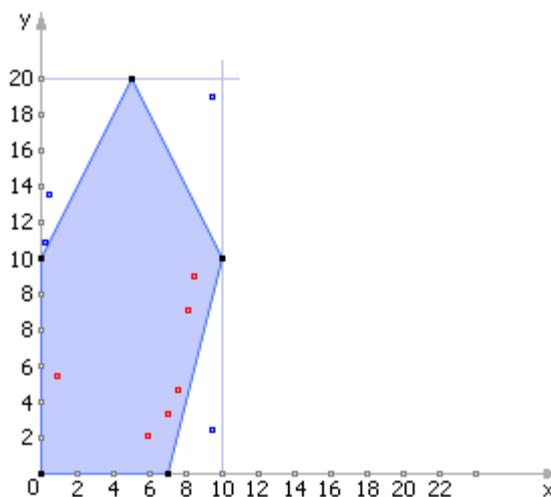


Рис. 21.9. График к решению задачи о площади фигуры методом Монте-Карло

Используем таблицу случайных чисел для генерации пар чисел R, G , равномерно распределенных в интервале от 0 до 1. Число R будет имитировать координату X ($0 \leq X \leq 10$), следовательно, $X = 10 \cdot R$. Число G будет имитировать координату Y ($0 \leq Y \leq 20$), следовательно, $Y = 20 \cdot G$. Сгенерируем по 10 чисел R и G и отобразим 10 точек $(X; Y)$ на рис. 21.9 и в табл. 21.5 [9].

Таблица 21.5.
Решение задачи методом Монте-Карло

Номер точки	R	G	X	Y	Точка (X; Y) попала в прямоугольник?	Точка (X; Y) попала в пятиугольник?
1	0.8109	0.3557	8.109	7.114	Да	Да
2	0.0333	0.5370	0.333	10.740	Да	Нет
3	0.1958	0.2748	1.958	5.496	Да	Да
4	0.6982	0.1652	6.982	3.304	Да	Да
5	0.9499	0.1090	9.499	2.180	Да	Нет
6	0.7644	0.2194	7.644	4.388	Да	Да
7	0.8395	0.4510	8.395	9.020	Да	Да
8	0.0415	0.6855	0.415	13.710	Да	Нет
9	0.5997	0.1140	5.997	2.280	Да	Да
10	0.9595	0.9595	9.595	19.190	Да	Нет
Всего:					10	6

Статистическая гипотеза заключается в том, что количество точек, попавших в контур фигуры, пропорционально площади фигуры: $6:10 = S:200$. То есть, по формуле метода Монте-Карло, получаем, что площадь S пятиугольника равна: $200 \cdot 6/10 = 120$.

Проследим, как менялась величина S от опыта к опыту (см. табл. 21.6).

Таблица 21.6.
Оценка точности ответа

Количество испытаний N	Оценка вероятности попадания случайной точки в испытуемую область	Оценка площади S методом Монте-Карло
1	$1/1 = 1.00$	200
2	$1/2 = 0.50$	100
3	$2/3 = 0.67$	133
4	$3/4 = 0.75$	150
5	$3/5 = 0.60$	120
6	$4/6 = 0.67$	133
7	$5/7 = 0.71$	143
8	$5/8 = 0.63$	125
9	$6/9 = 0.67$	133
10	$6/10 = 0.60$	120

В ответе все еще меняется значение второго разряда, то возможная неточность составляет пока больше 10%. Точность расчета может быть увели-

чена с ростом числа испытаний (см. рис. 21.10).

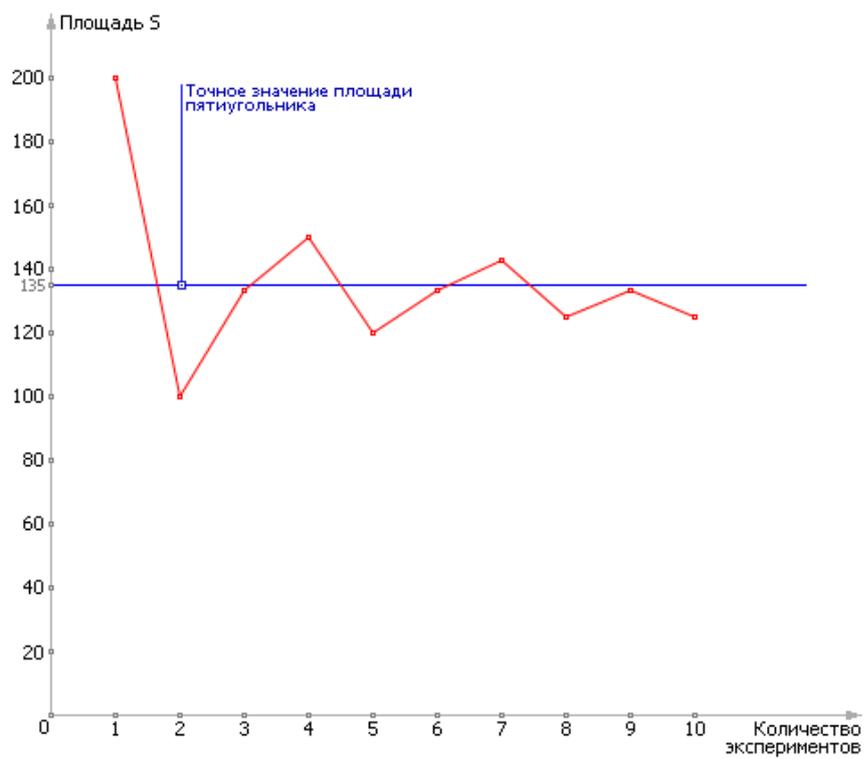


Рис. 21.10. Процесс сходимости определяемого экспериментально ответа к теоретическому результату

22.ГЕНЕРАТОРЫ СЛУЧАЙНЫХ ЧИСЕЛ

В основе метода Монте-Карло (Лекция 21.) лежит генерация случайных чисел, которые должны быть равномерно распределены в интервале (0; 1).

Если генератор выдает числа, смещенные в какую-то часть интервала (одни числа выпадают чаще других), то результат решения задачи, решаемой статистическим методом, может оказаться неверным. Поэтому проблема использования хорошего генератора действительно случайных и действительно равномерно распределенных чисел стоит очень остро.

Математическое ожидание m_r и дисперсия D_r такой последовательности, состоящей из n случайных чисел r_i , должны быть следующими (если это действительно равномерно распределенные случайные числа в интервале от 0 до 1):

$$m_r = \frac{\sum_{i=1}^n r_i}{n} = 0.5$$

$$D_r = \frac{\sum_{i=1}^n (r_i - m_r)^2}{n} = \frac{1}{12}$$

Если пользователю потребуется, чтобы случайное число x находилось в интервале $(a; b)$, отличном от $(0; 1)$, нужно воспользоваться формулой $x = a + (b - a) \cdot r$, где r — случайное число из интервала $(0; 1)$. Законность данного преобразования демонстрируется на рис. 22.1.

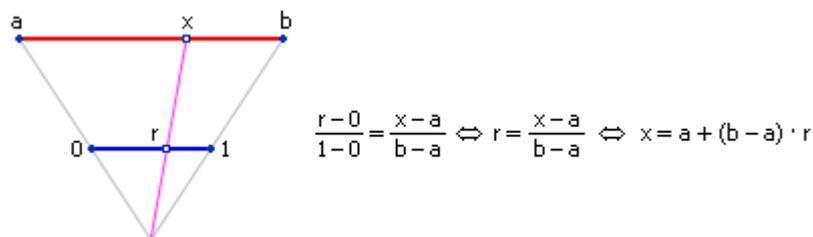


Рис. 22.1. Схема перевода числа из интервала (0; 1) в интервал (a; b)

Теперь x — случайное число, равномерно распределенное в диапазоне от a до b .

За **эталон генератора случайных чисел (ГСЧ)** принят такой генератор, который порождает **последовательность** случайных чисел с *равномерным* законом распределения в интервале $(0; 1)$. За одно обращение данный генератор возвращает одно случайное число. Если наблюдать такой ГСЧ достаточно длительное время, то окажется, что, например, в каждый из десяти интервалов $(0; 0.1)$, $(0.1; 0.2)$, $(0.2; 0.3)$, ..., $(0.9; 1)$ попадет практически одинаковое количество случайных чисел — то есть они будут распределены равномерно по всему интервалу $(0; 1)$. Если изобразить на графике $k = 10$ ин-

тервалов и частоты N_i попаданий в них, то получится экспериментальная кривая плотности распределения случайных чисел (см. рис. 22.2).

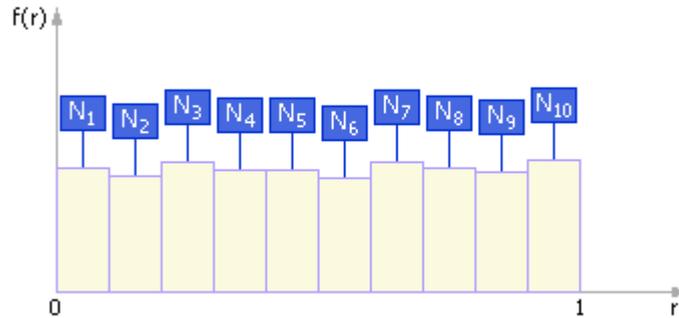


Рис. 22.2. Частотная диаграмма выпадения случайных чисел, порождаемых реальным генератором

В идеале кривая плотности распределения случайных чисел выглядела бы так, как показано на рис. 22.3. То есть в идеальном случае в каждый интервал попадает одинаковое число точек: $N_i = N/k$, где N — общее число точек, k — количество интервалов, $i = 1, \dots, k$.

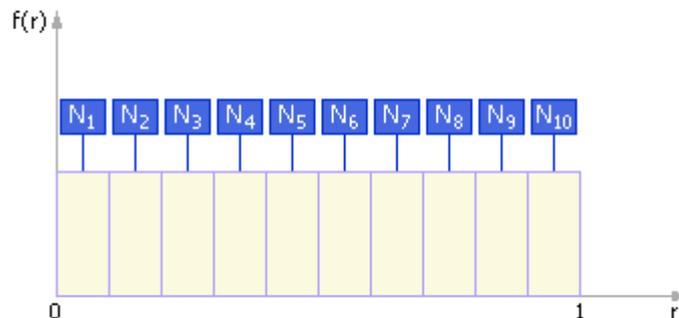


Рис. 22.3. Частотная диаграмма выпадения случайных чисел, порождаемых идеальным генератором теоретически

Следует помнить, что генерация произвольного случайного числа состоит из двух этапов [9]:

- генерация нормализованного случайного числа (то есть равномерно распределенного от 0 до 1);
- преобразование нормализованных случайных чисел r_i в случайные числа x_i , которые распределены по необходимому пользователю (произвольному) закону распределения или в необходимом интервале.

Генераторы случайных чисел по способу получения чисел делятся на:

- физические;
- табличные;
- алгоритмические.

Физические ГСЧ

Примером физических ГСЧ могут служить: монета («орел» — 1, «решка» — 0); игральные кости; поделенный на секторы с цифрами барабан со

стрелкой; аппаратный генератор шума (ГШ), в качестве которого используют шумящее тепловое устройство, например, транзистор (рис. 22.4–22.5).

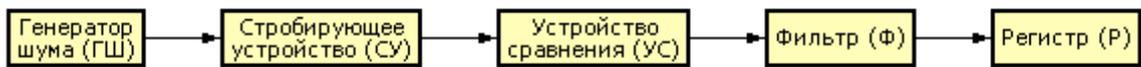


Рис. 22.4. Схема аппаратного метода генерации случайных чисел

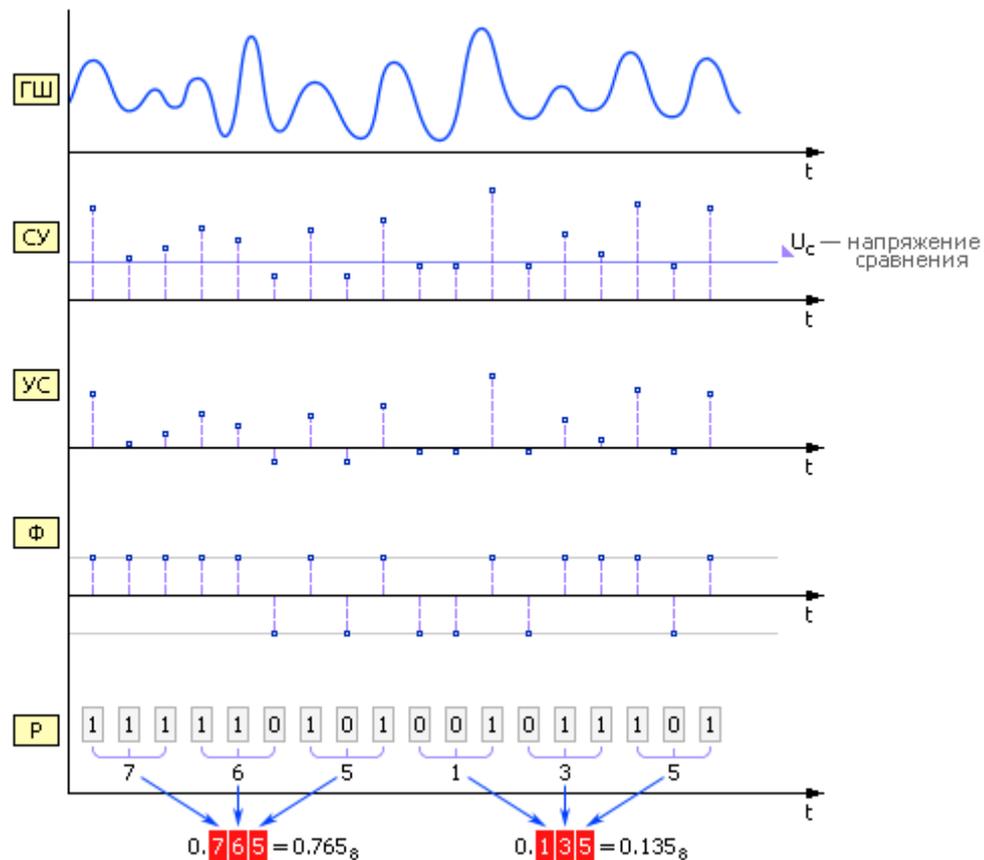


Рис. 22.5. Диаграмма получения случайных чисел аппаратным методом

Задача «Генерация случайных чисел при помощи монеты»

Сгенерируйте случайное трехразрядное число, распределенное по равномерному закону в интервале от 0 до 1, с помощью монеты. Точность — три знака после запятой.

Первый способ решения задачи

Подбросьте монету 9 раз, и если монета упала решкой, то запишите «0», если орлом, то «1». Итак, допустим, что в результате эксперимента получили случайную последовательность 100110100.

Начертите интервал от 0 до 1. Считывая числа в последовательности слева направо, разбивайте интервал пополам и выбирайте каждый раз одну из частей очередного интервала (если выпал 0, то левую, если выпала 1, то правую). Таким образом, можно добраться до любой точки интервала, сколько угодно точно.

1: интервал $[0; 1]$ делится пополам — $[0; 0.5]$ и $[0.5; 1]$, — выбирается правая половина, интервал сужается: $[0.5; 1]$. Следующее число, **0**: интервал $[0.5; 1]$ делится пополам — $[0.5; 0.75]$ и $[0.75; 1]$, — выбирается левая половина $[0.5; 0.75]$, интервал сужается: $[0.5; 0.75]$. Следующее число, **0**: интервал $[0.5; 0.75]$ делится пополам — $[0.5; 0.625]$ и $[0.625; 0.75]$, — выбирается левая половина $[0.5; 0.625]$, интервал сужается: $[0.5; 0.625]$. Следующее число, **1**: интервал $[0.5; 0.625]$ делится пополам — $[0.5; 0.5625]$ и $[0.5625; 0.625]$, — выбирается правая половина $[0.5625; 0.6250]$, интервал сужается: $[0.5625; 0.6250]$.

По условию точности задачи решение найдено: им является любое число из интервала $[0.5625; 0.6250]$, например, 0.625 .

В принципе, если подходить строго, то деление интервалов нужно продолжить до тех пор, пока левая и правая границы найденного интервала не СОВПАДУТ между собой с точностью до третьего знака после запятой. То есть с позиций точности сгенерированное число уже не будет отличимо от любого числа из интервала, в котором оно находится.

Второй способ решения задачи

Разобьем полученную двоичную последовательность 100110100 на триады: $100, 110, 100$. После перевода этих двоичных чисел в десятичные получаем: $4, 6, 4$. Подставив спереди «0.», получим: 0.464 . Таким методом могут получаться только числа от 0.000 до 0.777 (так как максимум, что можно «выжать» из трех двоичных разрядов — это $111_2 = 7_8$) — то есть, по сути, эти числа представлены в восьмеричной системе счисления. Для перевода *восьмеричного* числа в *десятичное* представление выполним: $0.464_8 = 4 \cdot 8^{-1} + 6 \cdot 8^{-2} + 4 \cdot 8^{-3} = 0.6015625_{10} = 0.602_{10}$.

Итак, искомое число равно: 0.602 .

Табличные ГСЧ

Табличные ГСЧ в качестве источника случайных чисел используют специальным образом составленные таблицы, содержащие проверенные некоррелированные, то есть никак не зависящие друг от друга, цифры. В табл. 22.1 приведен небольшой фрагмент такой таблицы. Обходя таблицу слева направо сверху вниз, можно получать равномерно распределенные от 0 до 1 случайные числа с нужным числом знаков после запятой (в нашем примере мы используем для каждого числа по три знака). Так как цифры в таблице не зависят друг от друга, то таблицу можно обходить разными способами, например, сверху вниз, или справа налево, или, скажем, можно выбирать цифры, находящиеся на четных позициях.

Таблица 22.1.
Случайные цифры. Равномерно распределенные от 0 до 1 случайные числа

Случайные цифры	Равномерно распределенные от 0 до 1 случайные числа
9 2 9 2 0 4 2 6	0.929
9 5 7 3 4 9 0 3	0.204
5 9 1 6 6 5 7 6	0.269



Достоинство данного метода в том, что он дает действительно случайные числа, так как таблица содержит проверенные некоррелированные цифры. Недостатки метода: для хранения большого количества цифр требуется много памяти; большие трудности порождения и проверки такого рода таблиц, повторы при использовании таблицы уже не гарантируют случайности числовой последовательности, а значит, и надежности результата.

Алгоритмические ГСЧ

Числа, генерируемые с помощью этих ГСЧ, всегда являются псевдослучайными (или квазислучайными), то есть каждое последующее сгенерированное число зависит от предыдущего:

$$r_{i+1} = f(r_i).$$

Последовательности, составленные из таких чисел, образуют петли, то есть обязательно существует цикл, повторяющийся бесконечное число раз. Повторяющиеся циклы называются **периодами**.

Достоинством данных ГСЧ является быстроедействие; генераторы практически не требуют ресурсов памяти, компактны. Недостатки: числа нельзя в полной мере назвать случайными, поскольку между ними имеется зависимость, а также наличие периодов в последовательности квазислучайных чисел.

Рассмотрим несколько алгоритмических методов получения ГСЧ:

- метод серединных квадратов;
- метод серединных произведений;
- метод перемешивания;
- линейный конгруэнтный метод.

Метод серединных квадратов

Имеется некоторое четырехзначное число R_0 . Это число возводится в квадрат и заносится в R_1 . Далее из R_1 берется середина (четыре средних цифры) — новое случайное число — и записывается в R_0 . Затем процедура повторяется (см. рис. 22.6). Отметим, что на самом деле в качестве случайного числа необходимо брать не **ghij**, а **0.ghij** — с приписанным слева нулем и десятичной точкой. Этот факт отражен как на рис. 22.6, так и на последующих подобных рисунках.

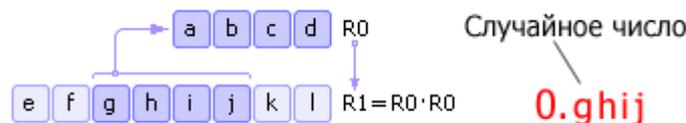


Рис. 22.6. Схема метода серединных квадратов

Недостатки метода: 1) если на некоторой итерации число R_0 станет равным нулю, то генератор вырождается, поэтому важен правильный выбор начального значения R_0 ; 2) генератор будет повторять последовательность через M^n шагов (в лучшем случае), где n — разрядность числа R_0 , M — основание системы счисления.

Для примера на **рис. 22.6**: если число R_0 будет представлено в двоичной системе счисления, то последовательность псевдослучайных чисел повторится через $2^4 = 16$ шагов. Заметим, что повторение последовательности может произойти и раньше, если начальное число будет выбрано неудачно.

Описанный выше способ был предложен Джоном фон Нейманом и относится к 1946 году. Поскольку этот способ оказался ненадежным, от него очень быстро отказались.

Метод серединных произведений

Число R_0 умножается на R_1 , из полученного результата R_2 извлекается середина R_2^* (это очередное случайное число) и умножается на R_1 . По этой схеме вычисляются все последующие случайные числа (см. рис. 22.7).

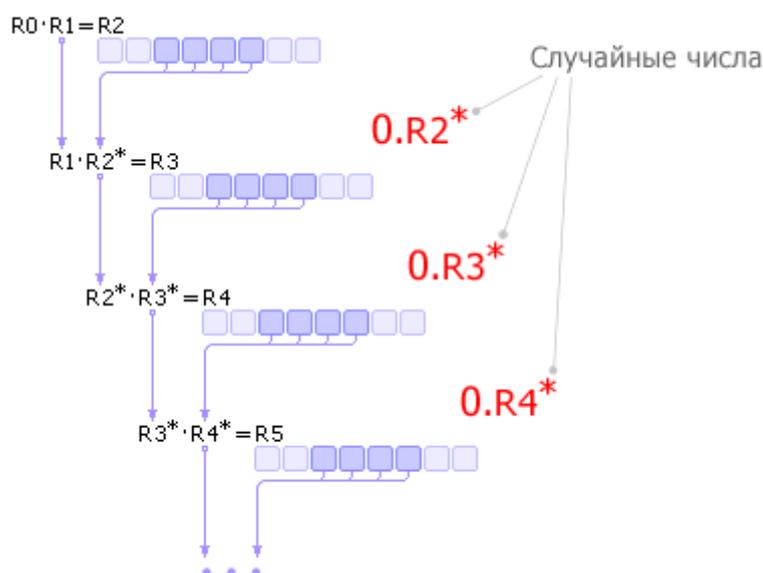


Рис. 22.7. Схема метода серединных произведений

Метод перемешивания

В методе перемешивания используются операции циклического сдвига содержимого ячейки влево и вправо. Идея метода состоит в следующем. Пусть в ячейке хранится начальное число R_0 . Циклически сдвигая содержимое ячейки влево на $1/4$ длины ячейки, получаем новое число R_0^* . Точно так же, циклически сдвигая содержимое ячейки R_0 вправо на $1/4$ длины ячейки, получаем второе число R_0^{**} . Сумма чисел R_0^* и R_0^{**} дает новое случайное число R_1 . Далее R_1 заносится в R_0 , и вся последовательность операций повторяется (см. рис. 22.8).

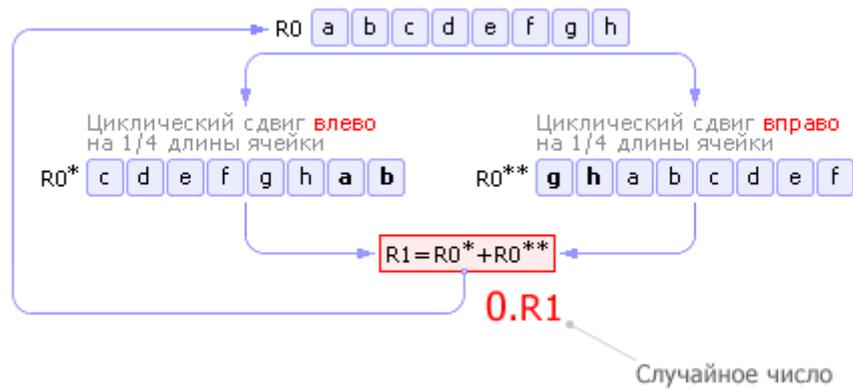


Рис. 22.8. Схема метода перемешивания

Обратите внимание, что число, полученное в результате суммирования $R0^*$ и $R0^{**}$, может не уместиться полностью в ячейке $R1$. В этом случае от полученного числа должны быть отброшены лишние разряды. Поясним это для **рис. 22.8**, где все ячейки представлены восемью двоичными разрядами. Пусть $R0^* = 10010001_2 = 145_{10}$, $R0^{**} = 10100001_2 = 161_{10}$, тогда $R0^* + R0^{**} = 100110010_2 = 306_{10}$. Как видим, число 306 занимает 9 разрядов (в двоичной системе счисления), а ячейка $R1$ (как и $R0$) может вместить в себя максимум 8 разрядов. Поэтому перед занесением значения в $R1$ необходимо убрать один «лишний», крайний левый бит из числа 306, в результате чего в $R1$ пойдет уже не 306, а $00110010_2 = 50_{10}$. Также заметим, что в таких языках, как Паскаль, «урезание» лишних битов при переполнении ячейки производится автоматически в соответствии с заданным типом переменной.

Линейный конгруэнтный метод

Линейный конгруэнтный метод является одной из простейших и наиболее употребительных в настоящее время процедур, имитирующих случайные числа. В этом методе используется операция $\text{mod}(x, y)$, возвращающая остаток от деления первого аргумента на второй. Каждое последующее случайное число рассчитывается на основе предыдущего случайного числа по следующей формуле:

$$r_{i+1} = \text{mod}(k \cdot r_i + b, M).$$

M — модуль ($0 < M$);

k — множитель ($0 \leq k < M$);

b — приращение ($0 \leq b < M$);

r_0 — начальное значение ($0 \leq r_0 < M$).

Последовательность случайных чисел, полученных с помощью данной формулы, называется **линейной конгруэнтной последовательностью**. Многие авторы называют линейную конгруэнтную последовательность при $b = 0$ **мультипликативным конгруэнтным методом**, а при $b \neq 0$ — **смешанным конгруэнтным методом**.

Для качественного генератора требуется подобрать подходящие коэффициенты. Необходимо, чтобы число M было довольно большим, так как период не может иметь больше M элементов. С другой стороны, деление, используемое в этом методе, является довольно медленной операцией, поэтому для двоичной вычислительной машины логичным будет выбор $M = 2^N$, поскольку в этом случае нахождение остатка от деления сводится внутри ЭВМ к двоичной логической операции «AND». Также широко распространен выбор наибольшего простого числа M , меньшего, чем 2^N : в специальной литературе доказывалось, что в этом случае младшие разряды получаемого случайного числа r_{i+1} ведут себя так же случайно, как и старшие, что положительно сказывается на всей последовательности случайных чисел в целом. В качестве примера можно привести одно из чисел Мерсенна, равное $2^{31} - 1$, и таким образом, $M = 2^{31} - 1$.

Одним из требований к линейным конгруэнтным последовательностям является как можно большая длина периода. Длина периода зависит от значений M , k и b . Теорема, которую мы приведем ниже, позволяет определить, возможно ли достижение периода максимальной длины для конкретных значений M , k и b .

Теорема. Линейная конгруэнтная последовательность, определенная числами M , k , b и r_0 , имеет период длиной M тогда и только тогда, когда:

- числа b и M взаимно простые;
- $k - 1$ кратно p для каждого простого p , являющегося делителем M ;
- $k - 1$ кратно 4, если M кратно 4.

Наконец, в заключение рассмотрим пару примеров использования линейного конгруэнтного метода для генерации случайных чисел [9].

Пример 1

$$M = 2^N$$

$$k = 3 + 8 \cdot q \text{ (или } k = 5 + 8 \cdot q)$$

$$b = 0$$

r_0 — нечетно

Было установлено, что ряд псевдослучайных чисел, генерируемых на основе данных из примера 1, будет повторяться через каждые $M/4$ чисел. Число q задается произвольно перед началом вычислений, однако при этом следует иметь в виду, что ряд производит впечатление случайного при больших k (а значит, и q). Результат можно несколько улучшить, если b нечетно и $k = 1 + 4 \cdot q$ — в этом случае ряд будет повторяться через каждые M чисел. После долгих поисков k исследователи остановились на значениях 69069 и 71365.

Пример 2

$$M = 2^{31} - 1$$

$$k = 1\ 220\ 703\ 125$$

$$b = 7$$

$$r_0 = 7$$

Генератор случайных чисел, использующий данные из примера 2, будет выдавать случайные неповторяющиеся числа с периодом, равным 7 миллионам.

Мультипликативный метод генерации псевдослучайных чисел был предложен Д. Г. Лехмером (D. H. Lehmer) в 1949 году.

Проверка качества работы генератора

От качества работы ГСЧ зависит качество работы всей системы и точность результатов. Поэтому случайная последовательность, порождаемая ГСЧ, должна удовлетворять целому ряду критериев.

Осуществляемые проверки бывают двух типов:

- проверки на равномерность распределения;
- проверки на статистическую независимость.

Проверки на равномерность распределения

1) ГСЧ должен выдавать близкие к следующим значения статистических параметров, характерных для равномерного случайного закона:

$$m_r = \frac{\sum_{i=1}^n r_i}{n} \approx 0.5 \quad \text{— математическое ожидание;}$$

$$D_r = \frac{\sum_{i=1}^n (r_i - m_r)^2}{n} \approx 0.0833 \quad \text{— дисперсия;}$$

$$\sigma_r = \sqrt{D_r} \approx 0.2887 \quad \text{— среднее квадратичное отклонение.}$$

2) Частотный тест

Частотный тест позволяет выяснить, сколько чисел попало в интервал $(m_r - \sigma_r; m_r + \sigma_r)$, то есть $(0.5 - 0.2887; 0.5 + 0.2887)$ или, в конечном итоге, $(0.2113; 0.7887)$. Так как $0.7887 - 0.2113 = 0.5774$, заключаем, что в хорошем ГСЧ в этот интервал должно попадать около 57.7% из всех выпавших случайных чисел (см. рис. 22.9).

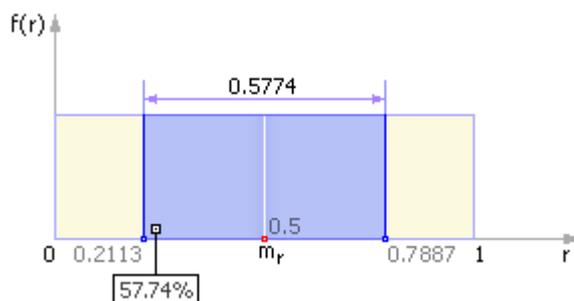


Рис. 22.9. Частотная диаграмма идеального ГСЧ в случае проверки его на частотный тест

Также необходимо учитывать, что количество чисел, попавших в интервал $(0; 0.5)$, должно быть примерно равно количеству чисел, попавших в интервал $(0.5; 1)$.

3) Проверка по критерию «хи-квадрат»

Критерий «хи-квадрат» (χ^2 -критерий) — это один из самых известных статистических критериев; он является основным методом, используемым в сочетании с другими критериями. Критерий «хи-квадрат» был предложен в 1900 году Карлом Пирсоном. Его замечательная работа рассматривается как фундамент современной математической статистики.

Для нашего случая проверка по критерию «хи-квадрат» позволит узнать, насколько созданный нами *реальный* ГСЧ близок к *эталону ГСЧ*, то есть удовлетворяет ли он требованию равномерного распределения или нет.

Частотная диаграмма *эталонного* ГСЧ представлена на рис. 22.10. Так как закон распределения эталонного ГСЧ равномерный, то (теоретическая) вероятность p_i попадания чисел в i -ый интервал (всего этих интервалов k) равна $p_i = 1/k$. И, таким образом, в каждый из k интервалов попадет *ровно* по $p_i \cdot N$ чисел (N — общее количество сгенерированных чисел).

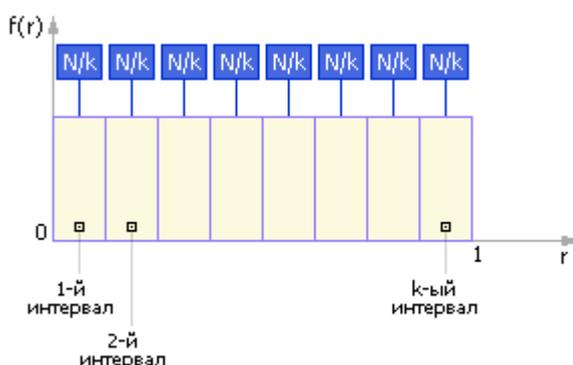


Рис. 22.10. Частотная диаграмма эталонного ГСЧ

Реальный ГСЧ будет выдавать числа, распределенные (причем, не обязательно равномерно!) по k интервалам и в каждый интервал попадет по n_i чисел (в сумме $n_1 + n_2 + \dots + n_k = N$). Как же нам определить, насколько испытываемый ГСЧ хорош и близок к эталонному? Вполне логично рассмотреть квадраты разностей между полученным количеством чисел n_i и «эталонным» $p_i \cdot N$. Сложим их, и в результате получим:

$$\chi_{\text{эксп.}}^2 = (n_1 - p_1 \cdot N)^2 + (n_2 - p_2 \cdot N)^2 + \dots + (n_k - p_k \cdot N)^2.$$

Из этой формулы следует, что чем меньше разность в каждом из слагаемых (а значит, и чем меньше значение $\chi_{\text{эксп.}}^2$), тем сильнее закон распределения случайных чисел, генерируемых реальным ГСЧ, тяготеет к равномерному.

В предыдущем выражении каждому из слагаемых приписывается одинаковый вес (равный 1), что на самом деле может не соответствовать действительности; поэтому для статистики «хи-квадрат» необходимо провести нормировку каждого i -го слагаемого, поделив его на $p_i \cdot N$:

$$\chi_{\text{эксп.}}^2 = \frac{(n_1 - p_1 \cdot N)^2}{p_1 \cdot N} + \frac{(n_2 - p_2 \cdot N)^2}{p_2 \cdot N} + \dots + \frac{(n_k - p_k \cdot N)^2}{p_k \cdot N}$$

Запишем полученное выражение более компактно и упростим его:

$$\chi_{\text{эксп.}}^2 = \sum_{i=1}^k \frac{(n_i - p_i \cdot N)^2}{p_i \cdot N} = \frac{1}{N} \sum_{i=1}^k \left(\frac{n_i^2}{p_i} \right) - N$$

В табл. 22.2 приведены *теоретические* значения «хи-квадрат» ($\chi^2_{\text{теор.}}$), где $\nu = N - 1$ — это число степеней свободы, \mathbf{p} — это доверительная вероятность, задаваемая пользователем, который указывает, насколько ГСЧ должен удовлетворять требованиям равномерного распределения, или \mathbf{p} — *это вероятность того, что экспериментальное значение $\chi^2_{\text{эксп.}}$ будет меньше табулированного (теоретического) $\chi^2_{\text{теор.}}$ или равно ему.*

Таблица 22.2.

Некоторые процентные точки χ^2 -распределения

	p = 1%	p = 5%	p = 25%	p = 50%	p = 75%	p = 95%	p = 99%
$\nu = 1$	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
$\nu = 2$	0.02010	0.1026	0.5754	1.386	2.773	5.991	9.210
$\nu = 3$	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
$\nu = 4$	0.2971	0.7107	1.923	3.357	5.385	9.488	13.28
$\nu = 5$	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
$\nu = 6$	0.8721	1.635	3.455	5.348	7.841	12.59	16.81
$\nu = 7$	1.239	2.167	4.255	6.346	9.037	14.07	18.48
$\nu = 8$	1.646	2.733	5.071	7.344	10.22	15.51	20.09
$\nu = 9$	2.088	3.325	5.899	8.343	11.39	16.92	21.67
$\nu = 10$	2.558	3.940	6.737	9.342	12.55	18.31	23.21
$\nu = 11$	3.053	4.575	7.584	10.34	13.70	19.68	24.72
$\nu = 12$	3.571	5.226	8.438	11.34	14.85	21.03	26.22
$\nu = 15$	5.229	7.261	11.04	14.34	18.25	25.00	30.58
$\nu = 20$	8.260	10.85	15.45	19.34	23.83	31.41	37.57
$\nu = 30$	14.95	18.49	24.48	29.34	34.80	43.77	50.89
$\nu = 50$	29.71	34.76	42.94	49.33	56.33	67.50	76.15
$\nu > 30$	$\nu + \sqrt{2\nu} \cdot x_p + 2/3 \cdot x_p^2 - 2/3 + O(1/\sqrt{\nu})$						
$x_p =$	-2.33	-1.64	-0.674	0.00	0.674	1.64	2.33

Приемлемым считают \mathbf{p} от 10% до 90%.

Если $\chi^2_{\text{эксп.}}$ много больше $\chi^2_{\text{теор.}}$ (то есть \mathbf{p} — велико), то генератор **не удовлетворяет** требованию равномерного распределения, так как наблюдаемые значения n_i слишком далеко уходят от теоретических $p_i \cdot N$ и не могут рассматриваться как случайные. Другими словами, устанавливается такой большой доверительный интервал, что ограничения на числа становятся очень нежесткими, требования к числам — слабыми. При этом будет наблюдаться очень большая абсолютная погрешность.

Еще Д. Кнут в своей книге «Искусство программирования» заметил, что иметь $\chi^2_{\text{эксп.}}$ маленьким тоже, в общем-то, нехорошо, хотя это и кажется, на первый взгляд, замечательно с точки зрения равномерности. Действительно, возьмите ряд чисел 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, ... — они идеальны с точки зрения равномерности, и $\chi^2_{\text{эксп.}}$ будет практически нулевым, но вряд ли вы их признаете случайными.

Если $\chi^2_{\text{эксп.}}$ много меньше $\chi^2_{\text{теор.}}$ (то есть \mathbf{p} — мало), то генератор **не удовлетворяет** требованию случайного равномерного распределения, так как

наблюдаемые значения n_i слишком близки к теоретическим $p_i \cdot N$ и не могут рассматриваться как случайные.

Если $\chi^2_{\text{эксп.}}$ лежит в некотором диапазоне, между двумя значениями $\chi^2_{\text{теор.}}$, которые соответствуют, например, $\mathbf{p} = 25\%$ и $\mathbf{p} = 50\%$, то можно считать, что значения случайных чисел, порождаемые датчиком, вполне являются случайными.

Дополнительно надо иметь в виду, что все значения $p_i \cdot N$ должны быть достаточно большими, например больше 5 (выяснено эмпирическим путем). Только тогда (при достаточно большой статистической выборке) условия проведения эксперимента можно считать удовлетворительными.

Процедура проверки имеет следующий вид.

1. Диапазон от 0 до 1 разбивается на k равных интервалов.
2. Запускается ГСЧ N раз (N должно быть велико, например, $N/k > 5$).
3. Определяется количество случайных чисел, попавших в каждый интервал: n_i , $i = 1, \dots, k$.
4. Вычисляется экспериментальное значение $\chi^2_{\text{эксп.}}$ по следующей формуле:

$$\chi^2_{\text{эксп.}} = \sum_{i=1}^k \frac{(n_i - p_i \cdot N)^2}{p_i \cdot N} = \frac{1}{N} \sum_{i=1}^k \left(\frac{n_i^2}{p_i} \right) - N$$

где $p_i = 1/k$ — теоретическая вероятность попадания чисел в k -ый интервал.

5. Путем сравнения экспериментально полученного значения $\chi^2_{\text{эксп.}}$ с теоретическим $\chi^2_{\text{теор.}}$ (из табл. 22.2) делается вывод о пригодности генератора для использования. Для этого: а) входим в табл. 22.2 (**строка = количество экспериментов – 1**); б) сравниваем вычисленное $\chi^2_{\text{эксп.}}$ с $\chi^2_{\text{теор.}}$, встречающимися в строке. При этом возможно три случая.

Первый случай: $\chi^2_{\text{эксп.}}$ много больше любого $\chi^2_{\text{теор.}}$ в строке — гипотеза о случайности равномерного генератора не выполняется (разброс чисел слишком велик, чтобы быть случайным).

Второй случай: $\chi^2_{\text{эксп.}}$ много меньше любого $\chi^2_{\text{теор.}}$ в строке — гипотеза о случайности равномерного генератора не выполняется (разброс чисел слишком мал, чтобы быть случайным).

Третий случай: $\chi^2_{\text{эксп.}}$ лежит между значениями $\chi^2_{\text{теор.}}$ двух рядом стоящих столбцов — гипотеза о случайности равномерного генератора выполняется с вероятностью \mathbf{p} (то есть в \mathbf{p} случаях из 100).

Заметим, что чем ближе получается \mathbf{p} к значению 50%, тем лучше.

Проверки на статистическую независимость

1) Проверка на частоту появления цифры в последовательности

Рассмотрим пример. Случайное число 0.2463389991 состоит из цифр 2463389991, а число 0.5467766618 состоит из цифр 5467766618. Соединяя последовательности цифр, имеем: 24633899915467766618.

Теоретическая вероятность p_i выпадения i -ой цифры (от 0 до 9) равна 0.1.

Следует вычислить частоту появления каждой цифры в выпавшей экспериментальной последовательности. Например, цифра 1 выпала 2 раза из 20, а цифра 6 выпала 5 раз из 20.

Потом считают оценку и принимают решение по критерию «хи-квадрат».

2) Проверка появления серий из одинаковых цифр

Обозначим через n_L число серий одинаковых подряд цифр длины L . Проверять надо все L от 1 до m , где m — это заданное пользователем число: максимально встречающееся число одинаковых цифр в серии.

В примере «24633899915467766618» обнаружены 2 серии длиной в 2 (33 и 77), то есть $n_2 = 2$ и 2 серии длиной в 3 (999 и 666), то есть $n_3 = 2$.

Вероятность появления серии длиной в L равна: $p_L = 9 \cdot 10^{-L}$ (теоретическая). То есть вероятность появления серии длиной в один символ равна: $p_1 = 0.9$ (теоретическая). Вероятность появления серии длиной в два символа равна: $p_2 = 0.09$ (теоретическая). Вероятность появления серии длиной в три символа равна: $p_3 = 0.009$ (теоретическая).

Например, вероятность появления серии длиной в один символ равна $p_L = 0.9$, так как всего может встретиться один символ из 10, а всего символов 9 (ноль не считается). А вероятность того, что подряд встретится два одинаковых символа «XX» равна $0.1 \cdot 0.1 \cdot 9$, то есть вероятность 0.1 того, что в первой позиции появится символ «X», умножается на вероятность 0.1 того, что во второй позиции появится такой же символ «X» и умножается на количество таких комбинаций 9.

Частость появления серий подсчитывается по ранее разобранный нами формуле «хи-квадрат» с использованием значений p_L .

Примечание: генератор может быть проверен многократно, однако проверки не обладают свойством полноты и не гарантируют, что генератор выдает случайные числа. Например, генератор, выдающий последовательность 12345678912345..., при проверках будет считаться идеальным, что, очевидно, не совсем так [9].

23.МОДЕЛИРОВАНИЕ СЛУЧАЙНОГО СОБЫТИЯ. МОДЕЛИРОВАНИЕ ПОЛНОЙ ГРУППЫ НЕСОВМЕСТНЫХ СОБЫТИЙ

Моделирование случайного события

Начнем с самого простого. Используем наше умение генерировать случайные числа для имитации выпадения случайных событий.

Случайное событие подразумевает, что у некоторого события есть несколько исходов и то, который из исходов произойдет в очередной раз, определяется только его вероятностью. То есть исход выбирается случайно с учетом его вероятности [9].

Нам известна вероятность выпуска бракованных изделий $P_{\text{б}} = 0.1$. Смоделировать выпадение этого события можно, разыграв равномерно распределенное случайное число из диапазона от 0 до 1 и установив, в какой из двух интервалов (от 0 до 0.1 или от 0.1 до 1) оно попало (см. рис. 23.1). Если число попадает в диапазон (0; 0.1), то выпущен брак, то есть событие произошло, иначе — событие не произошло (выпущено кондиционное изделие). При значительном числе экспериментов частота попадания чисел в интервал от 0 до 0.1 будет приближаться к вероятности $P = 0.1$, а частота попадания чисел в интервал от 0.1 до 1 будет приближаться к $P_{\text{к}} = 0.9$.

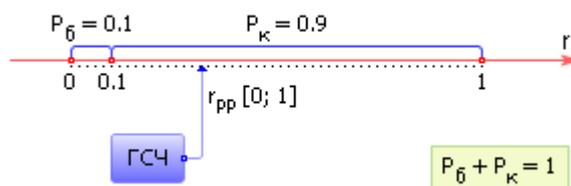


Рис. 23.1. Схема использования генератора случайных чисел для имитации случайного события
Фрагмент алгоритма представлен на рис. 23.2.

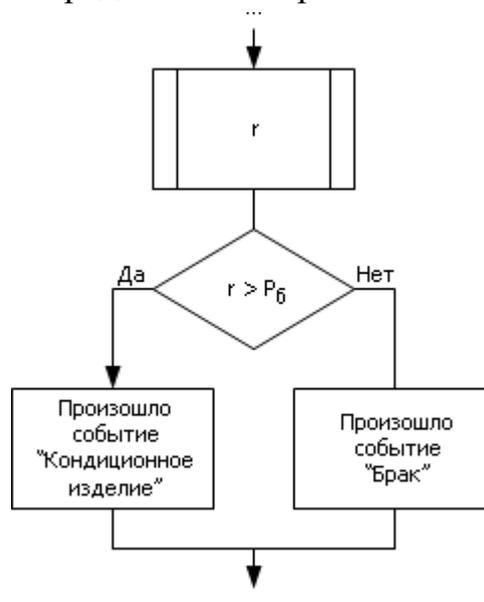


Рис. 23.2. Блок-схема алгоритма имитации случайного события

Не важно, как вы расположите на отрезке $[0; 1]$ интервал P_6 — в начале или в конце, поскольку метод Монте-Карло учитывает только частоту попадания случайных точек в интервал, а она зависит только от величины интервала и не зависит от его месторасположения.

Моделирование полной группы несовместных событий

События называются **несовместными**, если вероятность появления этих событий одновременно равна 0. Отсюда следует, что суммарная вероятность группы несовместных событий равна 1.

Обозначим через a_1, a_2, \dots, a_n события, а через P_1, P_2, \dots, P_n — вероятности появления отдельных событий.

Так как события несовместны, то сумма вероятностей их выпадения равна 1: $P_1 + P_2 + \dots + P_n = 1$.

Используем для имитации выпадения одного из событий генератор случайных чисел, значение которых также всегда находится в диапазоне от 0 до 1. Отложим на единичном интервале $[0; 1]$ отрезки P_1, P_2, \dots, P_n . Понятно, что в сумме отрезки составят точно единичный интервал. Точка, соответствующая выпавшему числу из ГСЧ на этом интервале, укажет на один из отрезков. Соответственно в большие отрезки случайные числа будут попадать чаще (вероятность появления этих событий больше!), в меньшие отрезки — реже (см. рис. 23.3).

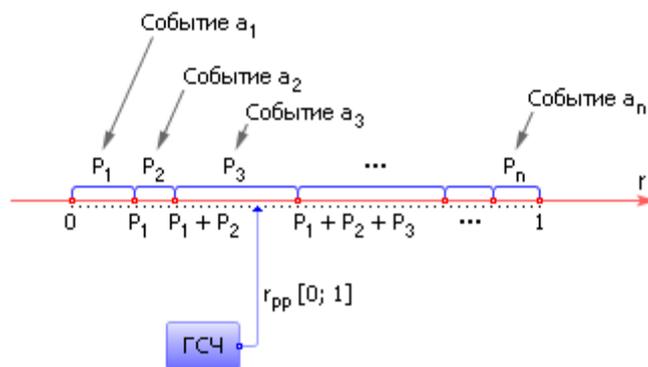


Рис. 23.3. Схема генерации несовместных случайных событий с помощью генератора случайных чисел

На рис. 23.4 показана блок-схема, которая реализует описанный алгоритм. Алгоритм определяет с помощью фильтра, построенного в виде последовательности условных операций (**IF**), в какой из интервалов — от 0 до P_1 , от P_1 до $(P_1 + P_2)$, от $(P_1 + P_2)$ до $(P_1 + P_2 + P_3)$ и так далее — попало число, сгенерированное генератором случайных чисел. Если число попало в какой-то из интервалов (что произойдет всегда и обязательно), то это соответствует выпадению связанного с ним события.

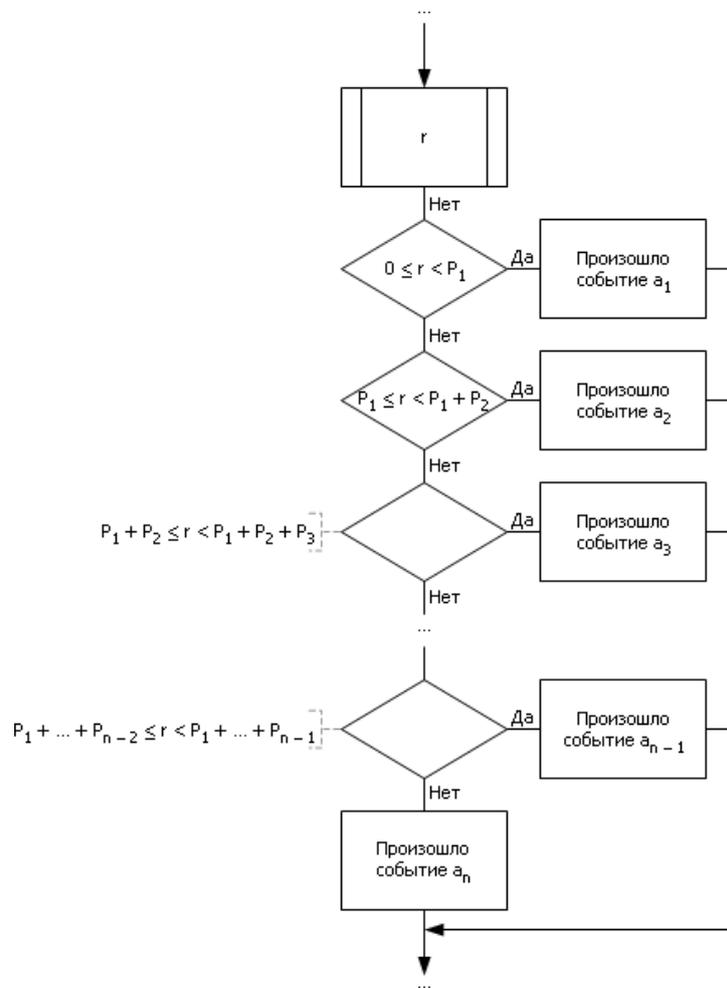


Рис. 23.4. Блок-схема алгоритма имитации случайных несовместных событий

Пример с возможным исходом четырех несовместных случайных событий.

Промоделируем выпадение последовательности событий — будем выбирать из колоды карт наугад карту (определять ее масть). Карты в колоду возвращать не будем.

В колоде 36 карт четырех мастей по 9 карт каждой масти. Интервал от 0 до 1 разделим на равные четыре части: $[0.00—0.25]$, $[0.25—0.50]$, $[0.50—0.75]$, $[0.75—1.00]$. Первая часть будет соответствовать картам масти червей (Ч), вторая — картам масти пик (П), третья — картам масти виной (В), четвертая — бубей (Б).

Взять случайное равномерно распределенное число в интервале от 0 до 1 из таблицы случайных чисел или стандартного ГСЧ. Пусть, например, это будет число 0.597. Данное число попадает в третий интервал, соответствующий масти В. Произошло случайное событие: «Масть выпавшей карты — В».

Теперь в колоде 9 карт масти Ч, 9 карт масти П, 8 карт масти В, 9 карт масти Б, то интервал от 0 до 1 будет разбит на отрезки длиной: $9/35$, $9/35$, $8/35$, $9/35$, то есть $[0.000—0.257]$, $[0.257—0.514]$, $[0.514—0.743]$, $[0.743—1.000]$. Разыграем случайное равномерно распределенное число в интервале от 0 до 1. Например, 0.321. Данное число попадает во второй интервал, соответствующий масти П.

Продолжая процесс, можно получить (в зависимости от конкретных случайных чисел), например, такую последовательность: В—П—В—Ч—Б—П—Ч—... (в качестве иллюстрации см. рис. 23.5).

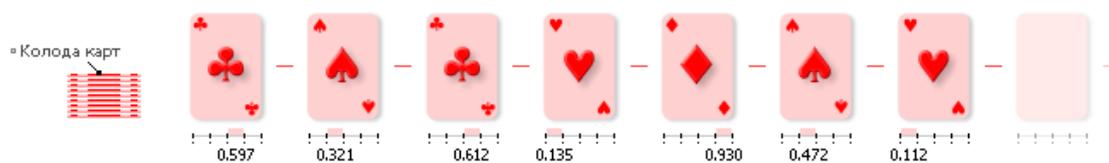


Рис. 23.5. Схема работы генератора случайных чисел на примере выбора карт из колоды

24. МОДЕЛИРОВАНИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ С ЗАДАНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ

Большой информативностью, по сравнению с такими статистическими характеристиками как математическое ожидание, дисперсия, для инженера обладает закон распределения вероятности случайной величины X . Представим, что X принимает случайные значения из некоторого диапазона. X — диаметр вытачиваемой детали. Диаметр может отклоняться от запланированного идеального значения под влиянием различных факторов, которые нельзя учесть, поэтому он является случайной слабо предсказуемой величиной. Но в результате длительного наблюдения за выпускаемыми деталями можно отметить, сколько деталей из 1000 имели диаметр X_1 (обозначим N_{X_1}), сколько деталей имели диаметр X_2 (обозначим N_{X_2}) и так далее. В итоге можно построить гистограмму частоты диаметров, откладывая для X_1 величину $N_{X_1}/1000$, для X_2 величину $N_{X_2}/1000$ и так далее. (Обратите внимание, если быть точным, N_{X_1} — это число деталей, диаметр которых не просто равен X_1 , а находится в диапазоне от $X_1 - \Delta/2$ до $X_1 + \Delta/2$, где $\Delta = X_1 - X_2$). Важно, что сумма всех частот будет равна 1 (суммарная площадь гистограммы неизменна). Если X меняется непрерывно, опытов проведено очень много, то в пределе $N \rightarrow \infty$ гистограмма превращается в график распределения вероятности случайной величины. На рис. 24.1, а показан пример гистограммы дискретного распределения, а на рис. 24.1, б показан вариант непрерывного распределения случайной величины.



Рис. 24.1. Сравнение дискретного и непрерывного законов распределения случайной величины

В примере закон распределения вероятности случайной величины показывает насколько вероятно то или иное значение диаметра выпускаемых деталей. Случайной величиной является диаметр детали [9].

В производстве и технике часто такие законы распределения заданы по условию задачи. Наша задача состоит в том, чтобы научиться имитировать появление конкретных случайных событий согласно вероятностям такого распределения.

Метод ступенчатой аппроксимации

Так как законы распределения вероятности событий могут быть различной формы, а не только равновероятными, то необходимо уметь превращать

равномерный ГСЧ в генератор случайных чисел с заданным произвольным законом распределения. На рис. 21.3 это соответствует двум первым блокам метода статистического моделирования. Для этого непрерывный закон распределения вероятности события дискретизируем, превратим в дискретный.

Обозначим: h_i — высота i -го столбца, $f(x)$ — распределение вероятности (показывает насколько вероятно некоторое событие x). Значение h_i операцией нормировки необходимо перевести в единицы вероятности появления значений x из интервала $x_i < x \leq x_{i+1}$: $P_i = h_i / (h_1 + h_2 + \dots + h_i + \dots + h_n)$.

Операция нормировки обеспечивает сумму вероятностей всех n событий равную 1:

$$\sum_{i=1}^n P_i = 1$$

На рис. 24.2 показаны графически переход от произвольного непрерывного закона распределения к дискретному (рис. 24.2, а), отображение полученных вероятностей на интервал $r_{pp}[0; 1]$ и генерация случайных событий с использованием эталонного равномерно распределенного ГСЧ (рис. 24.2, б).

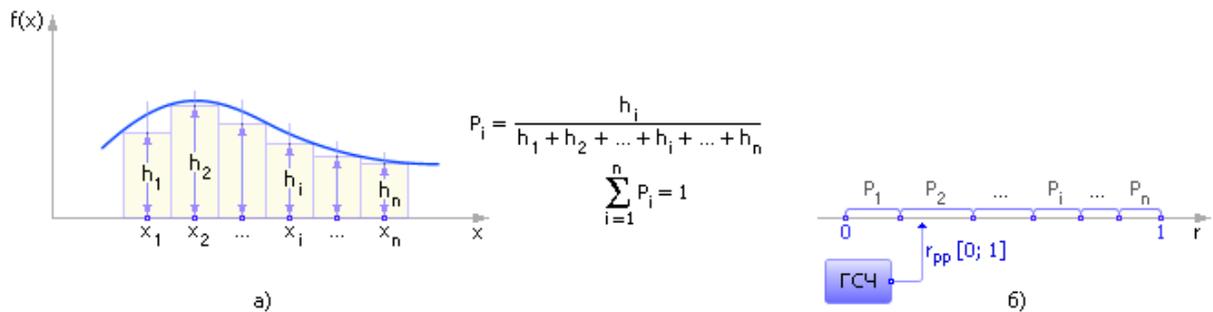


Рис. 24.2. Иллюстрация метода ступенчатой аппроксимации

Внутри интервала $x_i < x \leq x_{i+1}$ значение x теперь не различимо, одинаково. Метод огрубляет изначальную постановку задачи, переходя от непрерывного закона распределения к дискретному. Поэтому следует учитывать количество разбиений n из условий точности представления.

На рис. 24.3 показан фрагмент алгоритма, реализующего описанный метод. Алгоритм генерирует случайное число, равномерно распределенное от 0 до 1. Сравнивая границы отрезков, расположенных на интервале от 0 до 1, представляющих собой вероятности P выпадения тех или иных случайных величин X , определяет в цикле, какое из случайных событий i в результате этого выпадает.

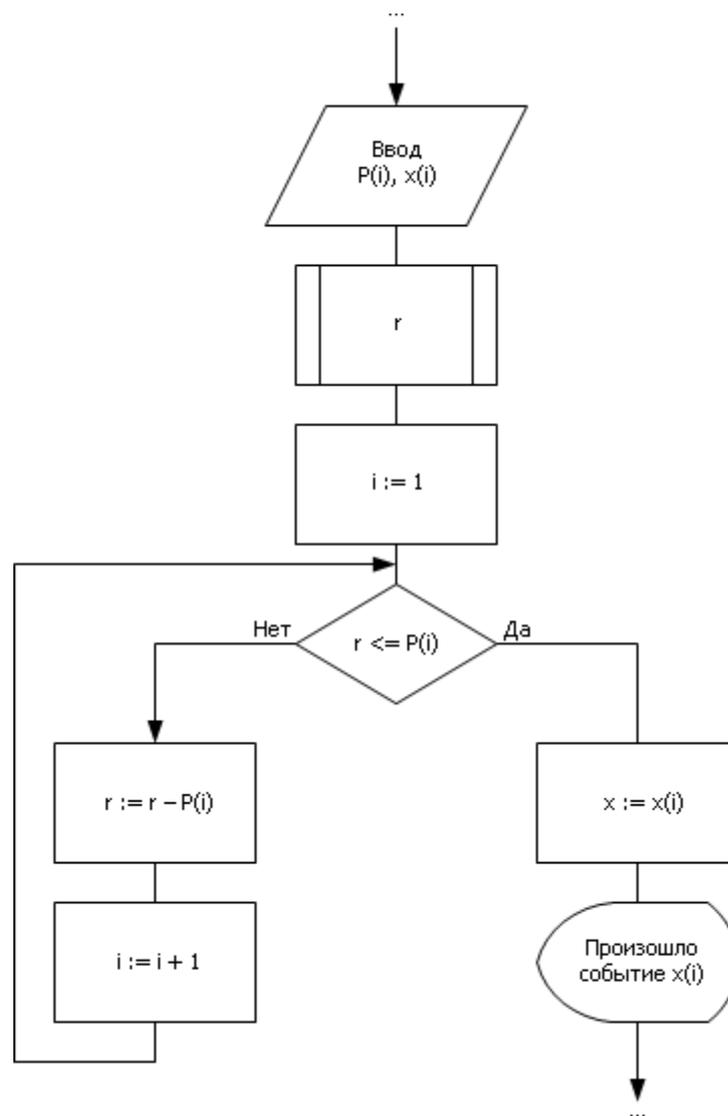


Рис. 24.3. Блок-схема алгоритма, реализующего метод ступенчатой аппроксимации

Внутри интервала $x_i < x \leq x_{i+1}$ значение x теперь не различимо, одинаково. Метод огрубляет изначальную постановку задачи, переходя от непрерывного закона распределения к дискретному. Поэтому следует учитывать количество разбиений n из условий точности представления.

Метод усечения

Метод используется в случае, когда функция задана аналитически (в виде формулы). График функции вписывают в прямоугольник (см. рис. 24.4). На ось Y подают случайное равномерно распределенное число из ГСЧ. На ось X подают случайное равномерно распределенное число из ГСЧ. Если точка в пересечении этих двух координат лежит ниже кривой плотности вероятности, то событие X произошло, иначе нет.

Недостатком метода является то, что те точки, которые оказались выше кривой распределения плотности вероятности, отбрасываются как ненужные, и время, затраченное на их вычисление, оказывается напрасным. Метод применим только для аналитических функций плотности вероятности.

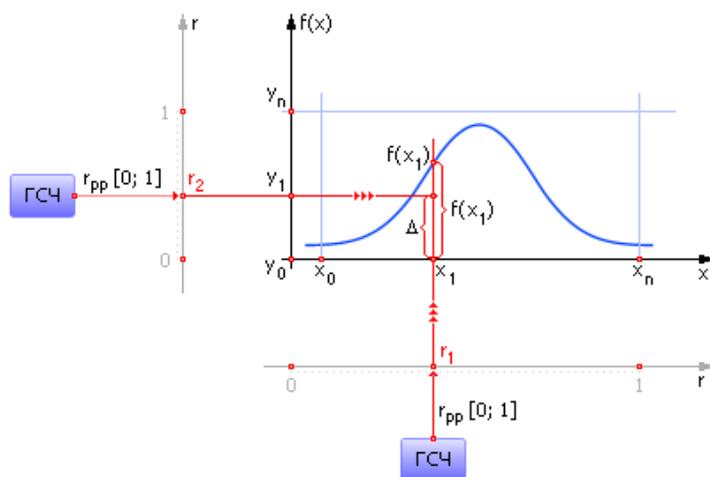


Рис. 24.4. Иллюстрация метода усечения

На рис. 24.5 показан алгоритм, реализующий метод усечения. В цикле генерируется два случайных числа из диапазона от 0 до 1. Числа масштабируются в шкалу X и Y и проверяется попадание точки со сгенерированными координатами под график заданной функции $Y = f(X)$. Если точка находится под графиком функции, то событие X произошло с вероятностью Y , иначе точка отбрасывается.

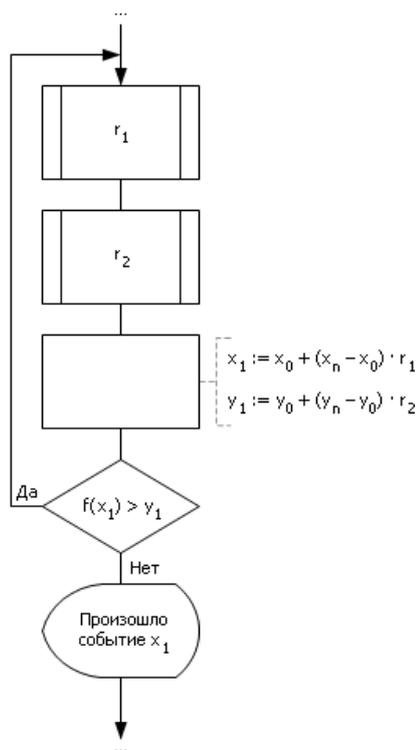


Рис. 24.5. Блок-схема алгоритма, реализующего метод усечения

Метод взятия обратной функции

Допустим, что нам задан интегральный закон распределения вероятности $F(x)$, где $f(x)$ — функция плотности вероятности и

$$F(x) = \int_{-\infty}^x f(x) dx$$

Достаточно разыграть случайное число, равномерно распределенное в интервале от 0 до 1. Поскольку функция F тоже изменяется в данном интервале, то случайное событие x можно определить взятием обратной функции

по графику или аналитически: $x = F^{-1}(r)$. Здесь r — число, генерируемое эталонным ГСЧ в интервале от 0 до 1, x_1 — сгенерированная в итоге случайная величина. Графически суть метода изображена на рис. 24.6.

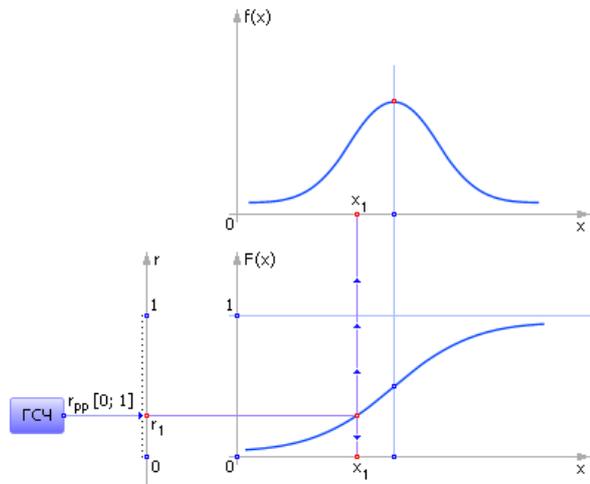


Рис. 24.6. Иллюстрация метода обратной функции для генерации случайных событий x , значения которых распределены непрерывно.

Данным методом особенно удобно пользоваться когда интегральный закон распределения вероятности задан аналитически и возможно аналитическое взятие обратной функции от него.

Пример 1. Примем к рассмотрению экспоненциальный закон распределения вероятности случайных событий $f(x) = \lambda \cdot e^{-\lambda x}$. Тогда интегральный закон распределения плотности вероятности имеет вид: $F(x) = 1 - e^{-\lambda x}$.

Так как r и F в данном методе предполагаются аналогичными и расположены в одном интервале, то, заменяя F на случайное число r , имеем:

$$r = 1 - e^{-\lambda x}.$$

Выражая искомую величину x из этого выражения (то есть, обращая функцию $\exp()$), получаем: $x = -1/\lambda \cdot \ln(1 - r)$.

Так как в статическом смысле $(1 - r)$ и r — это одно и то же, то $x = -1/\lambda \cdot \ln(r)$.

На рис. 24.7 показан фрагмент алгоритма, реализующего метод обратной функции для экспоненциального закона.

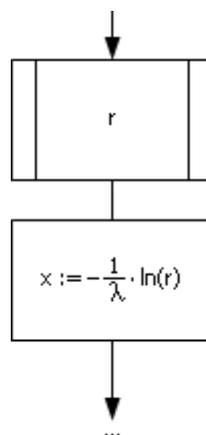


Рис. 24.7. Фрагмент блок-схемы алгоритма, реализующей метод обратной функции для экспоненциального закона

25. МОДЕЛИРОВАНИЕ НОРМАЛЬНО РАСПРЕДЕЛЕННЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Нормальный закон распределения встречается в природе весьма часто, поэтому для него разработаны отдельные эффективные методы моделирования. Формула распределения вероятности значений случайной величины x по нормальному закону имеет вид:

$$y = \frac{1}{\sigma_x \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x - m_x)^2}{2\sigma_x^2}}$$

Нормальное распределение имеет два параметра: математическое ожидание m_x и среднеквадратичное отклонение σ_x величины x от этого математического ожидания.

x — случайная величина;

$y(x)$ — вероятность принятия случайной величиной значения x ;

m_x — математическое ожидание;

σ_x — среднее квадратичное отклонение.

Нормализованным нормальным распределением называется такое нормальное распределение, у которого $m_x = 0$ и $\sigma_x = 1$. Из нормализованного распределения можно получить любое другое нормальное распределение с заданными m_x и σ_x по формуле: $z = m_x + x \cdot \sigma_x$.

Рассматривая последнюю формулу, вспомните формулы компьютерной графики: операция масштабирования выражается в математической модели через умножение (это соответствует изменению разброса величины, растягиванию геометрического образа), операция смещения выражается через сложение (это соответствует изменению значения наиболее вероятной величины, смещению геометрического образа) [9].

Функция нормального распределения имеет вид колокола. На рис. 25.1 показано нормализованное нормальное распределение.

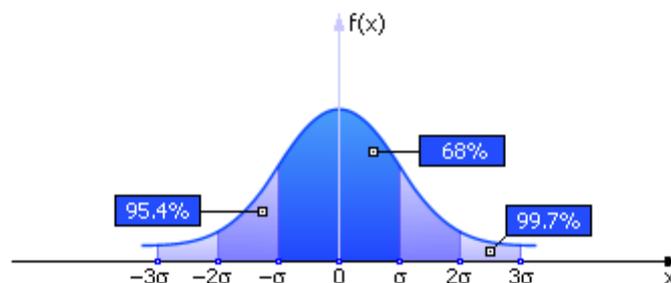


Рис. 25.1. Графический вид нормального закона распределения случайной величины x с параметрами $m_x = 0$ и $\sigma_x = 1$ (распределение нормализовано)

График на рис. 25.1 показывает, что в области $-\sigma < x < \sigma$ на графике сосредоточено 68% площади распределения, в области $-2\sigma < x < 2\sigma$ на графике сосредоточено 95.4% площади распределения, в области $-3\sigma < x < 3\sigma$ на графике сосредоточено 99.7% площади распределения («правило трех сигм»). Вспомните, пожалуйста, [рис. 2.7](#) из [лекции 02](#).

Пример. По нормальному распределению распределен рост людей, находящихся одновременно в большой аудитории. А именно: достаточно мало людей очень большого роста, и столь же мала вероятность встретить людей очень малого роста. В основном, легче встретить людей среднего роста — и вероятность этого велика.

Например, средний рост людей составляет, в основном, 170 см, то есть $m_x = 170$. Известно также, что $\sigma_x = 20$. На рис. 25.1 показано, что доля людей с ростом от 150 до 190 ($170 - 20 < 170 < 170 + 20$) составляет в обществе 68%. Доля людей от 130 см до 210 см ($170 - 2 \cdot 20 < 170 < 170 + 2 \cdot 20$) составляет в обществе 95.4%. Доля людей от 110 см до 230 ($170 - 3 \cdot 20 < 170 < 170 + 3 \cdot 20$) составляет в обществе 99.7%. Например, вероятность того, что человек окажется ростом меньше 110 см или больше 230 см составляет всего 3 человека на 1000.

Свойства нормального распределения

Изменение параметра нормального распределения m_x приводит к сдвигу кривой по оси x (см. рис. 25.2).

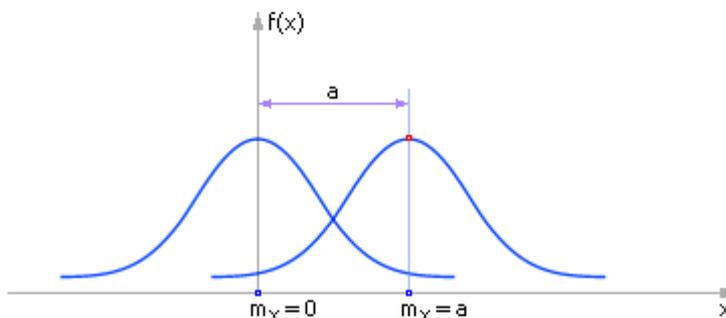


Рис. 25.2. Влияние параметра «математическое ожидание» на вид закона нормального распределения случайной величины x

Изменение параметра нормального распределения σ_x приводит к масштабированию формы по оси x (напоминаем, в любом случае всегда площадь под кривой плотности вероятности неизменна и равна 1).

Чем более не случаен процесс, тем меньше его среднеквадратичное отклонение, тем уже и выше колокол на графике. Изменение параметра нормального распределения σ_x приводит к масштабированию формы (см. рис. 25.3) по оси x (напоминаем, в любом случае всегда площадь под кривой плотности вероятности неизменна и равна 1).

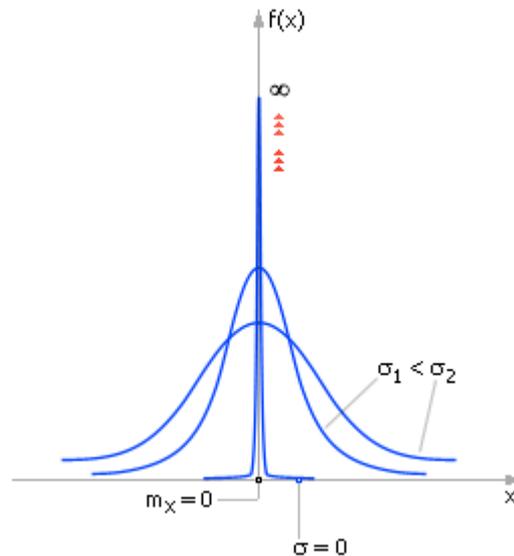


Рис. 25.3. Влияние параметра «среднеквадратичное отклонение» на вид закона нормального распределения случайной величины x

Чем более не случаен процесс, тем меньше его среднеквадратичное отклонение, тем уже и выше колокол на графике. Действительно, разброс случайности относительно математического ожидания становится все более минимальным. В пределе детерминированный процесс имеет вид, показанный на рис. 25.4.

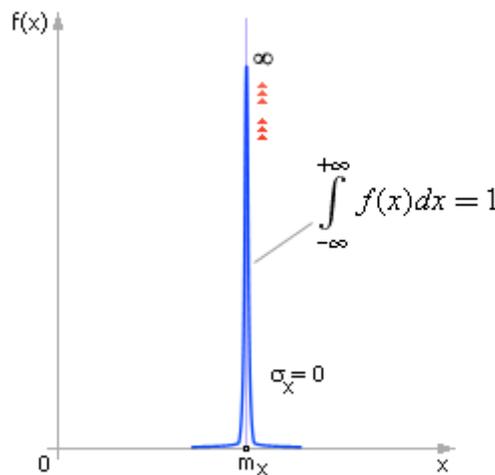


Рис. 25.4. Вид закона нормального распределения вероятности при переходе его к детерминированному случаю в пределе ($\sigma_x = 0$). Случайное событие становится детерминированным: $x = m_x \pm 0$ (разброса нет)

Изучать детерминированные процессы проще. Чем больше величина σ_x , тем менее закономерно поведение изучаемого объекта, так как возможны любые значения характеризующих его параметров, разброс величин относительно средней ожидаемой увеличивается. Прогнозирование и управление поведением объекта в этом случае затрудняется.

Вид интегральной кривой плотности распределения случайной величины, распределенной по нормальному закону. Вид ее приведен на рис. 25.5. F — интегральная функция Лапласа. Смысл интегральной функции — вероят-

ность того, что случайная величина примет значения из диапазона от $-\infty$ до x . Например, запись $F(170) = 0.5$ для нашего примера означает: вероятность того, что случайно выбранный из аудитории человек будет ростом не выше 170 см, составляет 0.5 (то есть каждый второй).

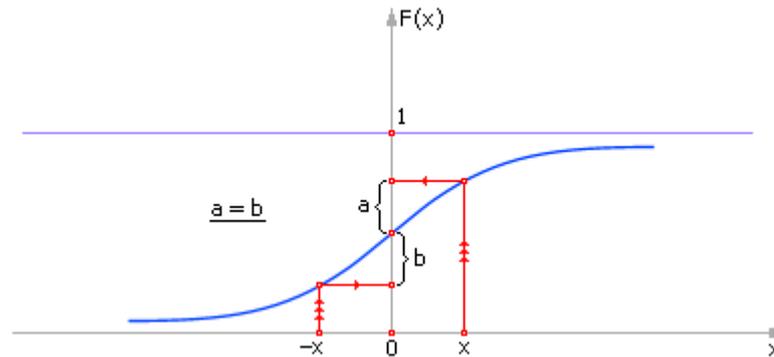


Рис. 25.5. Вид интегральной функции Лапласа $F(x)$

Данная функция задана интегралом от плотности вероятности нормального распределения:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

Интеграл не берется в общем виде, поэтому функция Лапласа задана в виде таблицы для $m_x = 0$ и $\sigma_x = 1$. Поскольку функция Лапласа симметрична относительно точки $(x = 0, y = 0.5)$ (как и функция самого нормального распределения), $F(-x) = 1 - F(x)$, то в таблице содержится только одна из ее симметричных частей.

Если задается интервал интегрирования функции Лапласа $[a; b]$, то:

$$P(a < x < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-m_x)^2}{2\sigma^2}} dx = \int_{\frac{a-m_x}{\sigma}}^{\frac{b-m_x}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt = F(b) - F(a)$$

Вероятность попадания X в интервал, симметричный относительно m_x :
 $P(|x - m_x| < L) = P((m_x - L) < x < (m_x + L)) = F((b - m_x)/\sigma) - F((a - m_x)/\sigma) =$
 $= F(L/\sigma) - F(-L/\sigma) = 2F(L/\sigma) - 1$

Например, для правила «трех сигм»: $P(|x - m_x| < 3\sigma) = 2 \cdot F(3) - 1 = 2 \cdot 0.9987 - 1 = 0.9973$ (как ранее мы и указывали). Число $F(3) = 0.9987$ взято из таблицы Лапласа.

Пример. Найти вероятность изготовления детали с ошибкой в ее размерах не более 15 мм, если известно, что изготовление детали с ошибкой распределено по нормальному закону $m = 0$ и $\sigma = 10$ мм.

$P(|x| < 15) = P(-15 < x < 15) = F((15 - 0)/10) - F((-15 - 0)/10) = F(1.5) - F(-1.5) = F(1.5) - (1 - F(1.5)) = 2 \cdot F(1.5) - 1 = 2 \cdot 0.9332 - 1 = 0.8664$. То есть 8664 детали из 10000 будут иметь ошибку в размерах не более 15 мм.

Важным свойством закона является то, что нормальное распределение является пределом для различного вида распределений, вытекающее из цен-

тральной предельной теоремы: «Для большого числа N случайных величин X с любым законом распределения их сумма есть случайное число с нормальным законом распределения».

$$\lim_{N \rightarrow \infty} P\left(\alpha < \sum_{i=1}^N x_i - N \cdot a \cdot \frac{1}{\sqrt{N} \cdot \sigma} < \beta\right) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{t^2}{2}} dt = F(\beta) - F(\alpha)$$

где a — математическое ожидание в законе распределения случайной величины X ; σ — среднеквадратичное отклонение в законе распределения случайной величины X ; N — количество случайных чисел.

Вспомним опыт «Доска Гальтона» из физики (см. рис. 25.6).

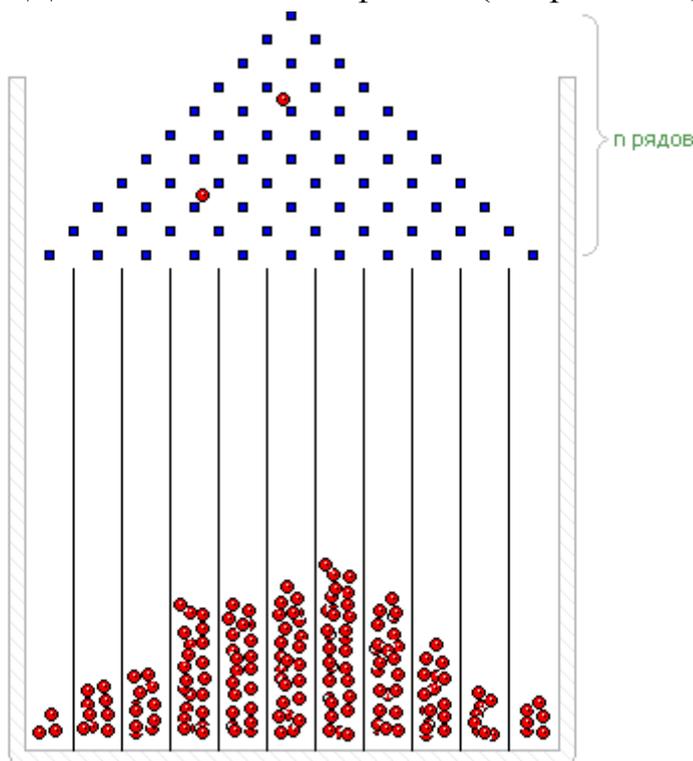


Рис. 25.6. Доска Гальтона. Шарики, падающие сверху в сосуд случайно распределяются в нем в согласии с нормальным законом распределения

Доска разделена на секции; в верхней части доски находятся особым образом расположенные стержни, ударяясь о которые, множество падающих сверху под влиянием силы тяжести шариков, испытывая соударения также и между собой, меняют свою траекторию полета. В результате в различные секции попадает различное количество шариков. Если дожидаться окончания опыта, то в каждой секции будет определенное количество шариков, конечно, каждый раз разное, так как процесс их столкновения случайный. Но интересно то, что распределение шариков по секциям будет образовывать нормальный закон распределения. Доска не меняется, шарики падают одни и те же, и тем не менее, во-первых, форма распределения слегка колеблется (случайность), и разные шарики попадают в разные секции, во-вторых, на макроуровне, где проявляется организация шариков как совокупности, всегда получается нормальный закон распределения (закономерность).

Почему малые отклонения в поведении элементов системы (шариков) ведут к большим разбросам в их координатах? Почему случайности не компенсируются? Насколько они не компенсируются?

Обозначим через D расстояние между стержнями доски Гальтона, r — диаметр стержня (см. рис. 25.7, а, б). Очевидно, что отклонение β шарика от отвесной траектории при малом угле α (изменение направления вследствие удара о стержень) можно посчитать как $\text{tg}(\alpha_1) = \beta_1/r$ или $\alpha_1 \approx \beta_1/r$ (при малых углах α_1). Далее шарик летит под углом α_1 к траектории свободного падения на расстояние D , пока не встретит новый стержень. За время перелета по прямой от стержня к стержню шарик отклонится от отвесной вертикальной линии на величину $\beta_2 = \alpha_1 \cdot D$. Далее снова, испытав соударение со стержнями, шарик отклонится на угол $\alpha_2 \approx \beta_2/r$ и $\beta_3 \approx \alpha_2 \cdot D = \beta_1 \cdot (D/r)^2$ и так далее.



Рис. 25.7. Взаимодействие шариков со стержнями доски (детализация)

Если шарик испытает n соударений, то он отклонится от отвесной траектории на расстояние $\beta_n \approx \beta_1 \cdot (D/r)^{n-1}$ (конечно, если ему «повезет» и он будет отклоняться все время в одну сторону).

Посмотрим, сколько это означает в цифрах, например, $D = 20r$, пусть стержней, от которых отклонился шарик в полете (соударений), $n = 10$, $\beta_1 = 1$ нм (расстояние меньше атома!!), то тогда легко вычислить, что $\beta_{10} = 10.24$ км. То есть, видно, что сверхмалые отклонения (1 нм и число соударений $n = 10$), фактически случайности, приводят к макроэффектам. На самом деле такой разброс наступает гораздо быстрее, так как шарики сталкиваются еще и между собой. Кстати, удаление какого-то шарика на 10 км — это тоже невероятное событие, так как вероятность, что шарик «повезет» 20 раз подряд, составляет крайне маловероятное число $P = 0.520 = 10^{-6}$.

Рассмотрим один эксперимент «Распределение Максвелла». В сосуде находится 40 синих и 40 красных шаров, каждые в своей половине сосуда. Половинки сосуда разделены перегородкой с отверстием, через которое могут проникать шары из одной части сосуда в другую (см. рис. 25.8). Количество шаров, как синих, так и красных, в каждой из частей сосуда подсчитываются.

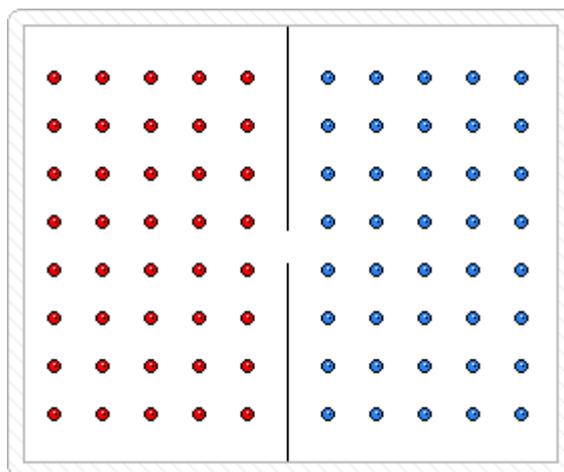


Рис. 25.8. Схема эксперимента, демонстрирующего диффузию (появление хаоса) в сложных системах

Шары имитируют броуновское движение молекул, соударяясь друг с другом упруго, обмениваясь при ударе энергиями, импульсами, направлениями скорости и также взаимодействуя со стенками сосуда. При смоделированном в опыте упругом ударе усредняется средняя кинетическая энергия встречающихся частиц. Несмотря на то, и в этом суть диффузии, что шары находились каждый цвет строго в своей половине, со временем примерно половина (50%) красных шаров окажется в первой половине сосуда, а половина (50%) — во второй его части, то же касается и синих шаров (50% : 50%). Употребленное здесь слово «примерно» означает, что время от времени совокупность шаров будет делиться не ровно на половины, а в пропорции $(50\% \pm h : 50\% \pm h)$, так как случайное количество шаров будет переходить в произвольные моменты времени из одной части сосуда в другую, и наоборот. Обнаруженная в наших опытах относительная флуктуация средних энергий достаточно велика, порядка 20%. Несмотря на то, что шары каждого цвета находились строго в своей половине (в этом и состоит суть диффузии), со временем примерно половина (50%) красных шаров окажется в первой половине сосуда, а половина (50%) — во второй его части, то же касается и синих шаров (50%:50%) (см. рис. 25.9).

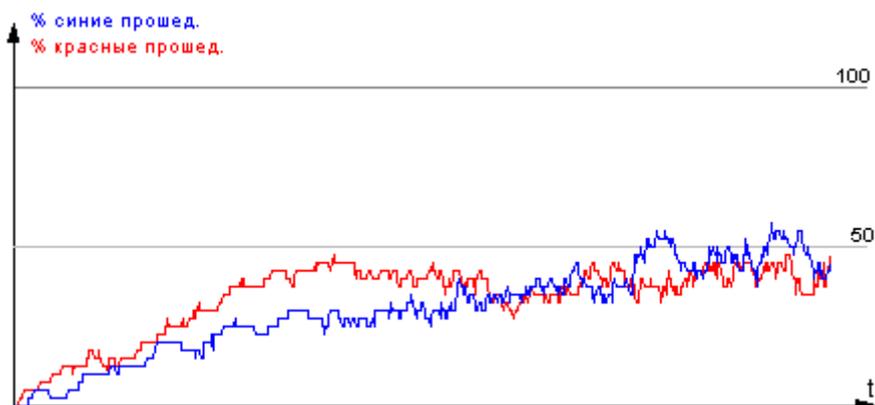


Рис. 25.9. Типичный вид поведения характеристик в системах с диффузией. На рисунке показано изменение количества синих и красных молекул в одной из половинок сосуда со временем

Особенность 1. Можете продолжать наблюдение сколь угодно долго, но при этом никогда первоначального распределения этих шаров достигнуто не

будет. При любом начальном порядке шаров система будет переходить к одному и тому же хаотическому состоянию. Почему? Объяснение этого надо искать в том, что время необратимо. Операция усреднения также необратима. Если я вам сообщу, что сумма двух чисел равна 8, то вряд ли вы сможете сказать, какие я выбрал для этого слагаемые. И множество вариантов слагаемых ведут к одной и той же сумме.

Хаотическая система не помнит свое прошлое. А именно: состояние такой системы по прошествии времени не зависит от того, как и из какого состояния система пришла в хаотическое. Из любого начального состояния наша система довольно быстро переходит в это хаотическое состояние, и находится в нем далее сколько угодно долго. То есть хаотическое состояние — устойчивое состояние. Интересно, что в этом устойчивом состоянии можно отыскать такой ряд параметров, который будет неизменен со временем, такие параметры описывают систему в целом и называются **макропараметрами**. Например, в нашем примере устойчивым макропараметром является распределение количества молекул n по модулю скорости v , которое демонстрирует закон Максвелла $n(v) = r \cdot v^2 \cdot e^{-c \cdot v^2/T}$ (r, c — коэффициенты, T — температура совокупности молекул).

Особенность 2. Если разделить сосуд на большее число частей, то шары распределятся между ними примерно в равных долях.

Особенность 3. Слово «примерно» пополам, говорит нам о том, что в ответе 50% : 50% всегда есть разброс, дисперсия, флуктуация. Этот разброс уменьшается, если увеличить число молекул в сосуде.

Особенность 4. Если запустить модель снова и попробовать шары сначала упорядочить, например, установить их ровно рядами в одной половине сосуда, или вверху, или в одном из углов, то очень быстро шары «примерно» равномерно распределятся по сосуду, и наступит хаос, то есть порядок исчезнет и, что интересно, никогда снова не наступит. Откуда возникает хаос, как система переходит от порядка к хаосу? Причина хаоса, как мы уже ранее показали в опыте «Доска Гальтона», появляется в момент **небольших отклонений** шаров (пусть, например, 10^{-9} м) от идеальной теоретической траектории, но это небольшое отклонение на пролете расстояния D приводит к существенному отклонению траектории от идеала и столкновению шаров, что **резко меняет** траектории соседей. Мгновенно появляется лавина соударений, и хаос наступает очень быстро.

Особенность 5. Ранее, считалось, что системы с малым числом элементов не способны проявлять случайность, хаотическое поведение. Преобладало рассуждение физиков о газе из миллиардов молекул. Проведите этот же эксперимент с пятью шарами, потом с тремя и, наконец, одним шаром. Распределение Максвелла в результате эксперимента проявляется достаточно упрямо, правда, для этого требуется ждать в несколько раз дольше. (Исключение составит опыт с одним шаром) **Опыт показывает, что усреднение по ансамблю молекул и по времени проводит к одинаковому результату.** Так что хаос может возникнуть и на небольших множествах.

Опыт «Доска Гальтона» и экспериментальное снятие распределения Максвелла приводят нас эмпирически к тому, что нормальное распределение широко распространено в окружающем нас мире. Суммирование случайных величин, сумма действий, совместное наложение эффектов на микроуровне часто приводит, усреднившись, к появлению нормального распределения на макроуровне.

Табличный метод генерации нормально распределенных чисел

Нормальное число можно взять из справочника в таблице функции Лапласа и получить случайное число по методу взятия обратной функции (см. лекцию 24): $x = F^{-1}(r)$, где F — интегральная функция Лапласа.

Технически это означает, что надо разыграть случайное равномерно распределенное число r из интервала $[0; 1]$ стандартным ГСЧ (см. таблицу абсолютно случайных проверенных чисел), найти равное ему число в таблице значений функции Лапласа в столбце F и по строке определить случайную величину x , соответствующую этому числу.

Недостатком метода является необходимость хранения в памяти компьютера всей таблицы чисел функции Лапласа.

Метод генерации нормально распределенных чисел, использующий центральную предельную теорему

Общая идея метода следующая: требуется сложить случайные числа с любым законом распределения, нормализовать их и перевести в нужный диапазон нормального распределения.

Допустим, что нам надо в целях имитации получить ряд случайных чисел x , распределенных по нормальному закону с заданными математическим ожиданием m_x и среднеквадратичным отклонением σ_x .

1. Сложим n случайных чисел, используя стандартный ГСЧ:

$$V = \sum_{i=1}^n r_i$$

Согласно ЦПТ числа V образуют ряд значений, распределенный по нормальному закону. Эти числа тем лучше описывают нормальный закон, чем больше параметр n . На практике n берут равными 6 или 12. Заметим, что закон распределения чисел V имеет математическое ожидание $m_V = n/2$, $\sigma_V = \sqrt{n/12}$. Поэтому он является смещенным относительно заданного произвольного.

2. С помощью формулы $z = (V - m_V)/\sigma_V$ нормализуем этот ряд. Получим нормализованный закон нормального распределения чисел Z . То есть $m_z = 0$, $\sigma_z = 1$.
3. Формулой (сдвиг на m_x и масштабирование на σ_x) преобразуем ряд Z в ряд x : $x = z \cdot \sigma_x + m_x$.

Пример. Смоделировать поток заготовок для обработки их на станке. Известно, что длина заготовки колеблется случайным образом. Средняя длина заготовки составляет 35 см, а среднеквадратичное отклонение реальной длины от средней составляет 10 см. То есть по условиям задачи $m_x = 35$, $\sigma_x = 10$. Тогда значение случайной величины будет рассчитываться по формуле: $V = r_1 + r_2 + r_3 + r_4 + r_5 + r_6$, где r — случайные числа из ГСЧ_{pp} $[0; 1]$,

$n = 6$.

$$X = \sigma_x \cdot (\text{sqrt}(12/n) \cdot (V - n/2)) + m_x = 10 \cdot \text{sqrt}(2) \cdot (V - 3) + 35$$

или

$$X = 10 \cdot \text{sqrt}(2) \cdot ((r_1 + r_2 + r_3 + r_4 + r_5 + r_6) - 3) + 35.$$

Метод Мюллера

Простым методом получения нормальных чисел является метод Мюллера, использующий формулы: $Z = \sqrt{-2 \cdot \text{Ln}(r_1)} \cdot \cos(2\pi \cdot r_2)$, где r_1 и r_2 — случайные числа из ГСЧ_{pp} [0; 1].

Можно также воспользоваться аналогичной формулой $Z = \sqrt{-2 \cdot \text{Ln}(r_1)} \cdot \sin(2\pi \cdot r_2)$, где r_1 и r_2 — случайные числа из ГСЧ_{pp} [0; 1].

Пример. Материал поступает в цех один раз в сутки по 10 штук сразу. Расход материала из цеха случайный по нормальному закону с математическим ожиданием $m = 10$ и среднеквадратичным отклонением $\sigma = 3.5$. Вычислить вероятность дефицита на складе при запасе материала в начальный момент времени 20 штук.

При реализации в среде моделирования Stratum решение задачи будет выглядеть следующим образом.

```
z := sqrt(-2 * ln(rnd)) * cos(2 * PI * rnd)
x := sigma * z + m
запас := запас + 10
запас := (запас - x) * ed(запас - x)
дефицит := дефицит + not(ed(запас - x))
N := N + 1
P := дефицит/N
stop(N > k)
```

z — нормальное нормализованное случайное число;

x — нормальное число, ежедневный расход материалов;

запас — состояние склада: начало дня, моделирование прихода;

запас — состояние склада: конец дня, моделирование расхода;

дефицит — счетчик дней, в течение которых наблюдался дефицит;

N — количество дней;

P — вероятность дефицита;

k — моделирование в течение k дней.

26. МОДЕЛИРОВАНИЕ СИСТЕМЫ СЛУЧАЙНЫХ ВЕЛИЧИН

На практике встречаются системы случайных величин, то есть такие две (и более) различные случайные величины X , Y (и другие), которые зависят друг от друга. Например, если произошло событие X и приняло какое-то случайное значение, то событие Y происходит хотя и случайно, но с учетом того, что X уже приняло какое-то значение.

Если в качестве X выпало большое число, то Y должно выпасть тоже достаточно большое число (если корреляция положительна). Весьма вероятно, что если человек имеет большой вес, то он, скорее всего, будет и большого роста. Хотя это не обязательно, это не закономерность, а корреляция случайных величин. Так как бывают, хотя и редко, люди с большим весом, но небольшого роста или с маленьким весом и высокие. И все таки основная масса тучных людей — высоки, а низких людей — имеют малый вес.

По определению, если случайные величины независимы, то

$$f(\mathbf{x}) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n).$$

x_i — случайная независимая величина;

$f(x_i)$ — плотность вероятности выпадения случайной независимой величины x_i ;

$f(\mathbf{x})$ — плотность вероятности выпадения вектора \mathbf{x} случайных независимых величин x_1, x_2, \dots, x_n .

Если случайные величины зависимы, то

$$f(\mathbf{x}) = f(x_1) \cdot f(x_2 | x_1) \cdot f(x_3 | x_2, x_1) \cdot \dots \cdot f(x_n | x_{n-1}, x_{n-2}, \dots, x_2, x_1).$$

$x_j | x_{j-1}, \dots, x_1$ — случайные зависимые величины: выпадение x_j при условии, что выпали x_{j-1}, \dots, x_1 ;

$f(x_j | x_{j-1}, \dots, x_1)$ — плотность условной вероятности появления x_j , если выпали x_{j-1}, \dots, x_1 ;

$f(\mathbf{x})$ — вероятность выпадения вектора \mathbf{x} случайных зависимых величин.

К примеру, имеется два зависимых события — X и Y , распределенных по нормальному закону. X имеет математическое ожидание m_x и среднеквадратическое отклонение σ_x . Y имеет математическое ожидание m_y и среднеквадратическое отклонение σ_y . Коэффициент корреляции — q — показывает, насколько тесно связаны события X и Y . Если коэффициент корреляции равен единице, то зависимость событий X и Y взаимно однозначная: одному значению X соответствует одно значение Y (см. рис. 26.1).

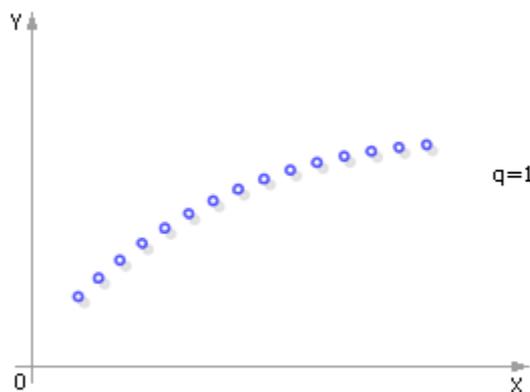


Рис. 26.1. Вид зависимости двух случайных величин при положительном коэффициенте корреляции ($q=1$)

При q близких к единице возникает картина, показанная на рис. 26.2, то есть одному значению X могут соответствовать уже несколько значений Y (точнее, одно из нескольких значений Y , определяемое случайным образом); в этом случае события X и Y менее коррелированы, менее зависимы друг от друга.

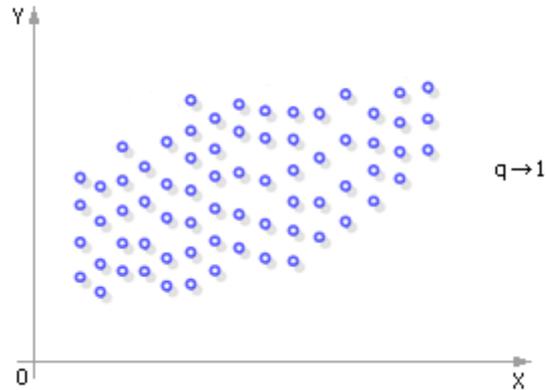


Рис. 26.2. Вид зависимости двух случайных величин при положительном коэффициенте корреляции ($0 < q < 1$)

Когда коэффициент корреляции стремится к нулю, возникает ситуация, при которой любому значению X может соответствовать любое значение Y , то есть события X и Y не зависят или почти не зависят друг от друга, не коррелируют друг с другом (см. рис. 26.3).

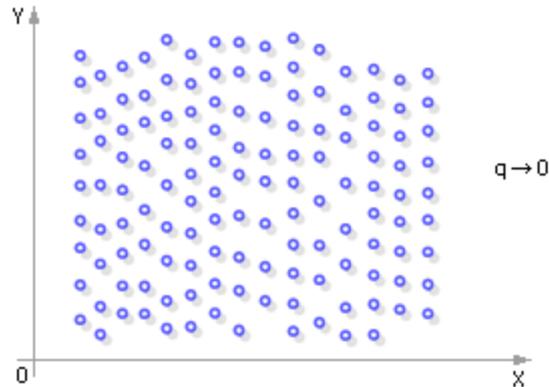


Рис. 26.3. Вид зависимости двух случайных величин при коэффициенте корреляции близком к нулю ($q \rightarrow 0$)

На всех графиках корреляция была принята положительной величиной. Если $q < 0$, то графики будут выглядеть так, как показано на рис. 26.4.

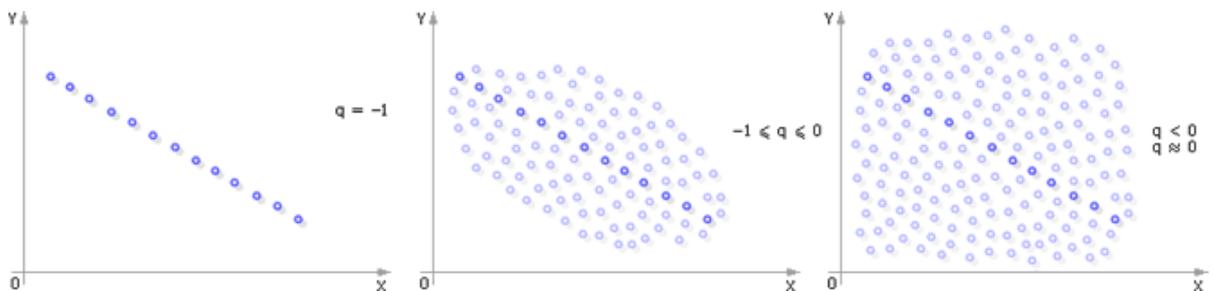


Рис. 26.4. Вид зависимости двух случайных величин при отрицательном коэффициенте корреляции
 а) $\rho = -1$; б) $-1 < \rho < 0$; в) $\rho \rightarrow 0$

На самом деле случайные события (X и Y) не могут принимать с равной вероятностью любые значения, как это имеет место на рис. 26.2. К примеру, в группе студентов не может быть людей сверхмалого или сверхбольшого роста; в основном, люди обладают неким средним ростом и разбросом вокруг этого среднего роста. Поэтому на одних участках оси X количество событий расположено гуще, на других — реже. (Плотность случайных событий, количество точек на графиках больше вблизи величин m_x). То же самое верно и для Y . И тогда рис. 26.2 можно изобразить более точно, так, как показано на рис. 26.5.

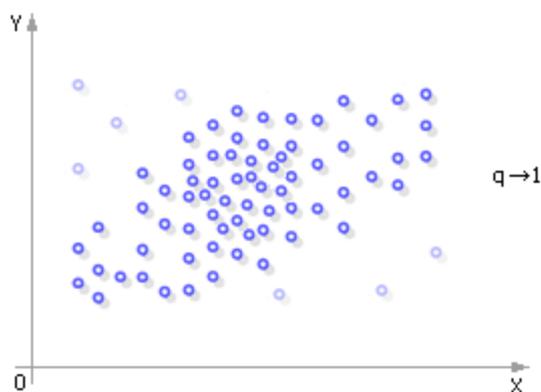


Рис. 26.5. Иллюстрация системы случайных зависимых величин

Возьмем нормальное распределение, как самое распространенное. Математическое ожидание указывает на самые вероятные события, здесь число событий больше и график событий гуще. Положительная корреляция указывает (см. рис. 26.2), что большие случайные величины X вызывают к генерации большие Y . Отрицательная корреляция указывает (см. рис. 26.4), что большие случайные величины X стимулируют к генерации меньшие случайные величины Y . Нулевая и близкая к нулю корреляция показывает (см. рис. 26.3), что величина случайной величины X никак не связана с определенным значением случайной величины Y . Легко понять сказанное, если представить себе сначала распределения $f(X)$ и $f(Y)$ отдельно, а потом связать их в систему (см. рис. 26.6, рис. 26.7 и рис. 26.8).

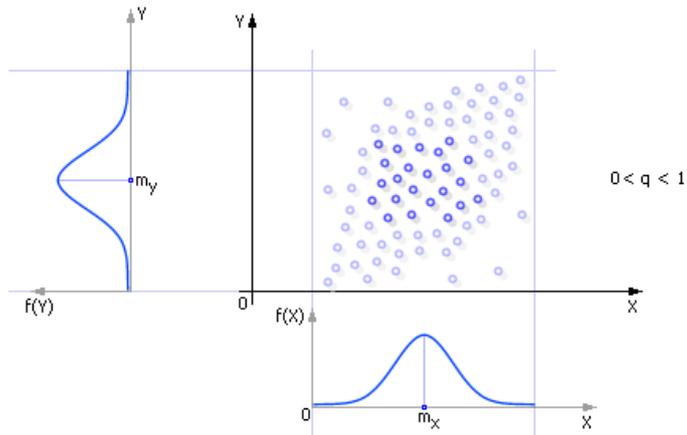


Рис. 26.6. Генерация системы случайных величин при положительном коэффициенте корреляции

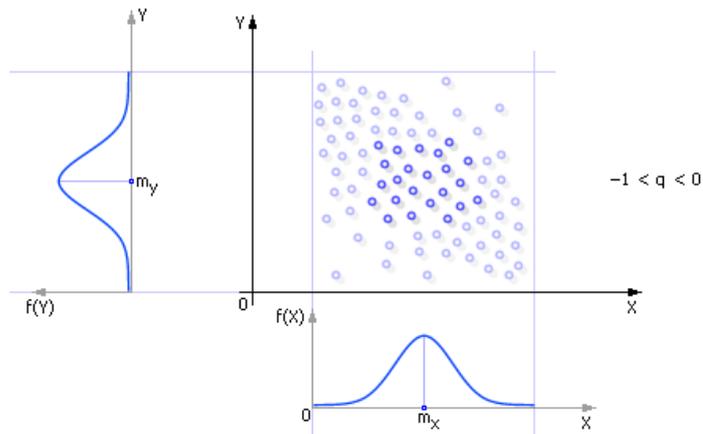


Рис. 26.7. Генерация системы случайных величин при отрицательном коэффициенте корреляции

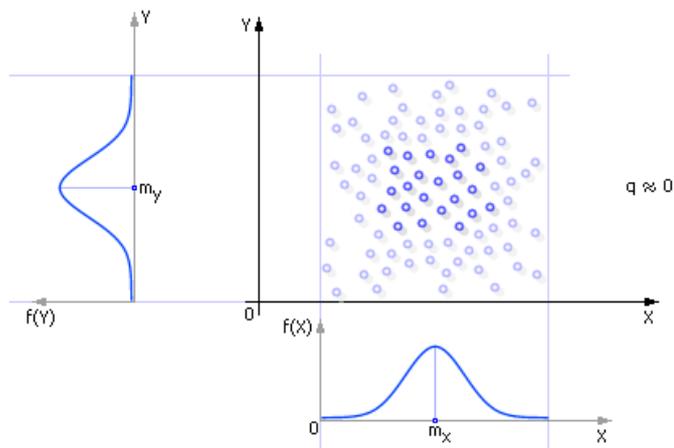


Рис. 26.8.

Пример реализации алгоритма моделирования двух зависимых случайных событий X и Y . Условие: допустим, что X и Y распределены по нормальному закону с соответствующими значениями m_x , σ_x и m_y , σ_y . Задан ко-

коэффициент корреляции двух случайных событий q , то есть случайные величины X и Y зависимы друг от друга, Y не совсем случайно.

Тогда алгоритм реализации модели будет следующим.

1. Разыгрывается шесть случайных равномерно распределенных на интервале $[0; 1]$ чисел $b_1, b_2, b_3, b_4, b_5, b_6$; находится их сумма S : $S = b_1 + b_2 + b_3 + b_4 + b_5 + b_6$. Находится нормально распределенное случайное число x по следующей формуле: $x = \sqrt{2} \cdot \sigma_x \cdot (S - 3) + m_x$, см. лекцию 25.
2. По формуле $m_{y/x} = m_y + q \cdot \sigma_y / \sigma_x \cdot (x - m_x)$ находится математическое ожидание $m_{y/x}$ (знак y/x означает, что y будет принимать случайные значения с учетом условия, что x уже принял какие-то определенные значения).
3. По формуле $\sigma_{y/x} = \sigma_y \cdot \sqrt{1 - q^2}$ находится среднеквадратическое отклонение $\sigma_{y/x}$ (знак y/x означает, что y будет принимать случайные значения с учетом условия, что x уже принял какие-то определенные значения).

Разыгрывается шесть случайных равномерно распределенных на интервале $[0; 1]$ чисел $r_1, r_2, r_3, r_4, r_5, r_6$; находится их сумма k : $k = r_1 + r_2 + r_3 + r_4 + r_5 + r_6$. Находится нормально распределенное случайное число y по следующей формуле: $y = \sqrt{2} \cdot \sigma_{y/x} \cdot (k - 3) + m_{y/x}$.

27. РАСПРЕДЕЛЕНИЕ ПУАССОНА

Наиболее общим случаем различного рода вероятностных распределений является биномиальное распределение. Воспользуемся его универсальностью для определения наиболее часто встречающихся на практике частных видов распределений.

Биномиальное распределение

Имеется некое событие А. Вероятность появления события А равна p , вероятность не появления события А равна $1 - p$, иногда ее обозначают как q . Пусть n — число испытаний, m — частота появления события А в этих n испытаниях.

Известно, что суммарная вероятность всех возможных комбинаций исходов равна единице, то есть:

$$1 = p^n + n \cdot p^{n-1} \cdot (1-p) + C_n^{n-2} \cdot p^{n-2} \cdot (1-p)^2 + \dots + C_n^m \cdot p^m \cdot (1-p)^{n-m} + \dots + (1-p)^n.$$

p^n — вероятность того, что в n испытаниях событие А произойдет n раз;

$n \cdot p^{n-1} \cdot (1-p)$ — вероятность того, что в n испытаниях событие А произойдет $(n-1)$ раз и не произойдет 1 раз;

$C_n^{n-2} \cdot p^{n-2} \cdot (1-p)^2$ — вероятность того, что в n испытаниях событие А произойдет $(n-2)$ раза и не произойдет 2 раза;

$P_m = C_n^m \cdot p^m \cdot (1-p)^{n-m}$ — вероятность того, что в n испытаниях событие А произойдет m раз и не произойдет $(n-m)$ раз;

$(1-p)^n$ — вероятность того, что в n испытаниях событие А не произойдет ни разу;

$$C_n^m = \frac{n!}{m! \cdot (n-m)!} \text{ — число сочетаний из } n \text{ по } m.$$

Математическое ожидание M биномиального распределения равно:

$$M = n \cdot p,$$

где n — число испытаний, p — вероятность появления события А.

Среднеквадратичное отклонение σ :

$$\sigma = \sqrt{n \cdot p \cdot (1-p)}.$$

Пример 1. Вычислить вероятность того, что событие, имеющее вероятность $p = 0.5$, в $n = 10$ испытаниях произойдет $m = 1$ раз. Имеем: $C_{10}^1 = 10$, и

далее: $P_1 = 10 \cdot 0.5^1 \cdot (1 - 0.5)^{10-1} = 10 \cdot 0.5^{10} = 0.0098$. Вероятность наступления этого события достаточно мала. Объясняется это, во-первых, тем, что абсолютно не ясно, произойдет ли событие или нет, поскольку вероятность равна 0.5 и шансы здесь «50 на 50»; а во-вторых, требуется исчислить то, что событие произойдет именно один раз (не больше и не меньше) из десяти.

Пример 2. Вычислить вероятность того, что событие, имеющее вероятность $p = 0.5$, в $n = 10$ испытаниях произойдет $m = 2$ раза. Имеем: $C_{10}^2 = 45$, и далее: $P_2 = 45 \cdot 0.5^2 \cdot (1 - 0.5)^{10-2} = 45 \cdot 0.5^{10} = 0.044$. Вероятность наступления этого события стала больше!

Пример 3. Увеличим вероятность наступления самого события. Сделаем его более вероятным. Вычислить вероятность того, что событие, имеющее вероятность $p = 0.8$, в $n = 10$ испытаниях произойдет $m = 1$ раз. Имеем: $C_{10}^1 = 10$, и далее: $P_1 = 10 \cdot 0.8^1 \cdot (1 - 0.8)^{10-1} = 10 \cdot 0.8^1 \cdot 0.2^9 = 0.000004$. Вероятность стала меньше, чем в первом примере! Ответ, на первый взгляд, кажется странным, но поскольку событие имеет достаточно большую вероятность, вряд ли оно произойдет только один раз. Более вероятно, что оно произойдет большее, чем один, количество раз. Действительно, подсчитывая $P_0, P_1, P_2, P_3, \dots, P_{10}$ (вероятность того, что событие в $n = 10$ испытаниях произойдет 0, 1, 2, 3, ..., 10 раз), мы увидим:

$$C_{10}^0 = 1, \quad C_{10}^1 = 10, \quad C_{10}^2 = 45, \quad C_{10}^3 = 120, \quad C_{10}^4 = 210, \quad C_{10}^5 = 252,$$

$$C_{10}^6 = 210, \quad C_{10}^7 = 120, \quad C_{10}^8 = 45, \quad C_{10}^9 = 10, \quad C_{10}^{10} = 1;$$

$$P_0 = 1 \cdot 0.8^0 \cdot (1 - 0.8)^{10-0} = 1 \cdot 1 \cdot 0.2^{10} = 0.0000\dots;$$

$$P_1 = 10 \cdot 0.8^1 \cdot (1 - 0.8)^{10-1} = 10 \cdot 0.8^1 \cdot 0.2^9 = 0.0000\dots;$$

$$P_2 = 45 \cdot 0.8^2 \cdot (1 - 0.8)^{10-2} = 45 \cdot 0.8^2 \cdot 0.2^8 = 0.0000\dots;$$

$$P_3 = 120 \cdot 0.8^3 \cdot (1 - 0.8)^{10-3} = 120 \cdot 0.8^3 \cdot 0.2^7 = 0.0008\dots;$$

$$P_4 = 210 \cdot 0.8^4 \cdot (1 - 0.8)^{10-4} = 210 \cdot 0.8^4 \cdot 0.2^6 = 0.0055\dots;$$

$$P_5 = 252 \cdot 0.8^5 \cdot (1 - 0.8)^{10-5} = 252 \cdot 0.8^5 \cdot 0.2^5 = 0.0264\dots;$$

$$P_6 = 210 \cdot 0.8^6 \cdot (1 - 0.8)^{10-6} = 210 \cdot 0.8^6 \cdot 0.2^4 = 0.0881\dots;$$

$$P_7 = 120 \cdot 0.8^7 \cdot (1 - 0.8)^{10-7} = 120 \cdot 0.8^7 \cdot 0.2^3 = 0.2013\dots;$$

$$P_8 = 45 \cdot 0.8^8 \cdot (1 - 0.8)^{10-8} = 45 \cdot 0.8^8 \cdot 0.2^2 = 0.3020\dots \text{ (самая большая вероятность!)};$$

$$P_9 = 10 \cdot 0.8^9 \cdot (1 - 0.8)^{10-9} = 10 \cdot 0.8^9 \cdot 0.2^1 = 0.2684\dots;$$

$$P_{10} = 1 \cdot 0.8^{10} \cdot (1 - 0.8)^{10-10} = 1 \cdot 0.8^{10} \cdot 0.2^0 = 0.1074\dots$$

$$\text{Разумеется, } P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 + P_9 + P_{10} = 1.$$

Нормальное распределение

Если изобразить величины $P_0, P_1, P_2, P_3, \dots, P_{10}$, которые мы подсчитали в примере 3, на графике, то окажется, что их распределение имеет вид, близкий к нормальному закону распределения (см. рис. 27.1) (см. лекцию 25. Моделирование нормально распределенных случайных величин).

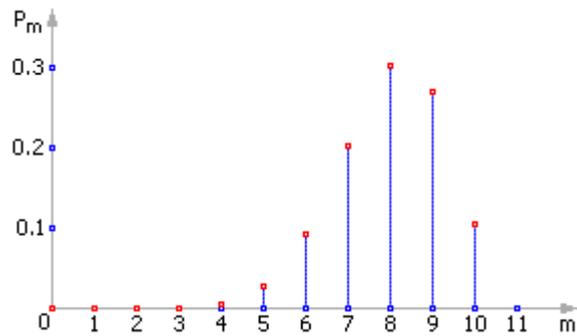


Рис. 27.1. Вид биномиального распределения вероятностей для различных m при $p = 0.8$, $n = 10$

Биномиальный закон переходит в нормальный, если вероятности появления и не появления события A примерно одинаковы, то есть, условно можно записать: $p \approx (1 - p)$. Для примера возьмем $n = 10$ и $p = 0.5$ (то есть $p = 1 - p = 0.5$).

Содержательно к такой задаче мы придем, если, например, захотим теоретически посчитать, сколько будет мальчиков и сколько девочек из 10 родившихся в роддоме в один день детей. Точнее, считать будем не мальчиков и девочек, а вероятность, что родятся только мальчики, что родится 1 мальчик и 9 девочек, что родится 2 мальчика и 8 девочек и так далее. Примем для простоты, что вероятность рождения мальчика и девочки одинакова и равна 0.5.

Распределение будет симметричное, так как вероятность рождения 3 мальчиков и 7 девочек равна вероятности рождения 7 мальчиков и 3 девочек. Наибольшая вероятность рождения будет у 5 мальчиков и 5 девочек. Эта вероятность равна 0.25, кстати, не такая уж она и большая по абсолютной величине. Далее, вероятность того, что родится сразу 10 или 9 мальчиков намного меньше, чем вероятность того, что родится 5 ± 1 мальчик из 10 детей. Как раз биномиальное распределение нам поможет сделать этот расчет [9].

$$C_{10}^0 = 1, C_{10}^1 = 10, C_{10}^2 = 45, C_{10}^3 = 120, C_{10}^4 = 210, C_{10}^5 = 252,$$

$$C_{10}^6 = 210, C_{10}^7 = 120, C_{10}^8 = 45, C_{10}^9 = 10, C_{10}^{10} = 1;$$

$$P_0 = 1 \cdot 0.5^0 \cdot (1 - 0.5)^{10-0} = 1 \cdot 1 \cdot 0.5^{10} = 0.000977\dots;$$

$$P_1 = 10 \cdot 0.5^1 \cdot (1 - 0.5)^{10-1} = 10 \cdot 0.5^{10} = 0.009766\dots;$$

$$P_2 = 45 \cdot 0.5^2 \cdot (1 - 0.5)^{10-2} = 45 \cdot 0.5^{10} = 0.043945\dots;$$

$$P_3 = 120 \cdot 0.5^3 \cdot (1 - 0.5)^{10-3} = 120 \cdot 0.5^{10} = 0.117188\dots;$$

$$P_4 = 210 \cdot 0.5^4 \cdot (1 - 0.5)^{10-4} = 210 \cdot 0.5^{10} = 0.205078\dots;$$

$$P_5 = 252 \cdot 0.5^5 \cdot (1 - 0.5)^{10-5} = 252 \cdot 0.5^{10} = 0.246094\dots;$$

$$P_6 = 210 \cdot 0.5^6 \cdot (1 - 0.5)^{10-6} = 210 \cdot 0.5^{10} = 0.205078\dots;$$

$$P_7 = 120 \cdot 0.5^7 \cdot (1 - 0.5)^{10-7} = 120 \cdot 0.5^{10} = 0.117188\dots;$$

$$P_8 = 45 \cdot 0.5^8 \cdot (1 - 0.5)^{10-8} = 45 \cdot 0.5^{10} = 0.043945\dots;$$

$$P_9 = 10 \cdot 0.5^9 \cdot (1 - 0.5)^{10-9} = 10 \cdot 0.5^{10} = 0.009766\dots;$$

$$P_{10} = 1 \cdot 0.5^{10} \cdot (1 - 0.5)^{10-10} = 1 \cdot 0.5^{10} = 0.000977\dots$$

Разумеется, $P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 + P_9 + P_{10} = 1$.

Отразим на графике величины $P_0, P_1, P_2, P_3, \dots, P_{10}$ (см. рис. 27.2).

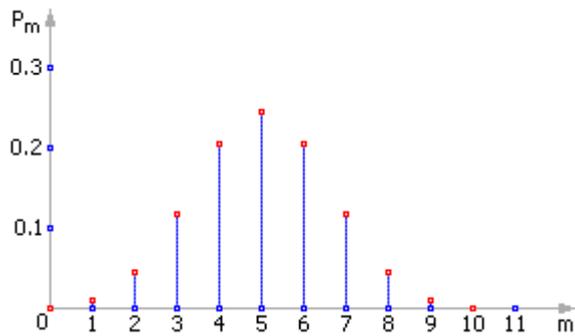


Рис. 27.2. График биномиального распределения при параметрах $p = 0.5$ и $n = 10$, приближающих его к нормальному закону

При условиях $m \approx n/2$ и $p \approx 1 - p$ или $p \approx 0.5$ вместо биномиального распределения можно использовать нормальное. При больших значениях n график сдвигается вправо и становится все более пологим, так как математическое ожидание и дисперсия возрастают с увеличением n : $M = n \cdot p$, $D = n \cdot p \cdot (1 - p)$.

Биномиальный закон стремится к нормальному и при увеличении n , что вполне естественно, согласно центральной предельной теореме (см. лекцию 34. Фиксация и обработка статистических результатов).

Рассмотрим, как изменится биномиальный закон в случае, когда $p \neq q$, то есть $p \rightarrow 0$. В этом случае применить гипотезу о нормальности распределения нельзя, и биномиальное распределение переходит в распределение Пуассона.

Распределение Пуассона

Распределение Пуассона — это частный случай биномиального распределения (при $n \gg 0$ и при $p \rightarrow 0$ (редкие события)).

Из математики известна формула, позволяющая примерно подсчитать значение любого члена биномиального распределения:

$$P_m = C_n^m \cdot p^m \cdot (1 - p)^{n-m} \cong \frac{a^m \cdot e^{-a}}{m!}$$

где $a = n \cdot p$ — параметр Пуассона (математическое ожидание), а дисперсия равна математическому ожиданию. Приведем математические выкладки, поясняющие этот переход. Биномиальный закон распределения

$$P_m = C_n^m \cdot p^m \cdot (1 - p)^{n-m}$$

может быть написан, если положить $p = a/n$, в виде

$$P_m = C_n^m \cdot \left(\frac{a}{n}\right)^m \cdot \left(1 - \frac{a}{n}\right)^{n-m}$$

или

$$P_m = \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{m-1}{n}\right) \cdot \left(1 - \frac{a}{n}\right)^{-m} \cdot \left(1 - \frac{a}{n}\right)^n \cdot \frac{a^m}{m!}$$

Так как p очень мало, то следует принимать во внимание только числа m , малые по сравнению с n . Произведение

$$\left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{m-1}{n}\right)$$

весьма близко к единице. Это же относится к величине

$$\left(1 - \frac{a}{n}\right)^{-m}$$

Величина

$$\left(1 - \frac{\alpha}{n}\right)^n$$

очень близка к $e^{-\alpha}$. Отсюда получаем формулу:

$$P_m = C_n^m \cdot p^m \cdot (1-p)^{n-m} \cong \frac{\alpha^m \cdot e^{-\alpha}}{m!}$$

Пример. В ящике находится $n = 100$ деталей, как качественных, так и бракованных. Вероятность достать бракованное изделие составляет $p = 0.01$. Допустим, что мы вынимаем изделие, определяем, бракованное оно или нет, и кладем его обратно. Поступая таким образом, получилось, что из 100 изделий, которые мы перебрали, два оказались бракованными. Какова вероятность этого?

По биномиальному распределению получаем:

$$\begin{aligned} P_2 &= C_{100}^2 \cdot p^2 \cdot (1-p)^{100-2} = C_{100}^2 \cdot 0.01^2 \cdot (1-0.01)^{100-2} = \\ &= \frac{100!}{2! \cdot 98!} \cdot 0.01^2 \cdot 0.99^{98} \cong 0.185 \end{aligned}$$

По распределению Пуассона получаем:

$$P_2 = \frac{\alpha^m \cdot e^{-\alpha}}{m!} = \frac{(n \cdot p)^m \cdot e^{-n \cdot p}}{m!} = \frac{(100 \cdot 0.01)^2 \cdot e^{-100 \cdot 0.01}}{2!} \cong 0.184$$

Величины получились близкими, поэтому в случае редких событий вполне допустимо применять закон Пуассона, тем более что он требует меньших вычислительных затрат.

Покажем графически вид закона Пуассона. Возьмем для примера параметры $p = 0.05$, $n = 10$. Тогда:

$$C_{10}^0 = 1, C_{10}^1 = 10, C_{10}^2 = 45, C_{10}^3 = 120, C_{10}^4 = 210, C_{10}^5 = 252,$$

$$C_{10}^6 = 210, C_{10}^7 = 120, C_{10}^8 = 45, C_{10}^9 = 10, C_{10}^{10} = 1;$$

$$P_0 = 1 \cdot 0.05^0 \cdot (1-0.05)^{10-0} = 1 \cdot 1 \cdot 0.95^{10} = 0.5987\dots;$$

$$P_1 = 10 \cdot 0.05^1 \cdot (1-0.05)^{10-1} = 10 \cdot 0.05^1 \cdot 0.95^9 = 0.3151\dots;$$

$$P_2 = 45 \cdot 0.05^2 \cdot (1-0.05)^{10-2} = 45 \cdot 0.05^2 \cdot 0.95^8 = 0.0746\dots;$$

$$P_3 = 120 \cdot 0.05^3 \cdot (1-0.05)^{10-3} = 120 \cdot 0.05^3 \cdot 0.95^7 = 0.0105\dots;$$

$$P_4 = 210 \cdot 0.05^4 \cdot (1-0.05)^{10-4} = 210 \cdot 0.05^4 \cdot 0.95^6 = 0.00096\dots;$$

$$P_5 = 252 \cdot 0.05^5 \cdot (1-0.05)^{10-5} = 252 \cdot 0.05^5 \cdot 0.95^5 = 0.00006\dots;$$

$$P_6 = 210 \cdot 0.05^6 \cdot (1-0.05)^{10-6} = 210 \cdot 0.05^6 \cdot 0.95^4 = 0.0000\dots;$$

$$P_7 = 120 \cdot 0.05^7 \cdot (1-0.05)^{10-7} = 120 \cdot 0.05^7 \cdot 0.95^3 = 0.0000\dots;$$

$$P_8 = 45 \cdot 0.05^8 \cdot (1-0.05)^{10-8} = 45 \cdot 0.05^8 \cdot 0.95^2 = 0.0000\dots;$$

$$P_9 = 10 \cdot 0.05^9 \cdot (1-0.05)^{10-9} = 10 \cdot 0.05^9 \cdot 0.95^1 = 0.0000\dots;$$

$$P_{10} = 1 \cdot 0.05^{10} \cdot (1-0.05)^{10-10} = 1 \cdot 0.05^{10} \cdot 0.95^0 = 0.0000\dots$$

Разумеется, $P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 + P_9 + P_{10} = 1$.

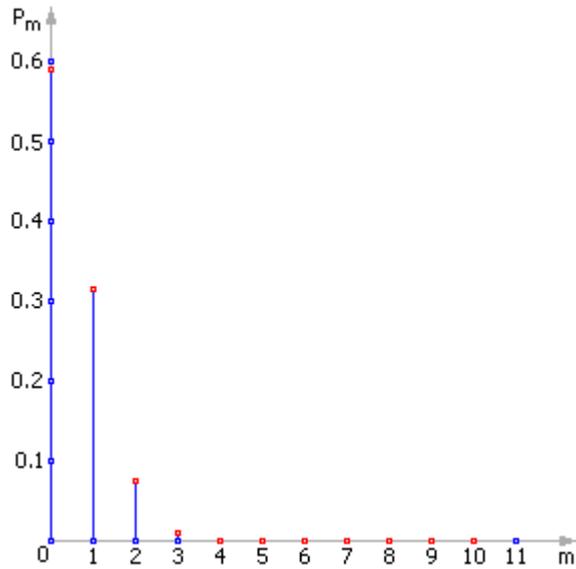


Рис. 27.3. График распределения Пуассона при $p = 0.05$ и $n = 10$

При $n \rightarrow \infty$ распределение Пуассона переходит в нормальный закон, согласно центральной предельной теореме (см. лекцию 34. Фиксация и обработка статистических результатов).

28. ПОТОК СЛУЧАЙНЫХ СОБЫТИЙ

Когда событий много и они следуют друг за другом, то они образуют **поток**. Заметим, что события при этом должны быть однородными, то есть похожими чем-то друг на друга. Появление водителей на АЗС, желающих заправить свой автомобиль. Однородные события образуют некий ряд. При этом считается, что статистическая характеристика этого явления (интенсивность потока событий) задана. Интенсивность потока событий указывает, сколько *в среднем* происходит таких событий за единицу времени. Но когда именно произойдет каждое конкретное событие надо определить методами моделирования. Когда мы сгенерируем, например, за 200 часов 1000 событий, их количество будет равно примерно величине средней интенсивности появления событий $1000/200 = 5$ событий в час, что является статистической величиной, характеризующей этот поток в целом [9].

Интенсивность потока в некотором смысле является математическим ожиданием количества событий в единицу времени. Но реально может так оказаться, что в один час появится 4 события, в другой — 6, хотя в среднем получается 5 событий в час, поэтому одной величины для характеристики потока недостаточно. Второй величиной, характеризующей насколько велик разброс событий относительно математического ожидания, является, как и ранее, дисперсия. Именно эта величина определяет случайность появления события, слабую предсказуемость момента его появления. Про эту величину мы расскажем в следующей лекции.

Поток событий — это последовательность однородных событий, наступающих одно за другим в случайные промежутки времени. На оси времени эти события выглядят как показано на рис. 28.1.

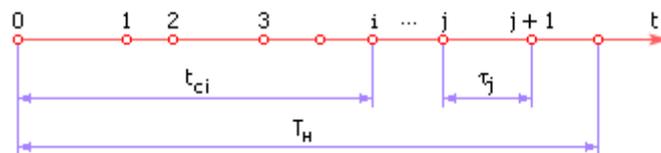


Рис. 28.1. Поток случайных событий

τ_j — интервал между событиями (случайная величина);

t_{ci} — момент совершения i -го события (отсчитывается от $t = 0$);

T_n — время наблюдения.

Примером потока событий могут служить последовательность моментов касания взлетной полосы самолетами, прилетающими в аэропорт.

Интенсивность потока λ — это среднее число событий в единицу времени. Интенсивность потока можно рассчитать экспериментально по формуле: $\lambda = N/T_n$, где N — число событий, произошедших за время наблюдения T_n .

Если интервал между событиями τ_j равен константе или определен какой-либо формулой в виде: $t_j = f(t_{j-1})$, то поток называется **детерминированным**. Иначе поток называется **случайным**.

Случайные потоки бывают:

- **ординарные:** вероятность одновременного появления двух и более событий равна нулю;
- **стационарные:** частота появления событий $\lambda(t) = \text{const}(t)$;
- **без последствия:** вероятность появления случайного события не зависит от момента совершения предыдущих событий.

Пуассоновский поток

За эталон потока в моделировании принято брать пуассоновский поток.

Пуассоновский поток — это ординарный поток без последствия.

Как ранее было указано, вероятность того, что за интервал времени $(t_0, t_0 + \tau)$ произойдет m событий, определяется из закона Пуассона:

$$P_m = \frac{a^m \cdot e^{-a}}{m!}$$

$$a = \int_{t_0}^{t_0 + \tau} \lambda(t) \cdot dt$$

где a — параметр Пуассона.

Если $\lambda(t) = \text{const}(t)$, то это **стационарный поток Пуассона** (простейший). В этом случае $a = \lambda \cdot t$. Если $\lambda = \text{var}(t)$, то это **нестационарный поток Пуассона**.

Для простейшего потока вероятность появления m событий за время τ равна:

$$P_m = \frac{(\lambda \cdot \tau)^m \cdot e^{-\lambda \tau}}{m!}$$

Вероятность неоявления (то есть ни одного, $m = 0$) события за время τ равна:

$$P_0 = \frac{(\lambda \cdot \tau)^0 \cdot e^{-\lambda \tau}}{0!} = e^{-\lambda \tau}$$

Рис. 28.2 иллюстрирует зависимость P_0 от времени. Очевидно, что чем больше время наблюдения, тем вероятность неоявления ни одного события меньше. Чем более значение λ , тем круче идет график, то есть быстрее убывает вероятность. Это соответствует тому, что если интенсивность появления событий велика, то вероятность неоявления события быстро уменьшается со временем наблюдения.

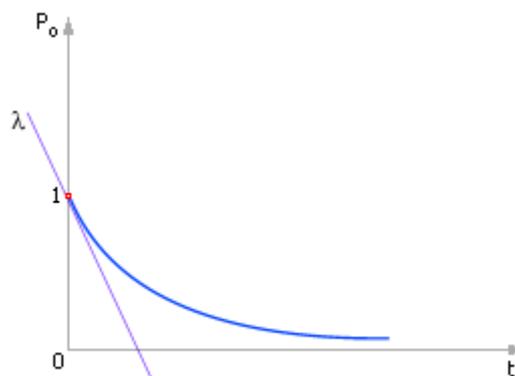


Рис. 28.2. График вероятности неоявления ни одного события во времени

Вероятность появления хотя бы одного события ($P_{\text{ХБ1С}}$) вычисляется так:

$$P_{\text{ХБ1С}} = 1 - P_0 = 1 - e^{-\lambda \tau} \quad (*)$$

так как $P_{\text{ХБИС}} + P_0 = 1$ (либо появится хотя бы одно событие, либо не появится ни одного, — другого не дано).

Из графика на рис. 28.3 видно, что вероятность появления хотя бы одного события стремится со временем к единице, то есть при соответствующем длительном наблюдении события такое обязательно рано или поздно произойдет. Чем дольше мы наблюдаем за событием (чем более t), тем больше вероятность того, что событие произойдет — график функции монотонно возрастает.

Чем больше интенсивность появления события (чем больше λ), тем быстрее наступает это событие, и тем быстрее функция стремится к единице. На графике параметр λ представлен крутизной линии (наклон касательной).

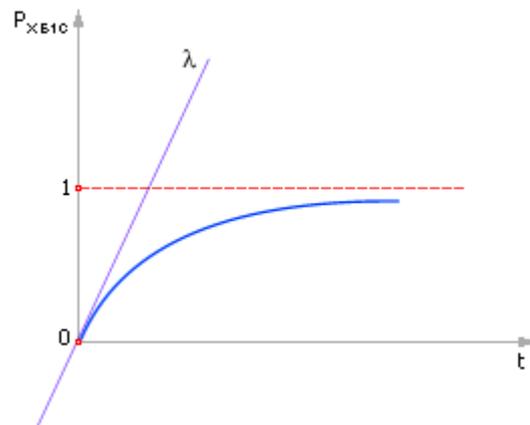


Рис. 28.3. График вероятности появления хотя бы одного события со временем

Если увеличивать λ , то при наблюдении за событием в течение одного и того же времени τ , вероятность наступления события возрастает (см. рис. 28.4). График исходит из 0, так как если время наблюдения бесконечно мало, то вероятность того, что событие произойдет за это время, ничтожна. Если время наблюдения бесконечно велико, то событие обязательно произойдет хотя бы один раз, значит, график стремится к значению вероятности равной 1.

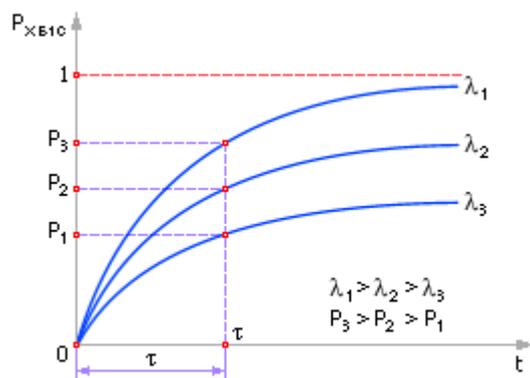


Рис. 28.4. Влияние величины интенсивности потока на вероятность появления события в течение заданного интервала времени τ

Изучая закон, можно определить, что: $m_x = 1/\lambda$, $\sigma = 1/\lambda$, то есть для простейшего потока $m_x = \sigma$. Равенство математического ожидания среднеквадратичному отклонению означает, что данный поток — поток без последовательности. Дисперсия (точнее, среднеквадратичное отклонение) такого потока велика. Физически это означает, что время появления события (расстояние между событиями) плохо предсказуемо, случайно, находится в интервале $m_x - \sigma < \tau_j < m_x + \sigma$. Хотя ясно, что в среднем оно примерно равно: $\tau_j = m_x = T_n/N$. Событие может появиться в любой момент времени, но в пределах разброса этого момента τ_j относительно m_x на $[-\sigma; +\sigma]$ (величину последовательности). На рис. 28.5 показаны возможные положения события 2 относительно оси времени при заданном σ . В данном случае говорят, что первое событие не влияет на второе, второе на третье и так далее, то есть последовательности отсутствует.

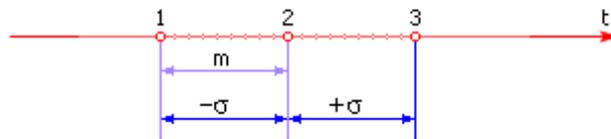


Рис. 28.5. Иллюстрация влияния величины σ на положение события на временной шкале

По смыслу P равно r (см. лекцию 23), поэтому, выражая τ из формулы (*), окончательно для определения интервалов между двумя случайными событиями имеем:

$$\tau = -1/\lambda \cdot \text{Ln}(r),$$

где r — равномерно распределенное от 0 до 1 случайное число, которое берут из ГСЧ, τ — интервал между случайными событиями (случайная величина τ_j).

Пример 1. Рассмотрим поток изделий, приходящих на технологическую операцию. Изделия приходят случайным образом — в среднем восемь штук за сутки (интенсивность потока $\lambda = 8/24$ [ед/час]). Необходимо промоделировать этот процесс в течение $T_n = 100$ часов. $m = 1/\lambda = 24/8 = 3$, то есть в среднем одна деталь за три часа. Заметим, что $\sigma = 3$. На рис. 28.6 представлен алгоритм, генерирующий поток случайных событий.

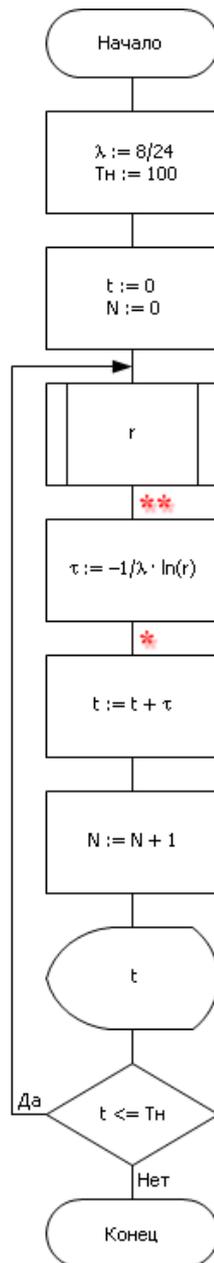


Рис. 28.6. Алгоритм, генерирующий поток случайных событий в заданном λ

На рис. 28.7 показан результат работы алгоритма — моменты времени, когда детали приходили на операцию. Как видно, всего за период $T_n = 100$ производственный узел обработал $N = 33$ изделия. Если запустить алгоритм снова, то N может оказаться равным, например, 34, 35 или 32. Но в среднем, за K прогонов алгоритма N будет равно $33.33\dots$. Если посчитать расстояния между событиями t_{ci} и моментами времени, определяемыми как $3 \cdot i$, то в среднем величина будет равна $\sigma = 3$.

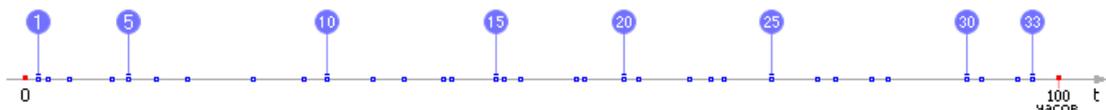


Рис. 28.7. Иллюстрация работы алгоритма, генерирующего поток случайных событий

Моделирование неординарных потоков событий

Если известно, что поток не является ординарным, то необходимо моделировать кроме момента возникновения события еще и число событий, которое могло появиться в этот момент. Например, вагоны на железнодорожную станцию прибывают в составе поезда в случайные моменты времени (ординарный поток поездов). В составе поезда может быть разное (случайное) количество вагонов. В этом случае о потоке вагонов говорят как о потоке неординарных событий.

Допустим, что $M_k = 10$, $\sigma = 4$ (то есть, в среднем в 68 случаях из 100 приходит от 6 до 14 вагонов в составе поезда) и их число распределено по нормальному закону. В место, отмеченное (*) в предыдущем алгоритме (см. рис. 28.6), нужно вставить фрагмент, показанный на рис. 28.8.

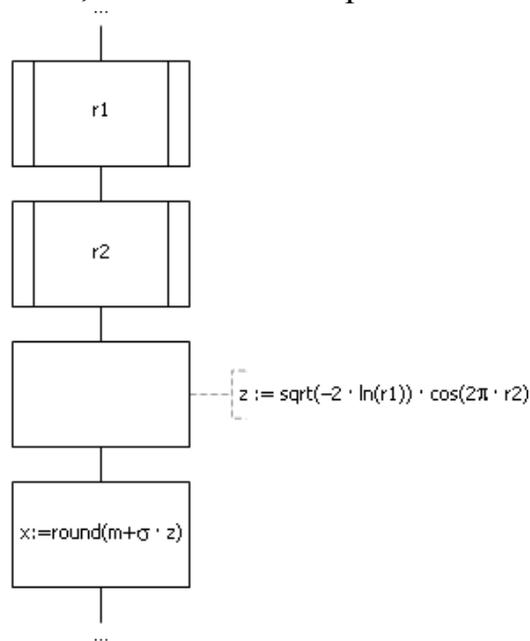
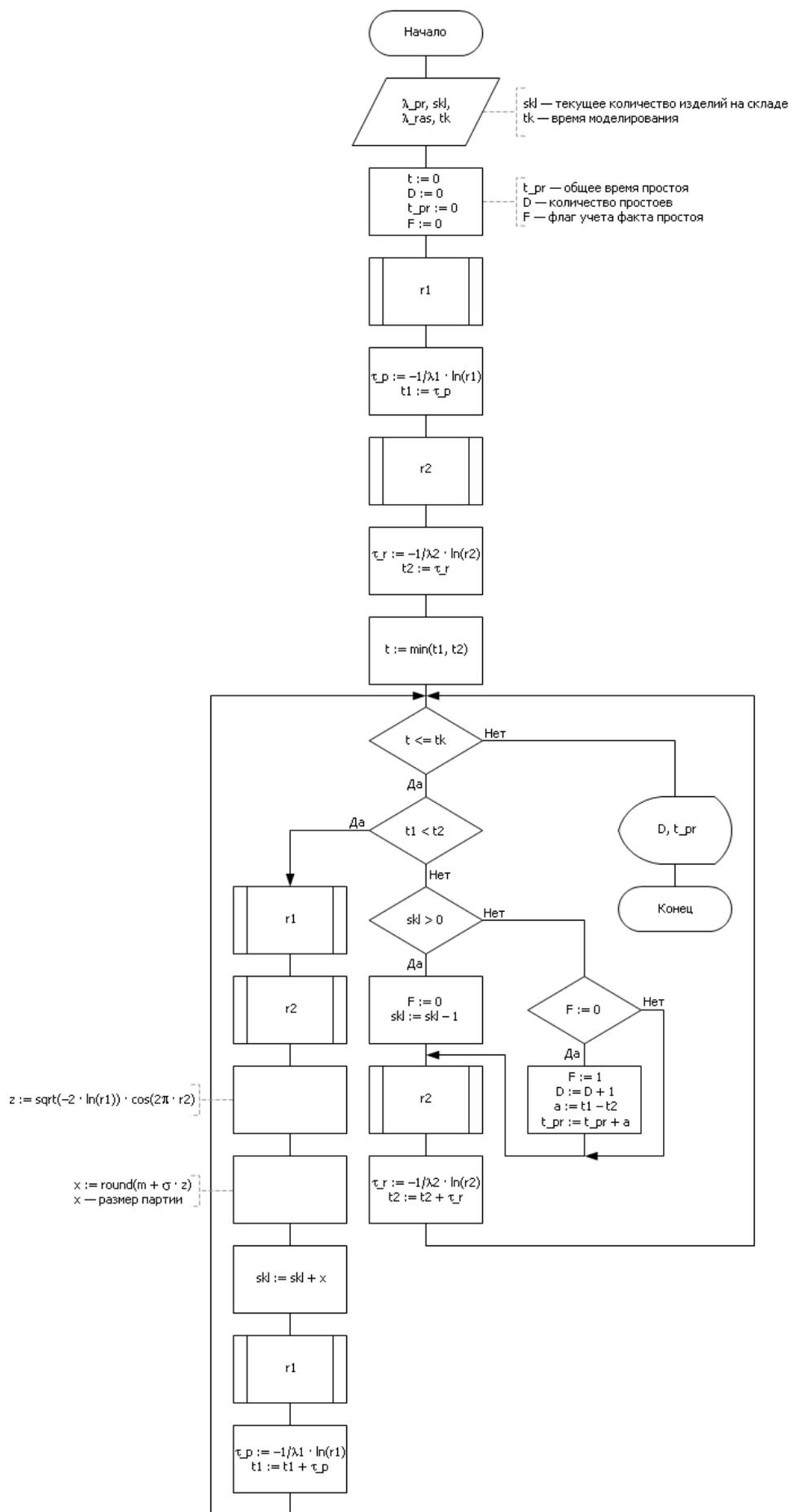


Рис. 28.8. Алгоритм, реализующий неординарный поток случайных событий

Пример 2. Очень полезным в производстве является решение следующей задачи. Каково среднее время суточного простоя оборудования технологического узла, если узел обрабатывает каждое изделие случайное время, заданное интенсивностью потока случайных событий λ_2 ? При этом экспериментально установлено, что привозят изделия на обработку тоже в случайные моменты времени, заданные потоком λ_1 партиями по 8 штук, причем размер партии колеблется случайно по нормальному закону с $m = 8$, $\sigma = 2$ (см. лекцию 25). До начала моделирования $T = 0$ на складе изделий не было. Необходимо промоделировать этот процесс в течение $T_n = 100$ часов. На рис. 28.9 представлен алгоритм, генерирующий случайным образом поток прихода партий изделий на обработку и поток случайных событий — выхода партий изделий с обработки.



29. ПОТОКИ С ПОСЛЕДЕЙСТВИЕМ (ПОТОКИ ЭРЛАНГА)

Интенсивность потока в некотором смысле является математическим ожиданием количества событий в единицу времени (обратная к ней величина указывает на среднее время между событиями). Второй величиной, характеризующей насколько велик разброс во времени прихода событий относительно математического ожидания, является дисперсия [9].

События в потоке следуют точно одно за другим каждые 12 минут без отклонений. Интенсивность такого потока будет равна 5 событий в час. Но если события будут идти случайно, например, 12 ± 2 минуты, то и они в среднем дадут также 5 событий в час. Например, за 200 часов произойдет 1000 событий, величина интенсивности $1000/200 = 5$ событий в час. Поэтому по данной характеристике потоки нельзя отличить друг от друга. Но очевидно, что второй поток все-таки будет более случайным, чем первый. Тем более, если в потоке события будут следовать друг за другом 12 ± 10 минут.

Первый поток мы назовем детерминированным, регулярным, второй и третий — случайными. Мера случайности с увеличением дисперсии (разброса величины интервала между событиями) будет возрастать. В первом потоке дисперсия равна нулю. Данный факт проиллюстрирован на рис. 29.1.

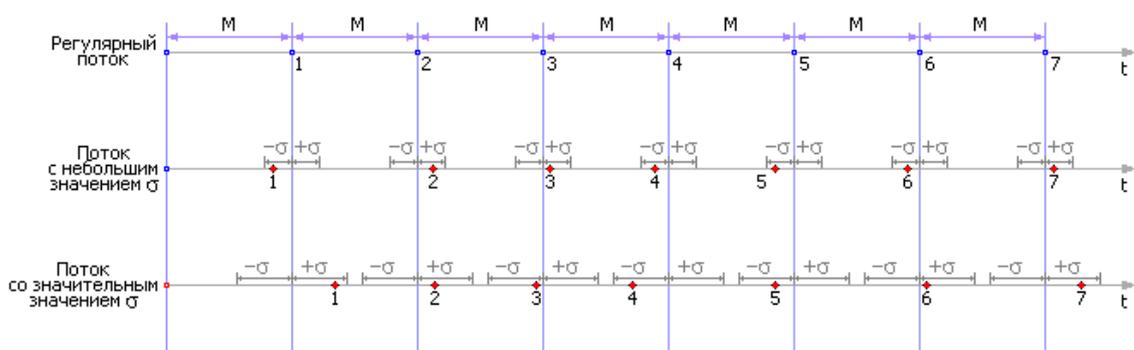


Рис. 29.1. Иллюстрация регулярного и случайного потоков

Именно дисперсия определяет случайность появления события, слабую предсказуемость момента его появления. Важно уметь управлять этой величиной при моделировании случайных потоков. Если предсказать каждое следующее событие трудно, то поток — без последствия (или с малым последствием, связь между событиями отсутствует, события случайны), если поток детерминирован, то последствие велико — каждое событие практически предсказывает момент появления следующего.

Поток Эрланга k -го порядка — это поток случайных событий, получающийся, если в простейшем (см. лекцию 28) случайном потоке сохранить каждое k -е событие, а остальные отбросить (см. рис. 29.2). Порядок потока — мера последствия потока. То есть обратной величиной к мере случайности потока является его порядок.

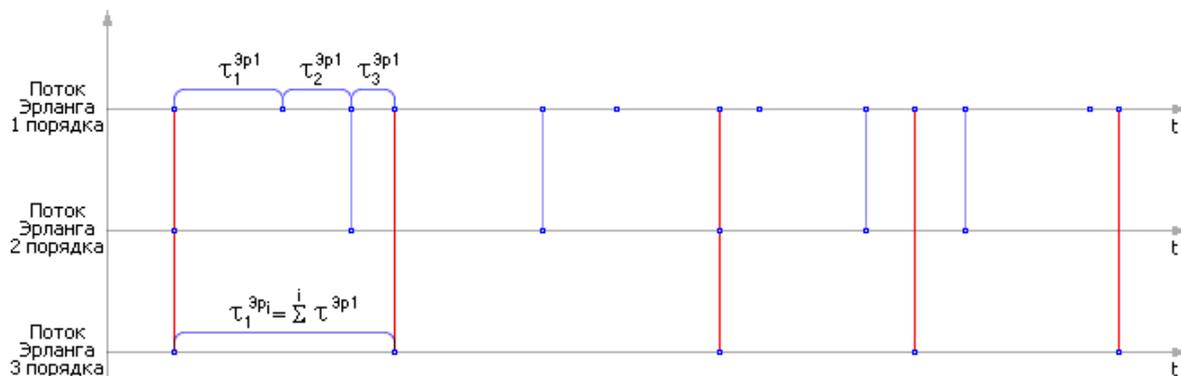


Рис. 29.2. Метод получения потоков Эрланга

Просеивание событий начинает приводить к тому, что между точками появляется последствие, детерминация, которая тем выше, чем больше k . С увеличением k точки ложатся на ось времени все более равномерно, разброс их уменьшается, регулярность увеличивается.

Основано это на том простом и ранее изученном нами факте, что сумма случайных величин есть величина неслучайная (центральная предельная теорема — см. лекцию 25). Чем больше мы сложим случайных величин, тем предсказуемее будет результат (их сумма).

Очевидно, что

$$\tau^{\text{Эр}k} = \sum_{i=1}^k \tau_i^{\text{Эр}1}$$

— интервал между событиями в потоке Эрланга k -го порядка.

Плотность вероятности распределения интервалов между случайными событиями в потоке Эрланга k -го порядка:

$$f(T_{\text{Эр}k}) = \frac{\lambda \cdot (\lambda \cdot t)^{k-1}}{(k-1)!} \cdot e^{-\lambda t}$$

$\lambda_k = \lambda/k$ — интенсивность потока Эрланга k -го порядка, где λ — интенсивность простейшего потока Пуассона, а λ_k интенсивность просеянного k раз потока, то есть в k раз меньше.

Параметры закона Эрланга вычисляются по формулам: $M_k = 1/\lambda_k$, $\sigma_k = 1/\sqrt{k}/\lambda_k$,

Обратите внимание, что в потоке Эрланга $M \neq \sigma$, то есть в потоках с последствием равенство M и σ невозможно.

Более того, при $k \rightarrow \infty$ событие происходит строго в размеренное время, так как $\sigma \rightarrow 0$.

Сравните:

Поток Эрланга 1-го порядка: $m = \sigma_1$ — поток без последствия;

Поток Эрланга i -го порядка: $m \neq \sigma_2$, при этом ($\sigma_2 > 0$) и ($\sigma_1 > \sigma_2$) разброс уменьшается, последствие увеличивается;

Поток Эрланга ∞ -го порядка: $m \neq \sigma = 0$ — регулярный поток.

Из этого следует, что порядок потока Эрланга — есть мера последствия потока.

Пример. Рассмотрим выход из строя лампочек на опоре уличного освещения. Примем время наблюдения 100 лет. Из паспортных данных на эти изделия известно, что среднее время работы изделия на отказ составляет 1,5 года; среднеквадратическое отклонение — 0,5 года.

То есть, задано: $M_k = 1.5$, $\sigma_k = 0.5$.

Поскольку $M_k \neq \sigma_k$, то $k \neq 1$, то есть мы имеем дело с потоком с последствием. Интенсивность этого потока $\lambda_k = 1/M_k = 1/1.5 = 0,67$. Вычисленная интенсивность потока говорит нам о том, что в течение года в среднем перегорает 0,67 лампочки или 67 лампочек за 100 лет.

Так как $\sigma_k = 1/\sqrt{k}/\lambda_k$, и равна 0,5, то вычислим порядок потока Эрланга: $k = 1/\sigma^2/\lambda_k^2 = 1/0.5^2/0.67^2 \approx 9$.

Вычислим интенсивность порождающего потока Пуассона $\lambda = \lambda_k \cdot k = 0,67 \cdot 9 = 6$.

На рис. 29.3 представлен пример алгоритма, реализующего моделирование описанного процесса. Обратите внимание, что берется каждое девятое событие, это обеспечивает достаточно высокую детерминированность потока (то есть малую дисперсию $\sigma_k = 0,5$).

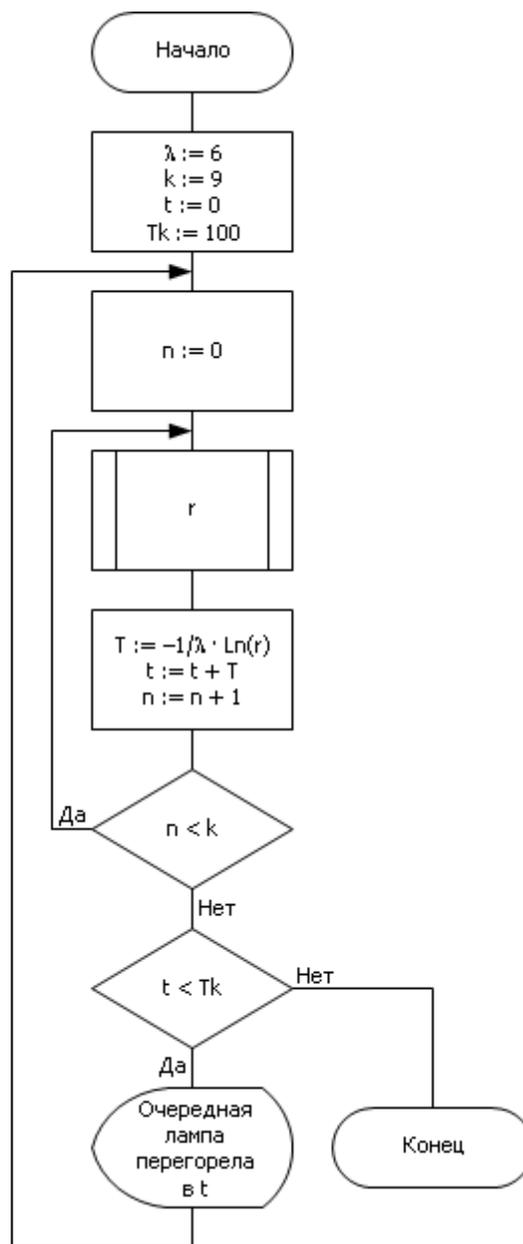


Рис. 29.3. Блок-схема алгоритма моделирования появления случайных событий в виде потока Эрланга

30.МОДЕЛИРОВАНИЕ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

Большой класс систем, которые сложно изучить аналитическими способами, но которые хорошо изучаются методами статистического моделирования, сводится к системам массового обслуживания (СМО).

В СМО подразумевается, что есть *типовые пути* (каналы обслуживания), через которые в процессе обработки проходят *заявки*. Принято говорить, что заявки *обслуживаются* каналами. Каналы могут быть разными по назначению, характеристикам, они могут сочетаться в разных комбинациях; заявки могут находиться в очередях и ожидать обслуживания. Часть заявок может быть обслужена каналами, а части могут отказать в этом. Заявки, с точки зрения системы, абстрактны: это то, что желает обслужиться, то есть пройти определенный путь в системе. Каналы являются также абстракцией: это то, что обслуживает заявки.

Заявки могут приходить неравномерно, каналы могут обслуживать разные заявки за разное время и так далее, количество заявок всегда весьма велико. Все это делает такие системы сложными для изучения и управления, и проследить все причинно-следственные связи в них не представляется возможным. Поэтому принято представление о том, что обслуживание в сложных системах носит случайный характер [9].

Примерами СМО (см. табл. 30.1) могут служить: автобусный маршрут и перевозка пассажиров; производственный конвейер по обработке деталей; влетающая на чужую территорию эскадрилья самолетов, которая «обслуживается» зенитками ПВО; ствол и рожок автомата, которые «обслуживают» патроны; электрические заряды, перемещающиеся в некотором устройстве и т. д.

Таблица 30.1.
Примеры систем массового обслуживания

СМО	Заявки	Каналы
Автобусный маршрут и перевозка пассажиров	Пассажиры	Автобусы
Производственный конвейер по обработке деталей	Детали, узлы	Станки, склады
Влетающая на чужую территорию эскадрилья самолетов, которая «обслуживается» зенитками ПВО	Самолеты	Зенитные орудия, радары, стрелки, снаряды
Ствол и рожок автомата, которые «обслуживают» патроны	Патроны	Ствол, рожок
Электрические заряды, перемещающиеся в некотором устройстве	Заряды	Каскады технического устройства

Но все эти системы объединены в один класс СМО, поскольку подход к их изучению един. Он состоит в том, что, во-первых, с помощью генератора случайных чисел разыгрываются случайные числа, которые имитируют случайные моменты появления заявок и время их обслуживания в каналах. Но в совокупности эти случайные числа, конечно, подчинены *статистическим закономерностям*.

Пусть сказано: «заявки в среднем приходят в количестве 5 штук в час». Это означает, что времена между приходом двух соседних заявок случайны, например: 0.1; 0.3; 0.1; 0.4; 0.2, как это показано на рис. 30.1, но в сумме они дают в среднем 1 (обратите внимание, что в примере это не точно 1, а 1.1 — но зато в другой час эта сумма, например, может быть равной 0.9); и только за достаточно большое время среднее этих чисел станет близким к одному часу.

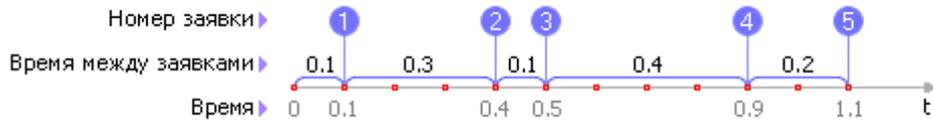
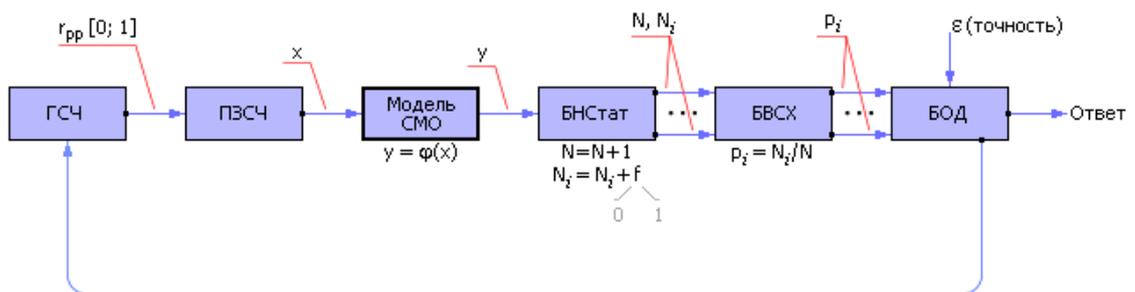


Рис. 30.1. Случайный процесс прихода заявок в СМО

Результат (например, пропускная способность системы), конечно, тоже будет случайной величиной на отдельных промежутках времени. Но измеренная на большом промежутке времени, эта величина будет уже, в среднем, соответствовать точному решению. Для характеристики СМО интересуются ответами в статистическом смысле.

Итак, систему испытывают случайными входными сигналами, подчиненными заданному статистическому закону, а в качестве результата принимают статистические показатели, усредненные по времени рассмотрения или по количеству опытов. Ранее, в лекции 21 (см. рис. 21.1), мы уже разработали схему для такого статистического эксперимента (см. рис. 30.2).



ГСЧ — генератор случайных чисел.
 ПЗСЧ — преобразователь закона случайных чисел.
 БНСтат — блок накопления статистики.
 БВСХ — блок вычисления статистических характеристик.
 БОД — блок оценки достоверности.

Рис. 30.2. Схема статистического эксперимента для изучения систем массового обслуживания

Во-вторых, все модели СМО собираются типовым образом из небольшого набора элементов (канал, источник заявок, очередь, заявка, дисциплина обслуживания, стек, кольцо и так далее), что позволяет имитировать эти задачи *типовым* образом. Для этого модель системы собирают из конструктора таких элементов. Неважно, какая конкретно система изучается, важно, что схема системы собирается из одних и тех же элементов. Разумеется, структура схемы будет всегда различной.

Перечислим некоторые основные понятия СМО.

Каналы — то, что обслуживает; бывают горячие (начинают обслуживать заявку в момент ее поступления в канал) и холодные (каналу для начала обслуживания требуется время на подготовку). **Источники заявок** — порождают заявки в случайные моменты времени, согласно заданному пользователем статистическому закону. **Заявки**, они же **клиенты**, входят в систему (порождаются источниками заявок), проходят через ее элементы (обслуживаются), покидают ее обслуженными или неудовлетворенными. Бывают **нетерпеливые заявки** — такие, которым надоело ожидать или находиться в системе и которые покидают по собственной воле СМО. Заявки образуют **потоки** — поток заявок на входе системы, поток обслуженных заявок, поток отказанных заявок. Поток характеризуется количеством заявок определенного сорта, наблюдаемым в некотором месте СМО за единицу времени (час, сутки, месяц), то есть поток есть величина статистическая.

Очереди характеризуются правилами стояния в очереди (дисциплиной обслуживания), количеством мест в очереди (сколько клиентов максимум может находиться в очереди), структурой очереди (связь между местами в очереди). Бывают ограниченные и неограниченные очереди. Перечислим важнейшие дисциплины обслуживания. FIFO (First In, First Out — первым пришел, первым ушел): если заявка первой пришла в очередь, то она первой уйдет на обслуживание. LIFO (Last In, First Out — последним пришел, первым ушел): если заявка последней пришла в очередь, то она первой уйдет на обслуживание (пример — патроны в рожке автомата). SF (Short Forward — короткие вперед): в первую очередь обслуживаются те заявки из очереди, которые имеют меньшее время обслуживания.

Пример, показывающий, как правильный выбор той или иной дисциплины обслуживания позволяет получить ощутимую экономию по времени.

Пусть имеется два магазина. В магазине № 1 обслуживание осуществляется в порядке очереди, то есть здесь реализована дисциплина обслуживания FIFO (см. рис. 30.3).

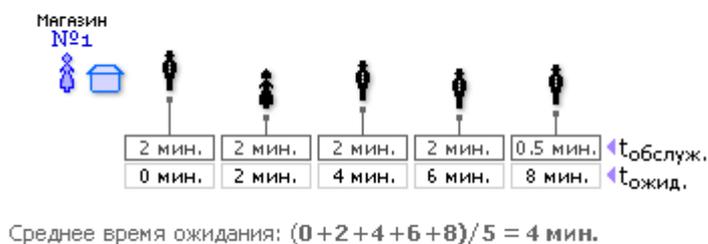


Рис. 30.3. Организация очереди по дисциплине FIFO

Время обслуживания $t_{\text{обслуж.}}$ на рис. 30.3 показывает, сколько времени продавец затратит на обслуживание одного покупателя. Понятно, что при покупке штучного товара продавец затратит меньше времени на обслуживание, чем при покупке, скажем, сыпучих продуктов, требующих дополнительных манипуляций (набрать, взвесить, высчитать цену и т. п.). Время ожида-

ния $t_{\text{ожид.}}$ показывает, через какое время очередной покупатель будет обслужен продавцом.

В магазине № 2 реализована дисциплина SF (см. рис. 30.4), означающая, что штучный товар можно купить вне очереди, так как время обслуживания $t_{\text{обслуж.}}$ такой покупки невелико.

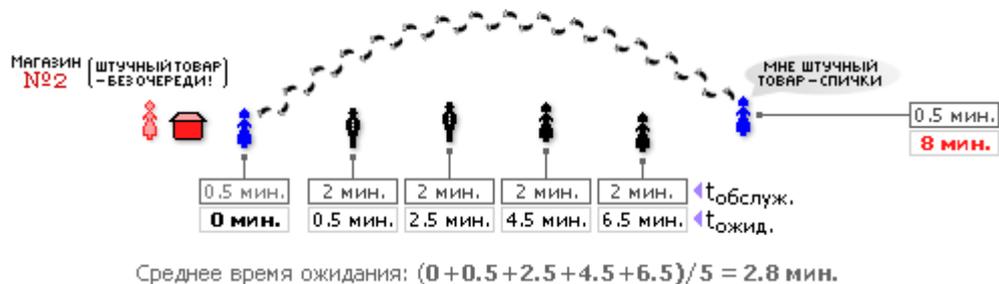


Рис. 30.4. Организация очереди по дисциплине SF

Как видно из обоих рисунков, последний (пятый) покупатель собирается приобрести штучный товар, поэтому время его обслуживания невелико — 0.5 минут. Если этот покупатель придет в магазин № 1, он будет вынужден выстоять в очереди целых 8 минут, в то время как в магазине № 2 его обслужат сразу же, вне очереди. Таким образом, среднее время обслуживания каждого из покупателей в магазине с дисциплиной обслуживания FIFO составит 4 минуты, а в магазине с дисциплиной обслуживания КВ — лишь 2,8 минуты. А общественная польза, экономия времени составит: $(1 - 2,8/4) \cdot 100\% = 30$ процентов. Итак, 30% сэкономленного для общества времени — и это лишь за счет правильного выбора дисциплины обслуживания.

При анализе результатов моделирования важно также указать интересы и степень их выполнения. Различают интересы клиента и интересы владельца системы. Заметим, что эти интересы совпадают не всегда.

Судить о результатах работы СМО можно по показателям. Наиболее популярные из них:

- вероятность обслуживания клиента системой;
- пропускная способность системы;
- вероятность отказа клиенту в обслуживании;
- вероятность занятости каждого из канала и всех вместе;
- среднее время занятости каждого канала;
- вероятность занятости всех каналов;
- среднее количество занятых каналов;
- вероятность простоя каждого канала;
- вероятность простоя всей системы;
- среднее количество заявок, стоящих в очереди;
- среднее время ожидания заявки в очереди;
- среднее время обслуживания заявки;
- среднее время нахождения заявки в системе.

Судить о качестве полученной системы нужно по совокупности значений показателей. При анализе результатов моделирования (показателей) важно так-

же обращать внимание на интересы клиента и интересы владельца системы, то есть минимизировать или максимизировать надо тот или иной показатель, а также на степень их выполнения. Заметим, что чаще всего интересы клиента и владельца между собой не совпадают или совпадают не всегда. Показатели будем обозначать далее $H = \{h_1, h_2, \dots\}$.

Параметрами СМО могут быть: интенсивность потока заявок, интенсивность потока обслуживания, среднее время, в течение которого заявка готова ожидать обслуживания в очереди, количество каналов обслуживания, дисциплина обслуживания и так далее. Параметры — это то, что влияет на показатели системы. Параметры будем обозначать далее как $R = \{r_1, r_2, \dots\}$.

Пример. Автозаправочная станция (АЗС).

1. Постановка задачи. На рис. 30.5 приведен план АЗС. Рассмотрим метод моделирования СМО на ее примере и план ее исследования. Водители, проезжая по дороге мимо АЗС по дороге, могут захотеть заправить свой автомобиль. Хотят обслужиться (заправить машину бензином) не все автомобилисты подряд; допустим, что из всего потока машин на заправку в среднем заезжает 5 машин в час.



Рис. 30.5. План моделируемой АЗС

На АЗС две одинаковые колонки, статистическая производительность каждой из которых известна. Первая колонка в среднем обслуживает 1 машину в час, вторая в среднем — 3 машины в час. Владелец АЗС заасфальтировал для машин место, где они могут ожидать обслуживания. Если колонки заняты, то на этом месте могут ожидать обслуживания другие машины, но не более двух одновременно. Очередь будем считать общей. Как только одна из колонок освободится, то первая машина из очереди может занять ее место на колонке (при этом вторая машина продвигается на первое место в очереди). Если появляется третья машина, а все места (их два) в очереди заняты, то ей отказывают в обслуживании, так как стоять на дороге запрещено (см. дорожные знаки около АЗС). Такая машина уезжает прочь из системы навсегда и как потенциальный клиент является потерянной для владельца АЗС. Можно

усложнить задачу, рассмотрев кассу (еще один канал обслуживания, куда надо попасть после обслуживания в одной из колонок) и очередь к ней и так далее. Но в простейшем варианте очевидно, что пути движения потоков заявок по СМО можно изобразить в виде эквивалентной схемы, а добавив значения и обозначения характеристик каждого элемента СМО, получаем окончательно схему, изображенную на рис. 30.6.

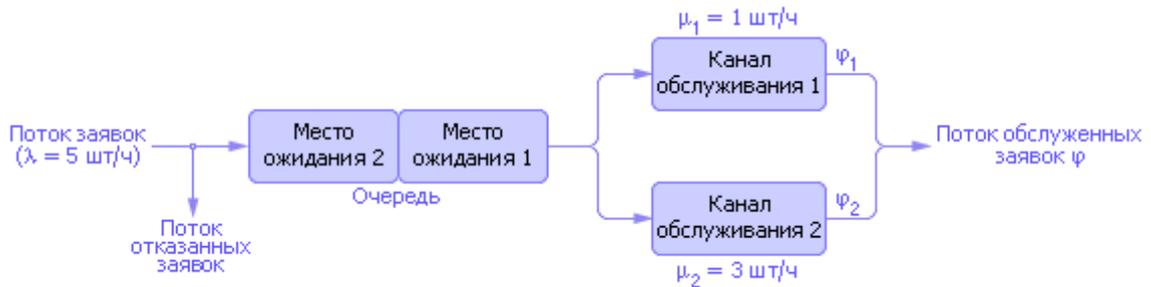


Рис. 30.6. Эквивалентная схема объекта моделирования

2. Метод исследования СМО. Применим в нашем примере принцип **последовательной проводки заявок** (подробно о принципах моделирования см. лекцию 32). Его идея заключается в том, что заявку проводят через всю систему от входа до выхода, и только после этого берутся за моделирование следующей заявки.

Построим временную диаграмму работы СМО, отражая на каждой линейке (ось времени t) состояние отдельного элемента системы. Временных линеек проводится столько, сколько имеется различных мест в СМО, потоков. В нашем примере их 7 (поток заявок, поток ожидания на первом месте в очереди, поток ожидания на втором месте в очереди, поток обслуживания в канале 1, поток обслуживания в канале 2, поток обслуженных системой заявок, поток отказанных заявок).

Для генерации времени прихода заявок используем формулу вычисления интервала между моментами прихода двух случайных событий (см. лекцию 28):

$$t = -\frac{1}{\lambda} \cdot \ln(r)$$

В этой формуле величина потока λ должна быть задана (до этого она должна быть определена экспериментально на объекте как статистическое среднее), r — случайное равномерно распределенное число от 0 до 1 из ГСЧ или таблицы, в которой случайные числа нужно брать подряд (не выбирая специально).

Задача. Сгенерируйте поток из 10 случайных событий с интенсивностью появления событий 5 шт/час.

Решение задачи. Возьмем случайные числа, равномерно распределенные в интервале от 0 до 1 (см. таблицу), и вычислим их натуральные логарифмы (см. табл. 30.2).

Таблица 30.2.

Фрагмент таблицы случайных чисел и их логарифмов

$r_{pp}[0; 1]$	$\ln(r_{pp}[0; 1])$
0.0333	-3.4022
0.3557	-1.0337
0.2172	-1.5269
0.5370	-0.6218

Формула пуассоновского потока определяет **расстояние между двумя случайными событиями** следующим образом: $t = -\ln(r_{pp})/\lambda$. Тогда, учитывая, что $\lambda = 5$, имеем расстояния между двумя случайными соседними событиями: 0,68, 0,21, 0,31, 0,12 часа. То есть события наступают: первое — в момент времени $t = 0$, второе — в момент времени $t = 0,68$, третье — в момент времени $t = 0,89$, четвертое — в момент времени $t = 1,20$, пятое — в момент времени $t = 1,32$ и так далее. События — приход заявок отразим на первой линейке (см. рис. 30.7).

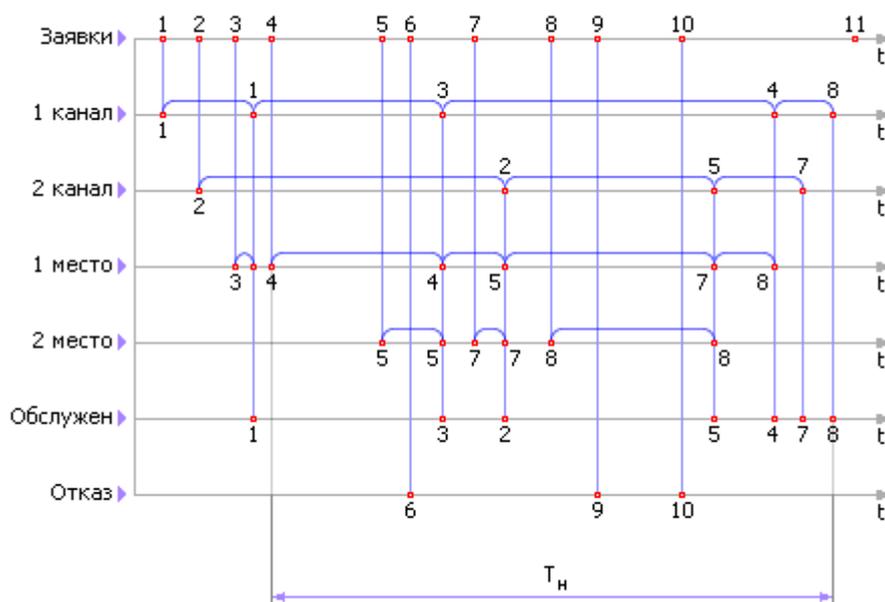


Рис. 30.7. Временная диаграмма работы СМО

Берется первая заявка и, так как в этот момент каналы свободны, устанавливается на обслуживание в первый канал. Заявка 1 переносится на линейку «1 канал».

Время обслуживания в канале тоже случайное и вычисляется по аналогичной формуле:

$$t = -\frac{1}{\lambda} \cdot \ln(r)$$

где роль интенсивности играет величина потока обслуживания μ_1 или μ_2 , в зависимости от того, какой канал обслуживает заявку. Находим на диаграмме момент окончания обслуживания, откладывая сгенерированное время обслуживания от момента начала обслуживания, и опускаем заявку на линейку «Обслуженные».

Заявка прошла в СМО весь путь. Теперь можно, согласно принципу последовательной проводки заявок, также проимитировать путь второй заявки.

Если в некоторый момент окажется, что оба канала заняты, то следует установить заявку в очередь. На рис. 30.7 это заявка с номером 3. По условиям задачи в очереди в отличие от каналов заявки находятся не случайное время, а ожидают, когда освободится какой-то из каналов. После освобождения канала заявка поднимается на линейку соответствующего канала и там организуется ее обслуживание.

Если все места в очереди в момент, когда придет очередная заявка, будут заняты, то заявку следует отправить на линейку «Отказанные». На рис. 30.7 это заявка с номером 6.

Процедуру имитации обслуживания заявок продолжают некоторое время наблюдения T_n . Чем больше это время, тем точнее в дальнейшем будут результаты моделирования. Реально для простых систем выбирают T_n , равное 50—100 и более часов, хотя иногда лучше мерить эту величину количеством рассмотренных заявок.

Анализ временной диаграммы

Анализ проведем на уже рассмотренном примере.

Сначала нужно дождаться установившегося режима. Откидываем первые четыре заявки как нехарактерные, протекающие во время процесса установления работы системы. Измеряем время наблюдения, допустим, что в нашем примере оно составит $T_n = 5$ часов. Подсчитываем из диаграммы количество обслуженных заявок $N_{обс.}$, времена простоя и другие величины. В результате можем вычислить показатели, характеризующие качество работы СМО.

1. Вероятность обслуживания: $P_{обс.} = N_{обс.}/N = 5/7 = 0.714$. Для расчета вероятности обслуживания заявки в системе достаточно разделить число заявок, которым удалось обслужиться за время T_n (см. линейку «Обслуженные») $N_{обс.}$, на число заявок N , которые хотели обслужиться за это же время. Как и раньше вероятность экспериментально определяем отношением свершившихся событий к общему числу событий, которые могли совершиться!
2. Пропускная способность системы: $A = N_{обс.}/T_n = 5/5 = 1.4$ [шт/час]. Для расчета пропускной способности системы достаточно разделить число обслуженных заявок $N_{обс.}$ на время T_n , за которое произошло это обслуживание (см. линейку «Обслуженные»).
3. Вероятность отказа: $P_{отк.} = N_{отк.}/N = 3/7 = 0.43$. Для расчета вероятности отказа заявке в обслуживании достаточно разделить число заявок $N_{отк.}$, которым отказали за время T_n (см. линейку «Отказанные»), на число заявок N , которые хотели обслужиться за это же время, то есть поступили в систему. $P_{отк.} + P_{обс.}$ в теории должно быть равно 1. На самом деле экспериментально получилось, что $P_{отк.} + P_{обс.} = 0.714 + 0.43 = 1.144$. Эта неточность объясняется тем, что время наблюдения T_n мало и статистика накоплена недостаточная для

- получения точного ответа. Погрешность это показателя сейчас составляет 14%.
4. Вероятность занятости одного канала: $P_1 = T_{\text{зан.}}/T_{\text{н}} = 0.05/5 = 0.01$, где $T_{\text{зан.}}$ — время занятости только одного канала (первого или второго). Измерениям подлежат временные отрезки, на которых происходят определенные события. Например, на диаграмме ищутся такие отрезки, во время которых заняты или первый или второй канал. В данном примере есть один такой отрезок в конце диаграммы длиной 0.05 часа. Доля этого отрезка в общем времени рассмотрения ($T_{\text{н}} = 5$ часов) определяется делением и составляет искомую вероятность занятости.
 5. Вероятность занятости двух каналов: $P_2 = T_{\text{зан.}}/T_{\text{н}} = 4.95/5 = 0.99$. На диаграмме ищутся такие отрезки, во время которых одновременно заняты и первый, и второй канал. В данном примере таких отрезков четыре, их сумма равна 4.95 часа. Доля продолжительности этих события в общем времени рассмотрения ($T_{\text{н}} = 5$ часов) определяется делением и составляет искомую вероятность занятости.
 6. Среднее количество занятых каналов: $N_{\text{ск}} = 0 \cdot P_0 + 1 \cdot P_1 + 2 \cdot P_2 = 0.01 + 2 \cdot 0.99 = 1.99$. Чтобы подсчитать, сколько каналов занято в системе в среднем, достаточно знать долю (вероятность занятости одного канала) и умножить на вес этой доли (один канал), знать долю (вероятность занятости двух каналов) и умножить на вес этой доли (два канала) и так далее. Полученная цифра 1.99 говорит о том, что из возможных двух каналов в среднем загружено 1.99 канала. Это высокий показатель загрузки, 99.5%, система хорошо использует ресурс.
 7. Вероятность простоя хотя бы одного канала: $P_1^* = T_{\text{простоя1}}/T_{\text{н}} = 0.05/5 = 0.01$.
 8. Вероятность простоя двух каналов одновременно: $P_2^* = T_{\text{простоя2}}/T_{\text{н}} = 0$.
 9. Вероятность простоя всей системы: $P_c^* = T_{\text{простоя сист.}}/T_{\text{н}} = 0$.
 10. Среднее количество заявок в очереди: $N_{\text{сз}} = 0 \cdot P_{0з} + 1 \cdot P_{1з} + 2 \cdot P_{2з} = 0.34 + 2 \cdot 0.64 = 1.62$ [шт]. Чтобы определить среднее количество заявок в очереди, надо определить отдельно вероятность того, что в очереди будет одна заявка $P_{1з}$, вероятность того, в очереди будет стоять две заявки $P_{2з}$ и т. д. и снова с соответствующими весами их сложить.
 11. Вероятность того, что в очереди будет одна заявка: $P_{1з} = T_{1з}/T_{\text{н}} = 1.7/5 = 0.34$ (всего на диаграмме четырех таких отрезка, в сумме дающих 1.7 часа).
 12. Вероятность того, в очереди будет стоять одновременно две заявки: $P_{2з} = T_{2з}/T_{\text{н}} = 3.2/5 = 0.64$ (всего на диаграмме три таких отрезка, в сумме дающих 3.25 часа).
 13. Среднее время ожидания заявки в очереди:

$$T_{\text{ср.ож.}} = \frac{\sum_{i=1}^N T_{\text{ож.}i}}{N}$$

(Сложить все временные интервалы, в течение которых какая-либо заявка находилась в очереди, и разделить на количество заявок). На временной диаграмме таких заявок 4.

14. Среднее время обслуживания заявки:

$$T_{\text{ср.обсл.}} = \frac{\sum_{i=1}^N T_{\text{обсл.}i}}{N}$$

(Сложить все временные интервалы, в течение которых какая-либо заявка находилась на обслуживании в каком-либо канале, и разделить на количество заявок).

15. Среднее время нахождения заявки в системе: $T_{\text{ср. сист.}} = T_{\text{ср. ож.}} + T_{\text{ср. обсл.}}$

16. Среднее количество заявок в системе:

$$N_{\text{ср.}} = \frac{\sum_{i=1}^K (N_{2i} + N_{3i} + N_{4i} + N_{5i})}{K}$$

Разобьем интервал наблюдения, например, на десятиминутки. Получится на пяти часах K подынтервалов (в нашем случае $K = 30$). В каждом подынтервале определим по временной диаграмме, сколько заявок в этот момент находится в системе. Смотреть надо на 2, 3, 4 и 5-ю линейки — какие из них заняты в данный момент. Затем сумму K слагаемых усреднить.

Далее оценить точность каждого из полученных результатов. То есть ответить на вопрос: насколько мы можем доверять этим значениям? Оценка точности проводится по методике, описанной в лекции 34.

Если точность не является удовлетворительной, то следует увеличить время эксперимента и тем самым улучшить статистику. Можно сделать и по-другому. Снова несколько раз запустить эксперимент на время $T_{\text{н}}$. А в последствии усреднить значения этих экспериментов. И снова проверить результаты на критерий точности. Эту процедуру следует повторять до тех пор, пока не будет достигнута требуемая точность.

Следует составить таблицу результатов и оценить значения каждого из них с точки зрения клиента и владельца СМО (см. табл. 30.3). В конце, учитывая сказанное в каждом пункте, следует сделать общий вывод. Таблица должна иметь примерно такой вид, какой показан в табл. 30.3.

Таблица 30.3.
Показатели СМО

Показатель	Формула	Значение	Интересы владельца СМО	Интересы клиента СМО
Вероятность обслуживания	$P_{\text{обс.}} = N_{\text{обс.}}/N$	0.714	Вероятность обслуживания мала, много клиентов уходят из системы неудовлетворенными, их деньги для владельца потеряны. Это «минус». Рекомендация: увеличить вероят-	Вероятность обслуживания мала, каждый третий клиент хочет, но не может обслужиться. Это «минус».

			ность обслуживания.	Рекомендация: увеличить вероятность обслуживания.
...
Среднее количество заявок в очереди	$N_{сз} = 0 \cdot P_{03} + 1 \cdot P_{13} + 2 \cdot P_{23}$	1.62	<p>Очередь практически все время вся забита. Все места в очереди используются достаточно эффективно. Вложения на организацию очереди окупают затраты на нее. Это «плюс».</p> <p>Клиенты, которые долго стоят в очереди, могут уйти, не дождавшись обслуживания. Клиенты, простаивая, могут нанести ущерб системе, ломать оборудование. Много отказов, потерянных клиентов. Это «минусы».</p> <p>Рекомендация: увеличить число мест в очереди, увеличить пропускную способность.</p>	<p>Очередь практически все время вся забита. Клиенту приходится стоять в очереди, прежде чем он попадет на обслуживание. Клиент может не попасть даже в очередь. Это «минус».</p> <p>Рекомендация: увеличить пропускную способность, увеличить число мест в очереди.</p>
Общий итог:			В интересах владельца: а) увеличить пропускную способность каналов, чтобы не терять клиентов (правда, модернизация каналов стоит денег); б) увеличить число мест в очереди (это тоже стоит денег), чтобы задержать потенциальных клиентов.	Клиенты заинтересованы в значительном увеличении пропускной способности для уменьшения времени ожидания и уменьшения отказов.

Синтез СМО

Мы проделали анализ существующей системы. Это дало возможность увидеть ее недостатки и определить направления улучшения ее качества. Но остаются непонятными ответы на конкретные вопросы, что именно надо сделать — увеличивать количество каналов или увеличивать их пропускную способность, или увеличивать количество мест в очереди, и, если увеличивать, то насколько? Есть и такие вопросы, что лучше - создать 3 канала с производительностью 5 шт/час или один с производительностью 15 шт/час?

Чтобы оценить чувствительность каждого показателя к изменению значения определенного параметра, поступают следующим образом. Фиксируют все параметры кроме одного, выбранного. Затем снимают значение всех показателей при нескольких значениях этого выбранного параметра. Конечно, приходится повторять снова и снова процедуру имитации и усреднять показатели при каждом значении параметра, оценивать точность. Но в результате получают надежные статистические зависимости характеристик (показателей) от параметра.

Например, для 12 показателей нашего примера можно получить 12 зависимостей от одного параметра: зависимость вероятности отказов $P_{отк.}$ от количества мест в очереди (КМО), зависимость пропускной способности A от количества мест в очереди, и так далее (см. рис. 30.8).

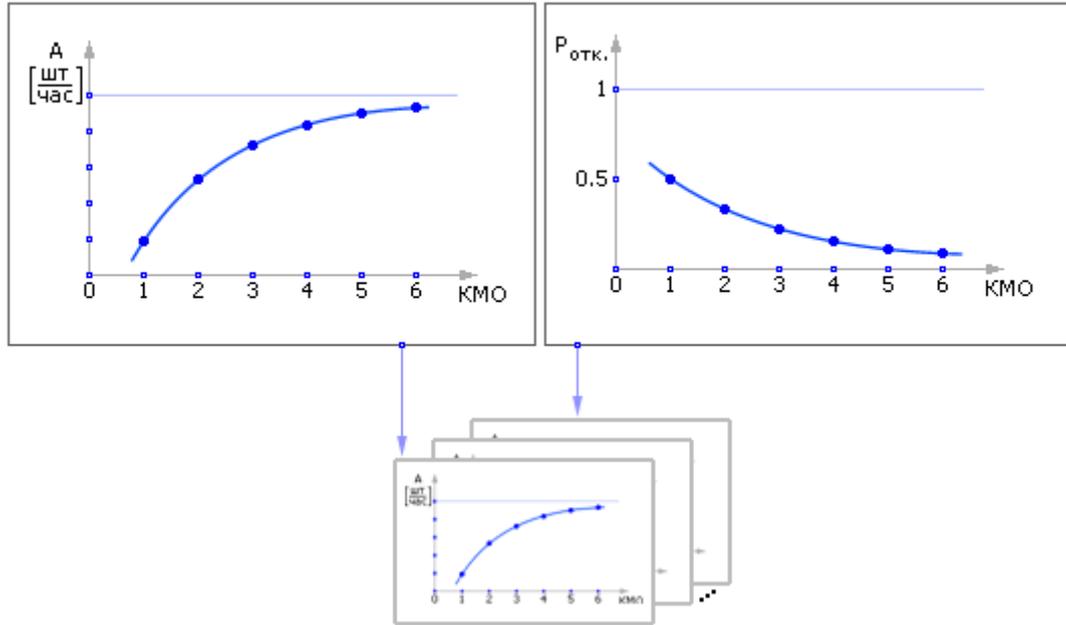


Рис. 30.8. Вид зависимостей показателей от параметров СМО

Затем так же можно снять еще 12 зависимостей показателей P от другого параметра R , зафиксировав остальные параметры. И так далее. Образуется своеобразная матрица зависимостей показателей P от параметров R , по которой можно провести дополнительный анализ о перспективах движения (улучшения показателей) в ту или иную сторону. Наклон кривых хорошо показывает чувствительность, эффект от движения по определенному показателю. В математике эту матрицу называют **якобианом J** , в которой роль наклона кривых играют значения производных $\Delta P_i / \Delta R_j$, см. рис. 30.9. (Напомним, что производная связана геометрически с углом наклона касательной к зависимости.)

$$\|J\| = \begin{vmatrix} J_{11} & J_{12} & J_{13} & \dots & J_{1,12} \\ J_{21} & J_{22} & J_{23} & \dots & J_{2,12} \\ \dots & \dots & \dots & \dots & \dots \\ J_{51} & J_{52} & J_{53} & \dots & J_{5,12} \end{vmatrix} = \begin{vmatrix} \frac{\Delta P_1}{\Delta R_1} & \frac{\Delta P_1}{\Delta R_2} & \frac{\Delta P_1}{\Delta R_3} & \dots & \frac{\Delta P_1}{\Delta R_{12}} \\ \frac{\Delta P_2}{\Delta R_1} & \frac{\Delta P_2}{\Delta R_2} & \frac{\Delta P_2}{\Delta R_3} & \dots & \frac{\Delta P_2}{\Delta R_{12}} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{\Delta P_5}{\Delta R_1} & \frac{\Delta P_5}{\Delta R_2} & \frac{\Delta P_5}{\Delta R_3} & \dots & \frac{\Delta P_5}{\Delta R_{12}} \end{vmatrix} =$$

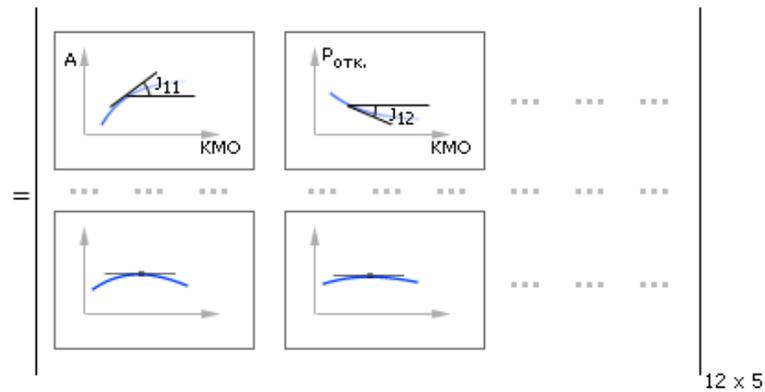


Рис. 30.9. Якобиан — матрица чувствительностей показателей в зависимости от изменения параметров СМО

Если показателей 12, а параметров, например, 5, то матрица имеет размерность 12×5 . Каждый элемент матрицы — кривая, зависимость i -го показателя от j -го параметра. Каждая точка кривой — среднее значение показателя на достаточно представительном отрезке T_n или усреднено по нескольким экспериментам.

Кривые снимались в предположении того, что все параметры кроме одного в процессе их снятия были неизменны. (Если бы все параметры меняли значения, то кривые были бы другими. Но так не делают, так как получится полная неразбериха и зависимостей не будет видно.)

Если на основании рассмотрения снятых кривых принимается решение о том, что некоторый параметр будет в СМО изменен, то все кривые для новой точки, в которой опять будет исследоваться вопрос о том, какой параметр следует изменить, чтобы улучшить показатели, следует снимать заново.

Так шаг за шагом можно попытаться улучшить качество системы. Но пока эта методика не может ответить на ряд вопросов. Дело в том, что, во-первых, если кривые монотонно растут, то возникает вопрос, где же все-таки следует остановиться. Во-вторых, могут возникать противоречия, один показатель может улучшаться при изменении выбранного параметра, в то время как другой будет одновременно ухудшаться. В-третьих, ряд параметров сложно выразить численно, например, изменение дисциплины обслуживания, изменение направлений потоков, изменение топологии СМО. Поиск решения в двух последних случаях проводится с применением методов экспертизы (см. лекцию 36. Экспертиза) и методами искусственного интеллекта (см. генетические алгоритмы в искусственном интеллекте).

Поэтому сейчас обсудим только первый вопрос. Как принять решение, каким должно быть все-таки значение параметра, если с его ростом показатель все время монотонно улучшается? Вряд ли значение бесконечности устроит инженера.

Параметр R — управление, это то, что находится в распоряжении владельца СМО (например, возможность заасфальтировать площадку и тем самым увеличить количество мест в очереди, поставить дополнительные каналы, увеличить поток заявок за счет увеличения затрат на рекламу и так далее). Меняя управление, можно влиять на значение показателя P , цель, критерий (вероятность отказов, пропускную способность, среднее время обслуживания и так далее). Из рис. 30.10 видно, что если увеличивать управление R , то можно добиться всегда улучшения показателя P . Но очевидно, что любое управление связано с затратами Z . И чем больше прилагают усилия для управления, чем больше значение управляющего параметра, тем больше затраты. Обычно затраты на управление растут линейно: $Z = C_1 \cdot R$. Хотя встречаются случаи, когда, например, в иерархических системах, они растут экспоненциально, иногда — обратно экспоненциально (скидки за опт) и так далее.

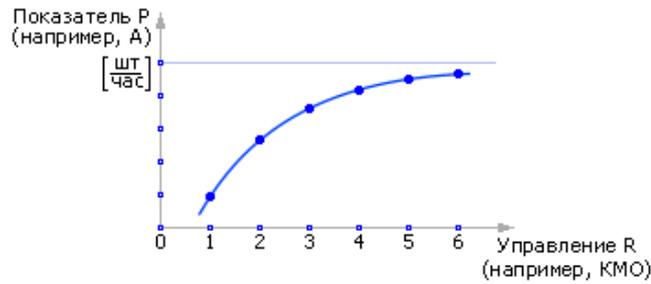


Рис. 30.10. Зависимость показателя P от управляемого параметра R (пример)

Когда-то вложение все новых затрат просто перестанет себя окупать. Например, эффект от заасфальтированной площадки размером в 1 км^2 вряд ли окупит затраты владельца бензоколонки в Урюпинске, там просто не будет столько желающих заправиться бензином. Показатель P в сложных системах не может расти бесконечно. Рано или поздно его рост замедляется. А затраты Z растут (см. рис. 30.11).

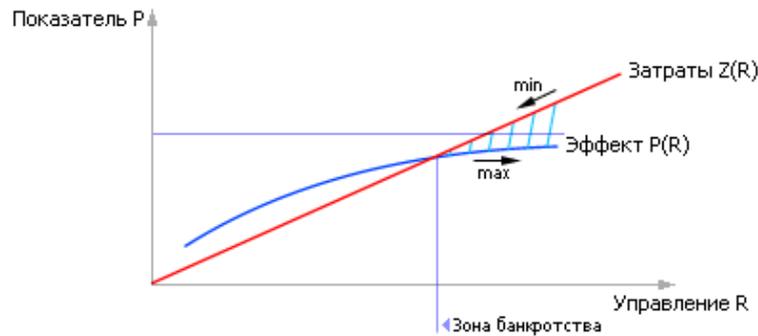


Рис. 30.11. Зависимости эффекта от применения показателя P и затрат Z на его получение как функции управляемого параметра R

Из рис. 30.11 видно, что при назначении цены C_1 за единицу затрат R и цены C_2 за единицу показателя P , эти кривые можно сложить. Кривые складывают, если их требуется одновременно минимизировать или максимизировать. Если одна кривая подлежит максимизации, а другая минимизации, то следует найти их разность, например по точкам. Тогда результирующая кривая (см. рис. 30.12), учитывающая и эффект от управления и затраты на это, будет иметь экстремум. Значение параметра R , доставляющего экстремум функции, и есть решение задачи синтеза.

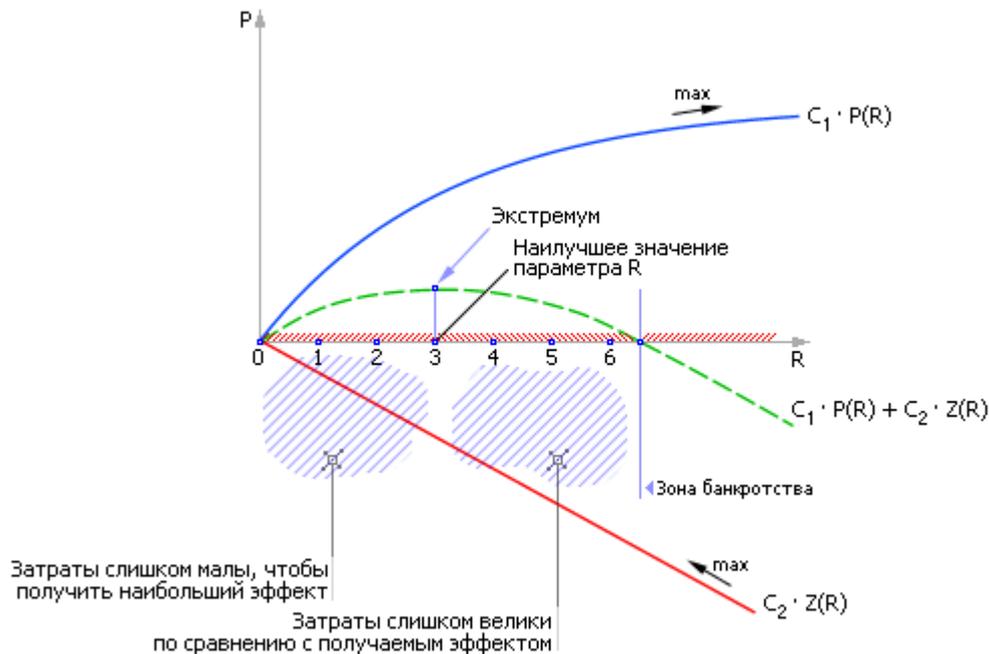


Рис. 30.12. Суммарная зависимость эффекта от применения показателя P и затрат Z на его получение как функции управляемого параметра R

Кроме управления R и показателя P в системах действует возмущение. Возмущения обозначим как $D = \{d_1, d_2, \dots\}$, см. рис. 30.13. Возмущение — это входное воздействие, которое, в отличие от управляющего параметра, не зависит от воли владельца системы. Например, низкие температуры на улице, конкуренция снижают, к сожалению, поток клиентов, поломки оборудования досадно снижают производительность системы. И управлять этими величинами непосредственно владелец системы не может. Обычно возмущение действует «назло» владельцу, снижая эффект P от управляющих усилий R . Это происходит потому, что, в общем случае, система создается для достижения целей, недостижимых самих по себе в природе. Человек, организовав систему, всегда надеется посредством ее достичь некоторой цели P . На это он затрачивает усилия R , идя наперекор природе. Система — организация доступных человеку, изученных им природных компонент для достижения некоторой новой цели, недостижимой ранее другими способами.

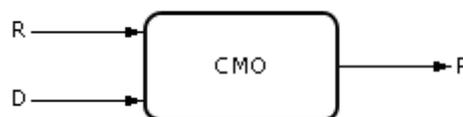


Рис. 30.13. Условное обозначение изучаемой системы, на которую воздействуют управляющие воздействия R и возмущения D

Если мы снимем зависимость показателя P от управления R еще раз (как показано на рис. 30.10), но в условиях появившегося возмущения D , то, возможно, характер кривой изменится. Скорее всего, показатель будет при одинаковых значениях управлений находиться ниже, так как возмущение носит

«противный» характер, снижая показатели системы (см. рис. 30.14). Система, предоставленная сама себе, без усилий управляющего характера, перестает обеспечивать цель, для достижения которой она была создана. Если, как и ранее, построить зависимость затрат, соотнести ее с зависимостью показателя от параметра управления, то найденная точка экстремума сместится (см. рис. 30.15) по сравнению со случаем «возмущение = 0» (см. рис. 30.12).

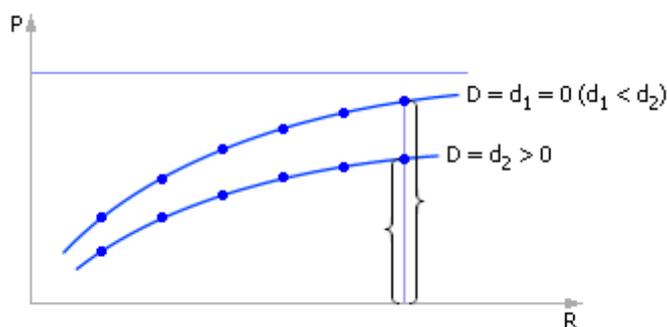


Рис. 30.14. Зависимость показателя P от управляющего параметра R при различных значениях действующих на систему возмущений D

Если снова увеличить возмущение, то кривые изменятся (см. рис. 30.14) и, как следствие, снова изменится положение точки экстремума (см. рис. 30.15).

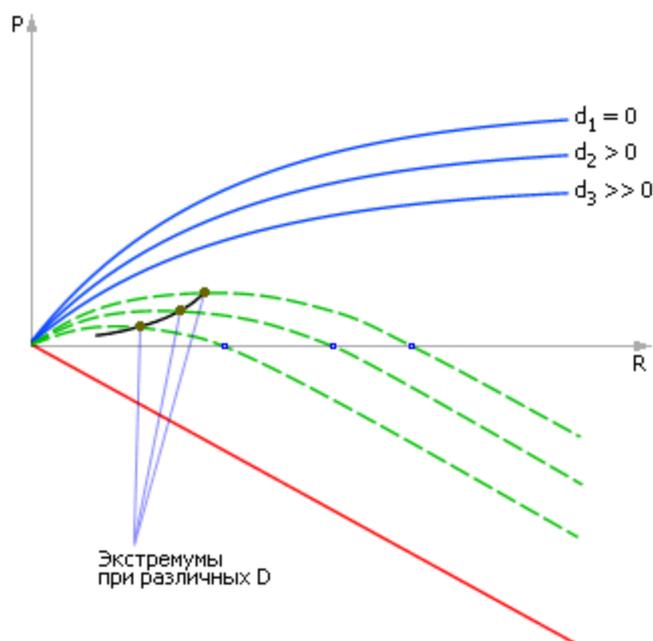


Рис. 30.15. Нахождение точки экстремума на суммарной зависимости при различных значениях действующего возмущающего фактора D

Все найденные положения точек экстремума переносятся на новый график, где образуют зависимость Показателя P от Управляющего параметра R при изменении Возмущений D (см. рис. 30.16).

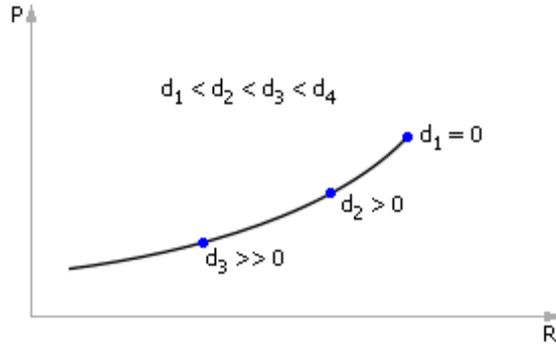


Рис. 30.16. Зависимость показателя P от управляющего параметра R при изменении значений возмущений D (кривая состоит только из точек экстремумов)

На самом деле на этом графике могут быть и другие рабочие точки (график пронизан как бы семействами кривых), но нанесенные нами точки задают такие координаты управляющего параметра, при которых при заданных возмущениях достигается наибольшее из возможных значение показателя P .

Этот график (см. рис. 30.16) связывает Показатель P , Управление (ресурс) R и Возмущение D в сложных системах, указывая, как действовать наилучшим образом ЛПР (лицу, принимающему решение) в условиях возникших возмущений. Теперь пользователь может, зная реальную обстановку на объекте (значение возмущения), быстро по графику определить, какое управляющее воздействие на объект необходимо, чтобы обеспечить наилучшее значение интересующего его показателя.

Если управляющее воздействие будет меньше оптимального, то суммарный эффект снизится, возникнет ситуация недополученной прибыли. Если управляющее воздействие будет больше оптимального, то эффект также снизится, так как заплатить за очередное увеличение управляющих усилий надо будет по величине больше, чем та, которую вы получите в результате ее использования (ситуация банкротства).

В тексте лекции мы использовали слова «управление» и «ресурс», то есть считали, что $R = U$. Управление действительно играет роль некоторой ограниченной ценности для владельца системы. То есть всегда является ценным для него ресурсом, за который всегда приходится платить, и которого всегда не хватает. Если бы эта величина не была ограничена, то мы бы могли достигать за счет бесконечной величины управлений бесконечно больших значений целей, а вот бесконечно больших результатов в природе явно не наблюдается.

Иногда различают собственно управление U и ресурс R , называя ресурсом некоторый запас, то есть границу возможного значения управляющего воздействия. В этом случае понятия ресурс и управление не совпадают: $U < R$. Иногда различают предельное значение управления $U \leq R$ и интегральный ресурс $\int U dt \leq R$.

31. МОДЕЛИРОВАНИЕ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ И СИСТЕМ

Проектирование технологических процессов, проверка свойств проекта на модели, прогнозирование хода производства, настройка модели для производственных процессов, управление производством имеют ряд особенностей.

Допустим, что некоторый производственный процесс нелинеен. Это значит, что между входом X (то, чем управляем) и выходом Y (то, что наблюдаем) имеется нелинейная зависимость. Как смоделировать такую нелинейность?

Пример 1. Логические функции. Для начала рассмотрим простую нелинейность релейного типа и допустим, что функция $Y(X)$ обратима (см. рис. 31.1). Такие зависимости называют *статическими* (значение на входе X однозначно, не зависимо от предыстории процесса определяет значение на выходе Y). Многие производственные процессы могут быть описаны таким способом [9].

Пусть зависимость состоит из трех участков (см. рис. 31.1). Часто характерной особенностью таких зависимостей является наличие в них участка, блокирующего выход, линейного участка и участка насыщения. Первый участок указывает на то, что, не подавая сырья ($X = 0$) на вход производственного процесса, нельзя надеяться на какой-либо результат ($Y = 0$). Второй участок указывает на тот факт, что, увеличивая количество сырья на входе, мы обеспечиваем увеличение готовой продукции на выходе. И, наконец, понятно, что если сырья будет очень много ($Y \gg 0$), то какие-либо ограничения производственного процесса (например, производительность оборудования, квалификация персонала, финансовые, энергетические ресурсы) все равно не позволят переработать все это сырье и выпустить соответствующее количество продукции. В сложных системах такой участок насыщения присутствует обязательно.

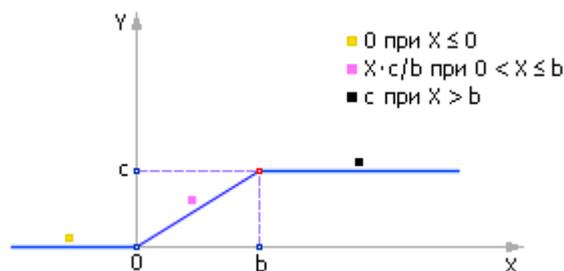


Рис. 31.1. Зависимость релейного типа

Каждый из участков можно легко описать в отдельности. Представленная на рис. 31.1 зависимость математически описывается тремя функциями. Три строки в записи означают, что в каждый отдельный момент (то есть, при определенном значении X) из описания используется одна из строк (первая,

вторая или третья). То есть подразумевается (но явно не записывается), что строки соединены знаком **ИЛИ**. Когда имеют дело с вычислительной техникой и языками программирования, где умалчивать — опасно, эту конструкцию записывают в строку:

$$Y := (0 \text{ при } X \leq 0) \text{ ИЛИ } (X \cdot c/b \text{ при } 0 < X \leq b) \text{ ИЛИ } (c \text{ при } X > b).$$

Если процессы аналоговые, то есть если мы имеем дело с любыми (рациональными) значениями X , роль операции **ИЛИ** играет знак «+»:

$$Y := (0 \text{ при } X \leq 0) + (X \cdot c/b \text{ при } 0 < X \leq b) + (c \text{ при } X > b).$$

Если условия исключают друг друга, то в любой момент времени может реализоваться только одно из трех условий. Оно и даст требуемый ответ в сумму. От остальных слагаемых в сумму придет ноль. Сложение искомого значения с нолями сформирует правильный ответ в целом.

Рассмотрим теперь отдельно одну из конструкций, например, c при $X > b$. Ее использование означает следующую фразу (которую не пишут и не произносят полностью исключительно в целях экономии места и времени): «если X больше b , то Y присвоить c ». Воспользуемся единичной функцией Хевисайда (см. рис. 31.2) для перевода данной фразы на формальный язык. Напомним, что значение единичной функции равно 0, если ее аргумент меньше или равен 0, и равно 1, если ее аргумент больше 0.

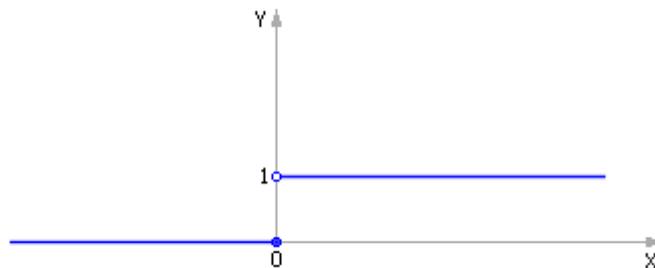


Рис. 31.2. Единичная функция Хевисайда $Y = \text{ed}(X)$

В результате получим: $Y := c \cdot \text{ed}(X - b)$. Проверим полученную формулу. Действительно:

- если $X < b$, то аргумент $(X - b)$ единичной функции отрицателен, следовательно, сама функция равна 0, поэтому $Y = 0$;
- если $X = b$, то аргумент $(X - b)$ единичной функции равен нулю, следовательно, сама функция равна 0, поэтому $Y = 0$;
- если $X > b$, то аргумент $(X - b)$ единичной функции положителен, следовательно, сама функция равна 1, поэтому $Y = c$.

В нашей формуле $Y := c \cdot \text{ed}(X - b)$ знак умножения играет роль логического **И**, поскольку для получения окончательного результата нам, во-первых, требуется значение « c » **И**, во-вторых, чтобы истиной (то есть, единцей) было значение выражения $\text{ed}(X - b)$.

Словесная логическая конструкция *если Условие, то Искомая_величина := Факт* соответствует типовой математической конструкции

$$\text{Искомая_величина} := \text{Факт} \cdot \text{ed}(\text{Условие}).$$

Неравенство превращается в единичную функцию (в математике единичная функция эквивалентна понятию **ПРЕДИКАТ**), логическое **ИЛИ** заменяется на знак сложения, логическое **И** заменяется на знак умножения. Знак « $:=$ » выражает причинно-следственную связь, увязывая причину и следствие в явном виде. В случае неявного задания функции роль этого знака будет играть знак уравнивания « $=$ ».

Теперь переведем вторую часть выражения, $Y := X \cdot c/b$ при $0 < X \leq b$, на формальный язык: $Y := X \cdot c/b \cdot \text{ed}(X) \cdot \text{ed}(b - X)$. Заметим, что правильный результат получится, если будет соблюдено уже три условия одновременно: $(X \cdot c/b)$ И $(X > 0)$ И $(X \leq b)$. Так же поступим и с оставшейся частью: $Y := 0$ при $0 \leq X$. Формальный перевод имеет вид: $Y := 0 \cdot \text{ed}(0 - X)$.

Собирая все три части вместе, имеем:

$$Y := 0 \cdot \text{ed}(0 - X) + X \cdot c/b \cdot \text{ed}(X) \cdot \text{ed}(b - X) + c \cdot \text{ed}(X - b).$$

Обратите внимание, что первое слагаемое вообще-то можно не писать, так как оно всегда равно 0. Остается:

$$Y := X \cdot c/b \cdot \text{ed}(X) \cdot \text{ed}(b - X) + c \cdot \text{ed}(X - b).$$

Если протестировать функцию в точке $X = b$, то окажется, что и первое слагаемое, и второе одновременно равны 0, а вся функция будет иметь вид, показанный на рис. 31.3. Такая ситуация называется **прощелкой**. (А есть ли щелчка между $0 \cdot \text{ed}(0 - X)$ и $X \cdot c/b \cdot \text{ed}(X) \cdot \text{ed}(b - X)$? Если «да», то при каком X ? Проверьте себя.)

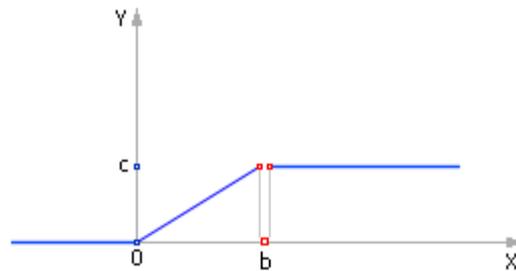


Рис. 31.3. Вид логической функции с ошибкой типа щелчки в описании

Чтобы избежать щелчек, надо писать формулы аккуратнее:

$$Y := X \cdot c/b \cdot \text{ed}(X) \cdot \text{not}(\text{ed}(X - b)) + c \cdot \text{ed}(X - b).$$

Действительно, в этом случае либо $\text{not}(\text{ed}(X - b))$ (из первого слагаемого), либо $\text{ed}(X - b)$ (из второго слагаемого) окажется обязательно равным 1, и щелчка исчезнет. Можно не вводить дополнительную операцию $\text{not}(q)$, а использовать ее аналог: $1 - \text{ed}(q)$.

Если в записи зависимости имеется n отрезков и при этом X на i -ом отрезке удовлетворяет условию: $a_i < X \leq a_{i+1}$, то имеем:

$$Y := \sum_{i=1}^n f_i(X) \cdot \text{ed}(X - a_i) \cdot \text{not}(\text{ed}(X - a_{i+1}))$$

Функция получилась однозначной, то есть одному и тому же X всегда соответствует одно и то же Y . Удобство таких описаний в том, что Y вычисляется в любой момент для любого X , достаточно подставить нужное любое значение X в формулу.

Вид записи зависит от системы, которой придется исполнять такую запись. Если система не понимает запись, то она не сможет ее исполнить, и такая запись бесполезна. Язык такой мертв и ничего не означает, так как не служит средством коммуникации, передачи информации от одного агента другому. К примеру, если исполнителем является алгоритмическая машина, то конструкция

$$Y := X \cdot c/b \cdot \text{ed}(X) \cdot \text{not}(\text{ed}(X - b)) + c \cdot \text{ed}(X - b).$$

будет иметь вид, показанный на рис. 31.4.

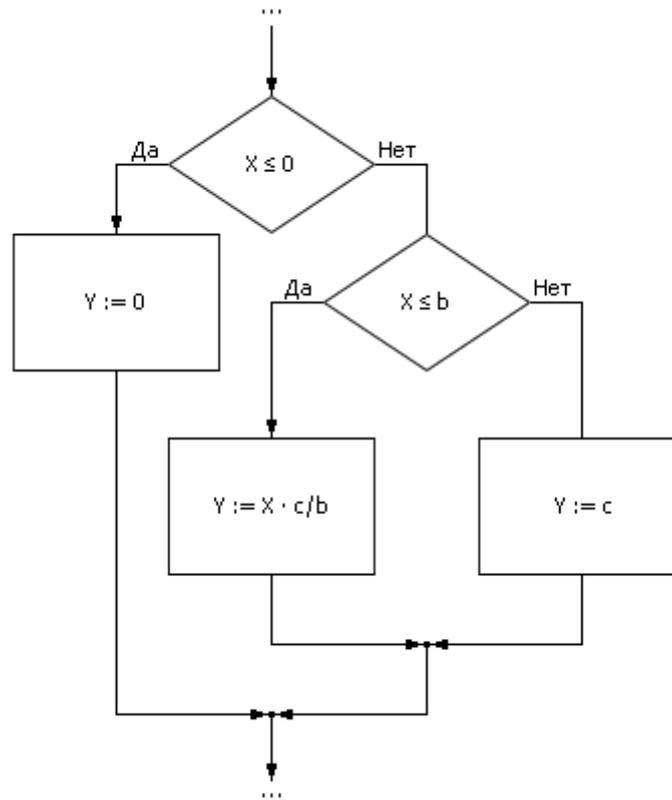


Рис. 31.4. Алгоритмическая реализация производственных моделей в виде логических функций с использованием условных конструкций

Есть соответствие в языках описания. Условная конструкция в алгоритмах соответствует «многоэтажной» формуле в математике или конструкции

$$\text{Искомая_величина} := \sum_{i=1}^n \left(\text{Факт}_i \cdot \prod_{j=1}^k \text{ed}(\text{Условие}_j) \right)$$

в формализованном языке моделирования. Или, с использованием логических функций **И**, **ИЛИ**, **НЕ**:

$$\text{Искомая_величина} := \text{ИЛИ}_{i=1}^n \left(\text{Факт}_i \cdot \text{И_если}_{j=1}^k (\text{Условие}_j) \right)$$

Пример 2. Генератор пилообразного сигнала (ГПС). Сконструируем периодическую пилообразную функцию (см. рис. 31.5) двумя различными способами.

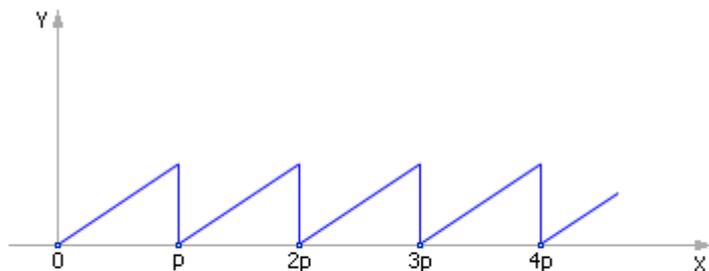


Рис. 31.5. Вид зависимости $Y(X)$ для ГПС

Вариант 2.1. ГПС на основе математической функции mod . Известная математическая функция $\text{mod}(X, a)$ возвращает нам остаток от деления X на a , а остаток, как известно, растет линейно с ростом X , а затем становится равным 0 (когда X и a становятся кратными), повторяя эту закономерность периодически. В среде Stratum-2000 это запишется так: $Y := X \% a$. Период и амплитуда «пилы» регулируются значениями X и a .

Вариант 2.2. ГПС на основе уравнения с памятью. $Y := (Y + d) \cdot \text{ed}(p - Y)$, где d — шаг изменения, p — период. Допустим, что сначала Y достаточно мало, поэтому $p > Y$. Если $p - Y > 0$, то для положительного аргумента единичная функция возвращает 1, и выражение имеет вид: $Y := (Y + d) \cdot 1$, то есть Y увеличивается на каждом такте на величину d , а значит — линейно растет.

Рано или поздно Y становится равно или больше p , и условие становится $p - Y \leq 0$. Для отрицательного и нулевого аргумента единичная функция возвращает 0, а значит выражение в этот момент принимает вид: $Y := (Y + d) \cdot 0$, то есть Y на данном шаге сбрасывается в ноль, и ситуация повторяется: снова Y становится малым и начинает расти, пока не достигнет значения p . При подаче на вертикальную координату осциллографа вычисленного значения Y , а на горизонтальную его координату значения X , сформированного как $X := X + 1$, можно увидеть «пилу» с периодом p (см. рис. 31.6).

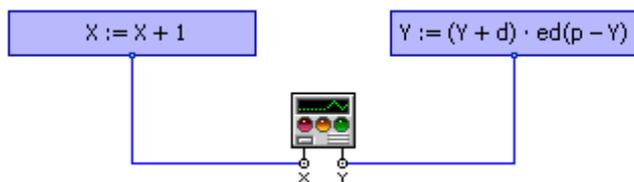


Рис. 31.6. Схема проекта в среде Stratum-2000, реализующего генератор пилообразного сигнала на основе уравнения с памятью

Во-первых, как видите, в выражении $Y := (Y + d) \cdot \text{ed}(p - Y)$, чтобы вычислить очередное значение Y , надо знать предыдущее. Поэтому говорят, что это **уравнение с обратной связью** или **уравнение с памятью**: чтобы вычислить N -е значение, требуется вычислять все предыдущие значения функции на ряде шагов. Во-вторых, функция получилась неоднозначная. То есть при разных начальных данных график функции будет различный. Одному и тому же X могут соответствовать разные Y . Это соответствует семейству графиков, порождаемых дифференциальным уравнением. Данная запись **НЕЯВНАЯ**,

так как она не содержит значение X непосредственно. Поэтому нельзя подставить произвольное значение X и тут же узнать, чему будет равно Y при этом значении. Чтобы определить значение Y , необходимо пройти (перебрать) все значения Y от 0 до искомого.

Первый вариант записи, $Y := X\%a$, представляет нам явную зависимость Y от X . В любой момент можно подставить любое значение X и получить соответствующее ему значение Y .

Каждый вариант записи имеет свои преимущества и недостатки. Не всегда удастся получить явную запись, так как записать закон функционирования системы обычно намного проще, чем найти его решение. Но и явный вариант имеет тот недостаток, что это только одно из решений, тогда как сам закон содержит потенциально все множество возможных других решений, многозначен.

Если реализовывать выражение $Y := (Y + d) \cdot \text{ed}(p - Y)$ на языке алгоритмов, то понадобится циклическая конструкция (см. рис. 31.7).

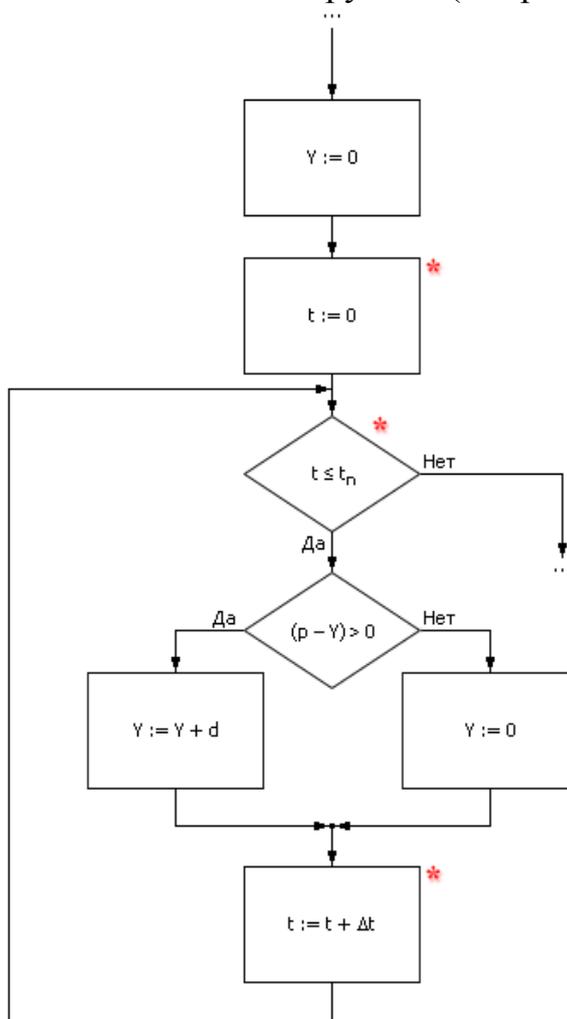


Рис. 31.7. Алгоритмическая реализация модели пилообразного сигнала с использованием циклической конструкции

Между системой с памятью и циклической конструкцией есть глубокая связь, система с памятью — это *динамическая система*, что соответствует в математике *дифференциальному уравнению*, пошаговый расчет которого на компьютере требует *циклического алгоритма*. Если решение дифференци-

ального уравнения можно найти аналитически, то есть выразить искомую величину в зависимости от остальных явно, то циклической конструкции не понадобится, достаточно алгоритма линейной структуры. Но это будет собственно не уравнение, а присвоение, то есть решение уравнения, точнее, одно из решений уравнения.

Пример 3. Генератор прямоугольных импульсов (ГПИ). Во многих приложениях требуется имитировать периодическое открывание и закрывание на некоторое время каких-то устройств, вентилях, каналов. Для этого удобно использовать генератор импульсов, который бы на выходе выдавал значения «0» и «1», то есть **генератор прямоугольных импульсов**. За параметры генератора удобно принять частоту изменений с 0 на 1 и обратно и скважность — соотношение времени удержания на выходе «1» ко времени всего цикла.

Вариант 3.1. ГПИ на основе периодической функции синус. Можно использовать несколько приемов моделирования генератора прямоугольных импульсов, имеющих свои достоинства и недостатки. Рассмотрим вначале генератор, использующий свойства периодической функции, например, математической функции синус (см. рис. 31.8).

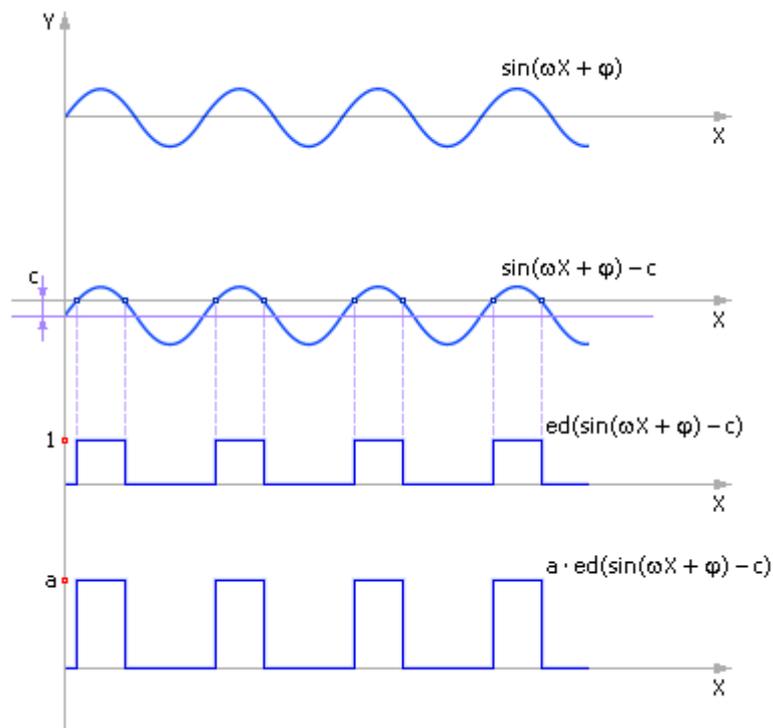


Рис. 31.8. Генерация последовательности прямоугольных импульсов средствами периодической функции

$Y := a \cdot \text{ed}(\sin(\omega \cdot X + \varphi) - c)$, где ω — частота генератора, задающая его период; φ — фаза генератора; a — амплитуда импульсов; c — задаваемый параметр, скважность. Скважность указывает на то, в течение какой части периода синуса Y будет равен a , и в течение какой части периода Y будет равен 0. Пока $\sin(\omega \cdot X + \varphi) > c$, единичная функция возвращает 1, $Y = a$; как только $\sin(\omega \cdot X + \varphi) \leq c$, единичная функция будет возвращать 0, $Y = 0$.

Явная запись, что определяет достоинства и недостатки этого варианта.

Вариант 3.2. ГПИ на основе уравнения с памятью. Другой вариант — использовать для генерации прямоугольных импульсов ранее созданный генератор пилообразного сигнала, который использовал последовательное вычисление значений от точки к точке (уравнение с памятью).

$Y := (Y + d) \cdot \text{ed}(p - Y)$, $Z := a \cdot \text{ed}(Y - c)$, где d — шаг изменения; p — период; a — амплитуда импульсов; c — задаваемый параметр, скважность.

Поскольку Y генерирует пилообразный сигнал, то достаточно сравнивать его значение с заданным значением c . Если $Y - c \leq 0$, то $Z = 0$, если $Y - c > 0$, то $Z = a \cdot 1$. В результате получается периодический единичный сигнал. При подаче на осциллограф на вертикальную координату вычисленного значения Z , а на горизонтальную координату — значения X , сформированного как $X := X + 1$, можно увидеть «забор» с периодом p (см. рис. 31.9).

Данная запись имеет неявный вид, что определяет достоинства и недостатки этого варианта.

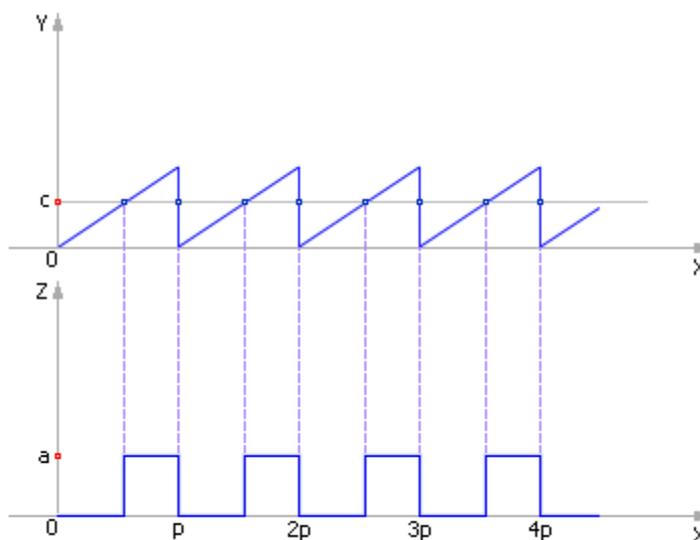


Рис. 31.9. Генерация последовательности прямоугольных импульсов с использованием выражения с памятью

Пример 4. Коммутатор. Часто используемым в производственных процессах технологическим устройством является коммутатор, целью которого является распределение какого-нибудь основного потока (материального, энергетического, информационного) на ряд частных потоков за счет переключения подачи с одного направления на другое (см. рис. 31.10). Это позволяет делить потоки, менять их направление.

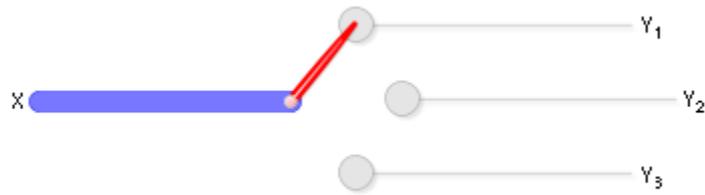


Рис. 31.10. Схема коммутатора

Пусть имеется один вход, помеченный как X , и, например, три выхода: Y_1 , Y_2 , Y_3 . Требуется сигнал X в первый момент времени транслировать на Y_1 , во второй момент времени — передавать на Y_2 , а в третий момент — на Y_3 , после чего циклически повторять этот процесс, начиная снова с Y_1 . (Естественно, возможны и другие варианты коммутаторов.) Временная диаграмма распределения основного потока на подпотоки дискретным коммутатором показана на рис. 31.11.

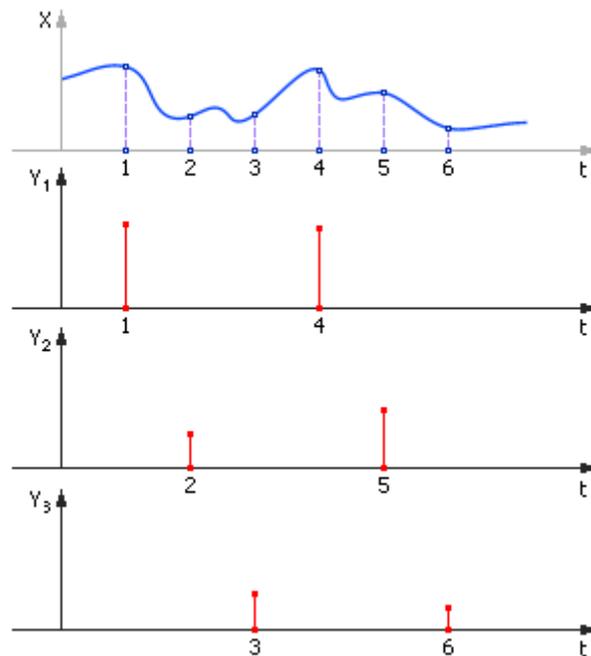


Рис. 31.11. Временная диаграмма работы дискретного коммутатора с одним входом и тремя выходами

Для этого понадобится дельта-функция Дирака: $Y = \delta(X)$. Эта функция возвращает 1, если ее аргумент равен 0, в остальных случаях она возвращает ноль (см. рис. 31.12).

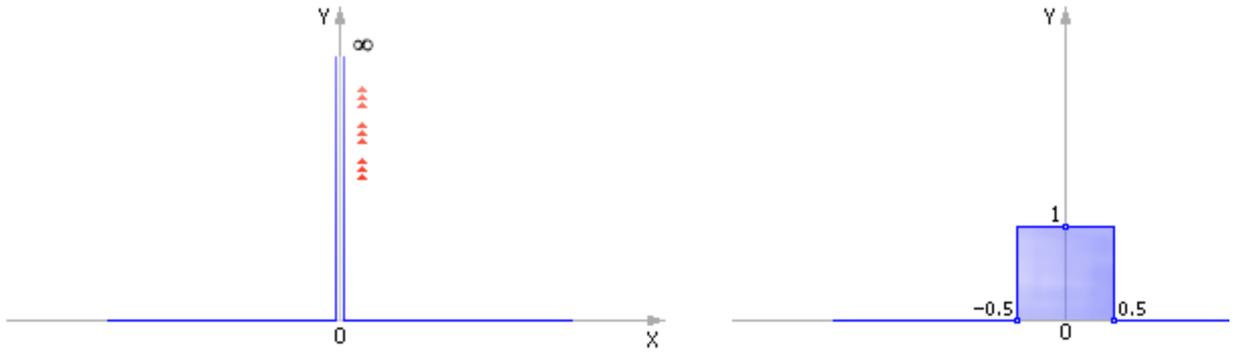


Рис. 31.12. Дельта-функция Дирака $Y = \delta(X)$ (слева — теоретический вид, справа — ее приближенный дискретный аналог)

С помощью дельта-функции выражения, имитирующие коммутатор, будут выглядеть так (в приведенной ниже записи переменная i играет роль времени t):

$$i := (i + 1) \cdot \text{ed}(2 - i)$$

$$Y_1 := X \cdot \delta(i - 0)$$

$$Y_2 := X \cdot \delta(i - 1)$$

$$Y_3 := X \cdot \delta(i - 2).$$

i , благодаря первому выражению, пробегает циклически значения 0, 1, 2, 0, 1, 2, 0, 1, 2, ... Если i оказалось равно 0, то $\delta(i - 0) = 1$ и $Y_1 := X \cdot 1$; при этом $\delta(i - 1) = 0$ и $Y_2 := X \cdot 0$ и $\delta(i - 2) = 0$ и $Y_3 := X \cdot 0$.

Если i оказалось равно 1, то $\delta(i - 1) = 1$ и $Y_2 := X \cdot 1$, а $\delta(i - 0) = 0$ и $Y_1 := X \cdot 0$ и $\delta(i - 2) = 0$ и $Y_3 := X \cdot 0$.

Если i оказалось равно 2, то $\delta(i - 2) = 1$ и $Y_3 := X \cdot 1$, а $\delta(i - 0) = 0$ и $Y_1 := X \cdot 0$ и $\delta(i - 1) = 0$ и $Y_2 := X \cdot 0$.

Для организации циклического коммутатора нам понадобилась «пила», которую мы построили ранее.

Лучше было бы сделать так: отдельно записать выражения для системы управления, где описываются информационные сигналы i, f_1, f_2, f_3 :

$$i := (i + 1) \cdot \text{ed}(2 - i);$$

$$f_1 := \delta(i - 0);$$

$$f_2 := \delta(i - 1);$$

$$f_3 := \delta(i - 2).$$

Флаги f_i принимают значения 0 и 1, как это и положено логическим переменным. Модель информационной части проекта показана на рис. 31.13.

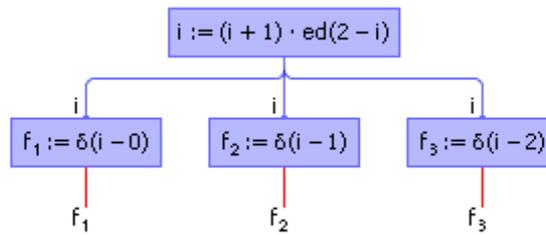


Рис. 31.13. Схема информационно-управляющих сигналов для коммутатора

Отдельно записать выражения для материальных потоков Y и X , которые перекрываются сигналами управления f_i (см. рис. 31.14):

$$Y_1 := X \cdot f_1;$$

$$Y_2 := X \cdot f_2;$$

$$Y_3 := X \cdot f_3.$$

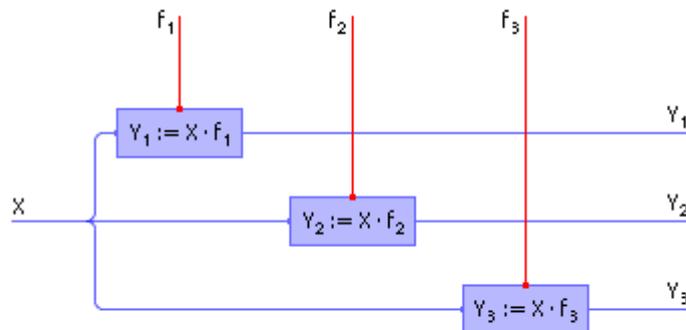


Рис. 31.14. Схема проекта, выполненного в среде Stratum-2000, реализующая коммутатор. Отражена часть проекта без формирователя флагов

Переменные Y , X принимают действительные значения, как это и положено переменным, которые имитируют материальные потоки. Можно представить себе аналогию: поток X переходит в поток Y , если кран f открыт ($f = 1$), или, наоборот, значения потока X не переходят на переменную Y , так как кран закрыт ($f = 0$). Такая схема в дальнейшем более предпочтительна, так как отделяет материальную систему от информационной, делая проект более понятным и прозрачным (см. также пример 5 ниже).

Объединим информационную схему со схемой, имитирующей материальные потоки (см. рис. 31.15).

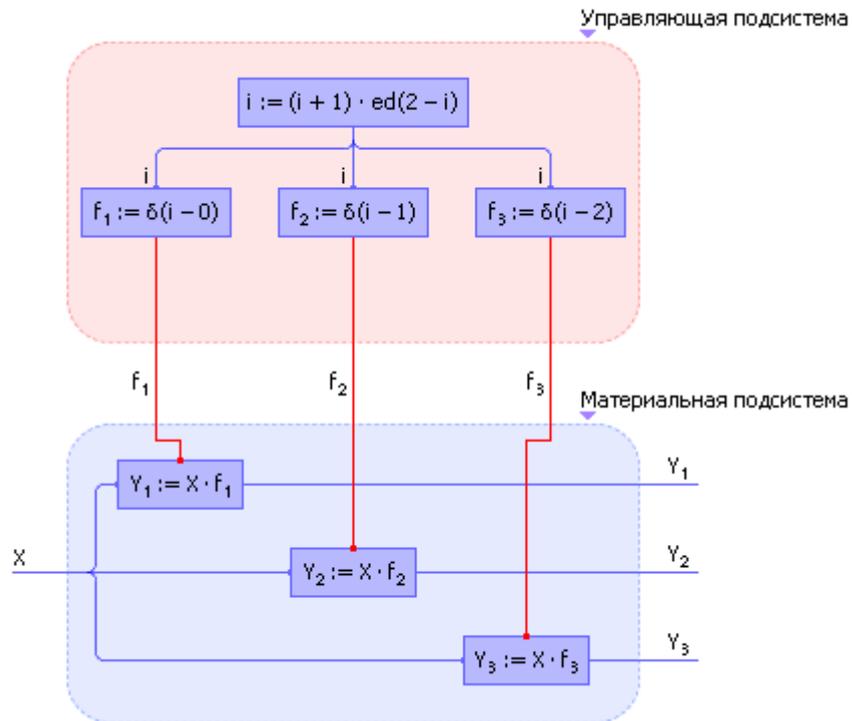


Рис. 31.15. Схема проекта «Модель коммутатора», представленного композицией информационной и материальной подсистем. Проект выполнен в среде Stratum-2000

На рис. 31.16 показано взаимодействие двух подсистем (информационной и материальной) между собой в общем виде, которого рекомендуется достигать в идеале для прозрачности процесса проектирования. Все зависимости и законы, касающиеся описания свойств материи и энергии, рекомендуется включать в состав модели материальных потоков (иногда в этой подсистеме дополнительно разделяют энергетические и материальные потоки), а законы измерения, управления, регулирования, формирования целей — включать в состав модели информационных сигналов.

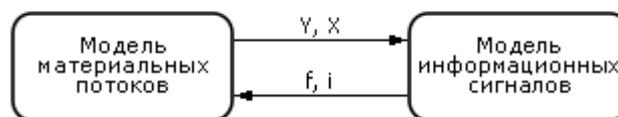


Рис. 31.16. Разделение материальной и управляющей моделей на отдельные подсистемы

Пример 5. Транспортировка изделий с одного узла на другой. Частый случай в производстве — организация транспортировки изделий с одного производственного узла на другой. Под **узлом** здесь будем понимать совокупность склада и органа, осуществляющего транспортировку (например, транспортное средство, канал). Склад хранит изделия, количество которых может уменьшаться или увеличиваться, если транспортный орган берет на себя их перемещение с узла на узел.

Допустим, что на первом узле имеется X_1 изделий, а на втором — X_2 изделий. Обозначим как U количество транспортируемых изделий за такт. То-

гда:

$$X_1 := X_1 - U \cdot f;$$

$$X_2 := X_2 + U \cdot f.$$

На рис. 31.17 показана связь двух узлов транспортным каналом, осуществляющим перекачку изделий со склада одного узла на склад другого узла с производительностью U штук в единицу времени.

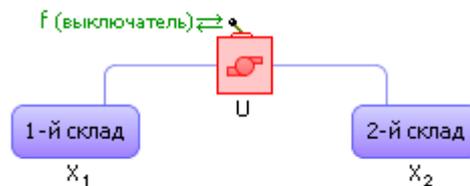


Рис. 31.17. Условное изображение модуля транспортировки с одного склада на другой

Если флаг f равен 1, то U изделий уходит с первого узла и столько же в этот момент приходит на второй узел. Если флаг f равен 0, то переброски изделий не происходит, то есть они не вычитаются в первом выражении и не складываются во втором, а X_1 и X_2 остаются неизменными величинами.

Если изделия необходимо транспортировать через цепь узлов (см. рис. 31.18), то, обозначая объемы перевозок между узлами U_1, U_2, \dots, U_{n-1} , имеем:

$$X_1 := X_1 - U_1 \cdot f;$$

$$X_2 := X_2 + U_1 \cdot f - U_2 \cdot f;$$

$$X_3 := X_3 + U_2 \cdot f - U_3 \cdot f;$$

...

$$X_n := X_n + U_{n-1} \cdot f.$$

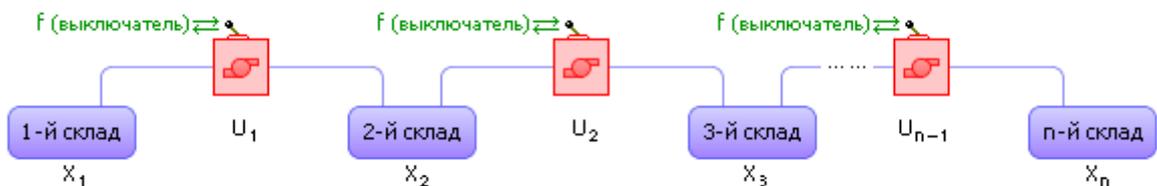


Рис. 31.18. Схема транспортировки изделий через цепь складов

Если задвижками f необходимо управлять по отдельности, то логично ввести переменные f_i , управляя задвижкой в каждом узле отдельно.

Следует заметить, что данную модель для улучшения ее адекватности можно модернизировать:

$$X_1 := X_1 - U_1 \cdot f \cdot \text{ed}(X_1);$$

$$X_2 := X_2 + U_1 \cdot f \cdot \text{ed}(X_1) - U_2 \cdot f \cdot \text{ed}(X_2);$$

$$X_3 := X_3 + U_2 \cdot f \cdot \text{ed}(X_2) - U_3 \cdot f \cdot \text{ed}(X_3);$$

$$\dots; \\ X_n := X_n + U_{n-1} \cdot f \cdot \text{ed}(X_{n-1}).$$

Выражение $\text{ed}(X_i)$ играет роль своеобразного флага разрешения транспортировки, запрещая ее в случае, если в узле нет изделий. Если такого флага нет, то возможны странные варианты переброски фиктивных изделий с узла на узел, например, появление отрицательных изделий.

Еще одно замечание. На цифровой технике уравнения решаются с некоторым тактом. Если в течение такта окажется, что U_i больше X_i , то за такт из узла исчезнет больше изделий, чем в нем было, и количество изделий станет меньше нуля.

Во-первых, на это можно не обращать внимания, поскольку это погрешность численного расчета. Второй вариант — предотвратить все-таки такие коллизии. Для этого есть два способа. Первый состоит в том, чтобы уменьшить время dt такта расчета (до этого мы принимали его равным 1): $X_1 := X_1 - U_1 \cdot f \cdot \text{ed}(X_1) \cdot dt$. Второй способ: вычитать и прибавлять значение U_i , если $U_i < X_i$, или вычитать и прибавлять X_i , если $U_i \geq X_i$, то есть вычитать и прибавлять минимальное из двух чисел U_i и X_i : $X_1 := X_1 - U_1 \cdot f \cdot \text{ed}(X_1) \cdot \text{ed}(X_1 - U_1) - X_1 \cdot f \cdot \text{ed}(X_1) \cdot \text{not}(\text{ed}(X_1 - U_1))$. В данной записи выражение $\text{ed}(X_i)$, играющее роль своеобразного флага разрешения транспортировки, становится излишним, поэтому его можно опустить: $X_1 := X_1 - U_1 \cdot f \cdot \text{ed}(X_1 - U_1) - X_1 \cdot f \cdot \text{not}(\text{ed}(X_1 - U_1))$. С учетом всего вышесказанного, напомним выражение для

$$X_2: \\ X_2 := X_2 + U_1 \cdot f \cdot \text{ed}(X_1 - U_1) + X_1 \cdot f \cdot \text{not}(\text{ed}(X_1 - U_1)) - U_2 \cdot f \cdot \text{ed}(X_2 - U_2) - X_2 \cdot f \cdot \text{not}(\text{ed}(X_2 - U_2)).$$

Если представить себе, что мы имеем дело с производственной технологической линией, то в качестве складов в этой цепочке выступают запасы незавершенной продукции (промежуточные склады в производстве), а роль транспортных артерий — станки, которые обрабатывают изделия, тем самым как бы «проталкивая» их с одного участка на другой.

Тогда U играет в этих выражениях роль производительности станков: чем больше U , тем больше изделий за одну единицу времени переходит с одного склада на другой, а X играет роль количества изделий на промежуточных складах в производстве. Таким образом, U характеризует динамику производственного процесса, а множество X — его статику. То есть X характеризует, сколько изделий находится в производстве в разной стадии готовности, а U — сколько изделий и с какой скоростью обрабатывается в каждый момент времени, как быстро перетекают изделия с узла на узел. Задавая различные значения U , можно имитировать различные производственные ситуации.

Иногда при описании производства задают закон движения U . Например, $U = k \cdot X$. Логика действия такого закона состоит в том, что чем больше изделий X скапливается на складе, тем больше должна быть интенсивность их обработки U , чтобы избежать чрезмерного накопления изделий на промежуточных складах. Как ранее было описано, закон движения, в отличие от

уравнений материального баланса, представляет собой информационный поток, это хорошо видно на рис. 31.19.

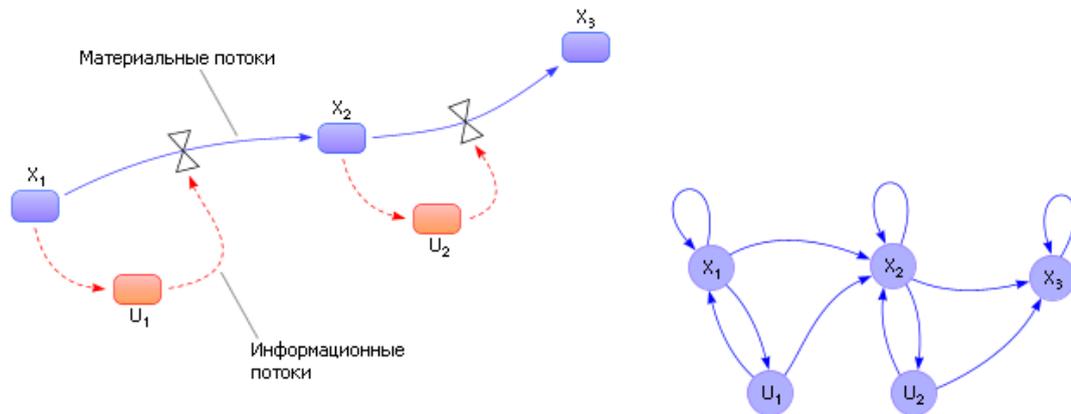


Рис. 31.19. Два варианта изображения схемы модели технологической линии с информационными и материальными связями с различным уровнем детализации. Слева — обозначения по Форрестеру с разделением материальных и информационных потоков, справа — граф зависимостей переменных

Можно задать и другие законы, например, $U_i := U_i + k \cdot (X_i - X_{i+1})$. Логика этого закона состоит в том, что если на $i + 1$ -ом складе затоваривание ($X_{i+1} > X_i$), то надо притормаживать U_i , если недостаток, то есть $X_{i+1} < X_i$, то разгонять производительность данного узла.

Измерительная часть системы управления может фиксировать общее количество изделий P_1 на всех складах, их суммарную стоимость P_2 , общую производительность станков P_3 , неравномерность распределения изделий в производстве P_4 , выполнение плана X_z — P_5 и так далее:

$$P_1 = X_1 + X_2 + X_3 + \dots + X_n;$$

$$P_2 = X_1 \cdot c_1 + X_2 \cdot c_2 + X_3 \cdot c_3 + \dots + X_n \cdot c_n;$$

$$P_3 = U_1 + U_2 + U_3 + \dots + U_n;$$

$$P_4 = \text{abs}(X_1 - X_2) + \text{abs}(X_2 - X_3) + \dots + \text{abs}(X_{n-1} - X_n);$$

$$P_5 = X_n - X_z.$$

По данным показателям можно настраивать регуляторы в производстве, добиваясь приемлемых значений у этих показателей.

С помощью такого подхода, используя логические комбинации (**И**, **ИЛИ**, **НЕ**) и условия, отражаемые единичными функциями, можно описать разделения, слияния потоков, переход партий, сборку, брак, помехи, задержки, переход изделий с линии на линию, и вообще любую производственную топологию.

Задачи, которые можно решить на такой модели:

- какие регуляторы подобрать для того, чтобы обеспечить интересующее значение показателей P_1 — P_5 ;
- за какое время можно выполнить заданный план, какие управления U для этого следует задать;
- как перейти от одной картины заполнения складов ($X_1, X_2, X_3, \dots, X_n$) в производстве к другой ($X_1^*, X_2^*, X_3^*, \dots, X_n^*$), изменяя множество U ;

- компенсируется ли и как регуляторами случайное возмущение типа «брак»;
- какой уровень запасов изделий следует иметь в производстве, чтобы исключить дефицит;
- насколько можно остановить какой-то из станков (поломка), чтобы не сорвать план;
- какие управляющие величины U следует применить при заданном заполнении складов ($X_1, X_2, X_3, \dots, X_n$), чтобы обеспечить равномерность движения и распределения изделий по технологической линии.

Пример 6. Обработка партии изделий в термической печи. Рассмотрим обработку партии изделий в термической печи. Если температура в печи до 50 градусов, то партия снова поступает в обработку. Если температура в печи между 50 и 100 градусами, то партия уходит на склад готовых изделий. Если температура в печи больше 100 градусов, то партия поступает в брак.

Блок информационных сигналов имеет вид:

$$f_1 := \text{ed}(50 - T);$$

$$f_2 := \text{not}(\text{ed}(50 - T)) \cdot \text{ed}(100 - T);$$

$$f_3 := \text{not}(\text{ed}(100 - T)).$$

Блок имитации материальных процессов:

$$\text{Обр} := \text{Партия} \cdot f_1;$$

$$\text{Готовые} := \text{Готовые} + \text{Партия} \cdot f_2;$$

$$\text{Брак} := \text{Брак} + \text{Партия} \cdot f_3.$$

Обратите внимание: на схеме модели (рис. 31.20) отдельно сформированы все возможные и существенные для процесса условия «ed(Условие)», отдельно сформированы информационные сигналы (f_1, f_2, f_3), отдельно описаны материальные потоки.

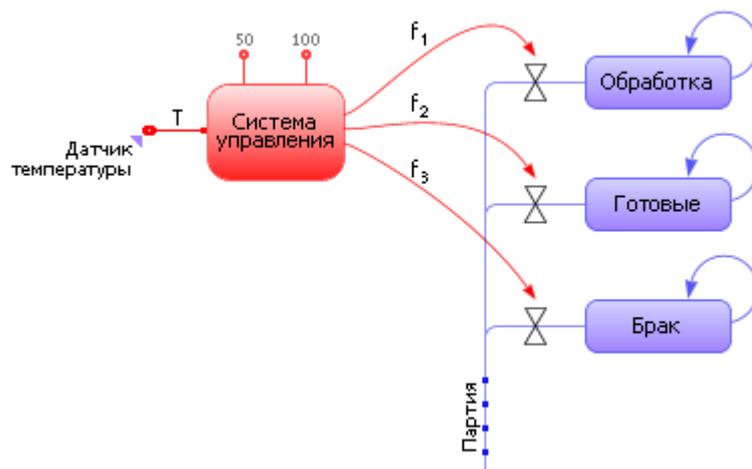


Рис. 31.20. Схема модели «Обработка изделий в термической печи»

Пример 7. Погрузка железнодорожного состава. Рассмотрим процесс погрузки железнодорожного состава, состоящего из электровоза и прицепленных к нему трех вагонов (см. рис. 31.21). Над первым вагоном находится бункер, из которого насыпается груз. Вагон стоит на весах, измеряющих массу насыпанного груза. Грузоподъемность вагонов задана и известна системе

управления из справочника в базе данных АСУ. Система управления (СУ) должна определить момент (и выдать управляющий сигнал электровозу), когда следует поставить под загрузку следующий вагон. После загрузки последнего вагона следует закрыть заслонку бункера.

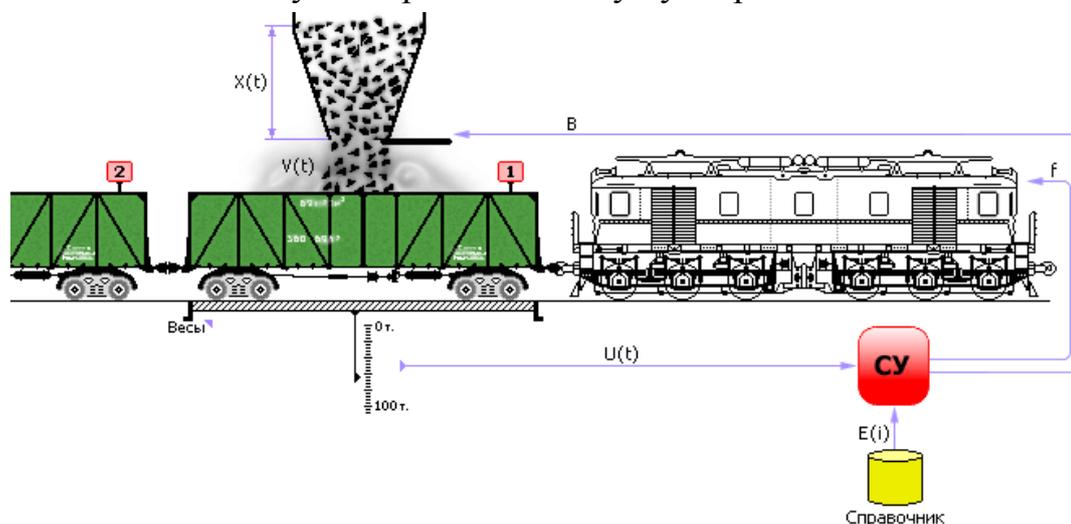


Рис. 31.21. Погрузка железнодорожного состава и система управления ею

Пусть $U(t)$ — текущая загрузка вагона, значение на весах; $V(t)$ — интенсивность поступления груза в вагон из бункера, количество за такт; i — номер вагона (1, 2, 3); $E(i)$ — грузоподъемность i -го вагона (задана); $X(t)$ — уровень сырья в бункере; f — сигнал машинисту электровоза: «продернуть состав на один вагон вперед».

Выражения, моделирующие погрузку и реакции системы управления, должны имитировать процесс погрузки (законы изменения $U(t)$, $X(t)$ и $V(t)$), сигналы АСУ (f и i).

$U(t) := (U(t) + V(t)) \cdot \text{ed}(E(i) - U(t))$ — загрузка вагона увеличивается, если вагон не полон, и останавливается и сбрасывается в ноль, если вагон загружен до нормы $E(i)$;

$f := \text{not}(\text{ed}(E(i) - U(t)))$ — сигнал машинисту электровоза: «продерни состав»;

$i := i + f$ — счетчик вагонов увеличивается в момент продергивания состава;

$X(t) := X(t) - V(t)$ — уровень сырья в бункере понижается;

$a := k \cdot X(t)$ — чем меньше уровень X , тем медленнее высыпается сырье, так как верхние слои меньше давят на нижние;

$V(t) := a \cdot B$ — интенсивность поступления груза в вагон из бункера;

$B := \text{ed}(3 - i + 1) \cdot \text{ed}(X(t))$ — сырье поступает в вагон, если его номер не больше третьего **И** сырье в бункере есть.

На основе построенной модели можно решить ряд задач, например, определить время, за которое будет загружен весь состав, или выяснить другие какие-то другие вопросы. В общем случае, то, какая задача будет решаться на модели, во многом зависит от целей человека, использующего эту модель.

Пример 8. Триггер. Часто для управления информационными сигналами требуется триггер. Таблица истинности триггера представлена в табл. 31.1.

Таблица 31.1.
Таблица истинности триггера

Входы			Новое состояние триггера, выход
Сброс триггера	Старое состояние	Информационный вход	
R	Q	S	Q
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	0

Это устройство, как видно из табл. 31.1, легко описывается логическими функциями, но при этом одновременно является системой с памятью (внутри триггера находится цепь обратной связи, математически: $Q := F(R, S, Q)$, см. рис. 31.22).

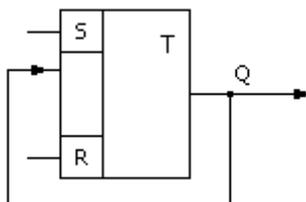


Рис. 31.22. Условное обозначение триггера на схемах

Представим модель триггера, используя логические функции его состояния. В табл. 31.1 выделим все строки с $Q = 1$ (столбец «Новое состояние триггера, выход»). Их три. Запишем три ситуации через операцию **ИЛИ**: $Q := c_1$ ИЛИ c_2 ИЛИ c_3 . Каждая ситуация, чтобы строка выдала 1 на выходе, должна быть одновременной комбинацией сигналов R , Q , S , то есть R И Q И S . Если входной сигнал в строке равен 1, то переменную надо взять «как есть», если равен 0, то инвертировать функцией НЕ. Например, строка

R	Q	S	Q
0	0	1	1

дает: $\text{НЕ}(R)$ И $\text{НЕ}(Q)$ И S И Q , или, используя функции в записи, как их обычно реализуют в формальных языках, имеем: $\text{not}(R) \cdot \text{not}(Q) \cdot S \cdot Q$, или еще вариант: $\text{and}(\text{and}(\text{not}(R), \text{not}(Q)), \text{and}(S, Q))$. Вся таблица в виде логической функции опишется так:

$Q := (\text{not}(R) \cdot \text{not}(Q) \cdot S) \text{ ИЛИ } (\text{not}(R) \cdot Q \cdot \text{not}(S)) \text{ ИЛИ } (\text{not}(R) \cdot Q \cdot S).$

В табл. 31.2 для справки приведены значения общеупотребимых логических функций **И**, **ИЛИ**, **НЕ**. Окончательно, приводя подобные, вынося за скобки и используя формальную общепринятую языковую запись, получим:

$$Q := \text{and}(\text{not}(R), \text{or}(S, \text{and}(Q, \text{not}(S)))).$$

Таблица 31.2.
Значения логических функций И, ИЛИ, НЕ

A	B	A and B	A or B	not A
0	0	0	0	1
0	1	0	1	1
1	0	0	1	0
1	1	1	1	0

Пример 9. СТАРТ-СТОП или Ловушка. Теперь допустим, что надо смоделировать такой процесс, который начинается при выполнении какого-то сигнала Y , затем идет какое-то время T_p и останавливается сам по себе по прохождении такого времени. Если во время выполнения процесса поступает снова сигнал Y , то процесс игнорирует такой сигнал Y . Если сигнал Y приходит вновь, когда процесс остановлен, то процесс готов воспринять сигнал и отреагировать на него снова.

Ловушкой такая конструкция называется потому, что модель на время захватывает входной сигнал и не реагирует в течение этого времени на другие входные сигналы. С помощью таких конструкций часто имитируют процесс обработки изделий в отдельных узлах технологической линии. Изделие попадает в узел, как в ловушку, и выходит из него только через определенное время. В течение всего этого времени узел занят и не готов принять другие изделия (входные сигналы). Получается, что и изделие занято узлом, и узел занят изделием.

Назовем для наглядности флаги: «Можно_начать» — флаг устанавливается в 1, если можно начать процесс, а процесс можно начать, если пришел сигнал Y **И** процесс сейчас не идет; «Процесс_идет» — флаг устанавливается в 1, если процесс происходит, то есть уже начался **И** пока не закончился. Процесс идет, если можно начать процесс **ИЛИ** если процесс уже идет **И** пока не остановлен; «Стоп» — флаг устанавливается в 1, если надо остановить процесс, а это происходит, если процесс уже или пока не идет **ИЛИ** время протекания процесса окончилось.

Теперь запишем уравнения согласно логике (в записи использованы знаки $\&$ (И) и $|$ (ИЛИ)):

$$\text{Можно_начать} := \text{ed}(Y) \& \text{not}(\text{Процесс_идет});$$

$$\text{Процесс_идет} := \text{Можно_начать} | (\text{Процесс_идет} \& \text{not}(\text{Стоп}));$$

$$t := (t + h) \cdot \text{ed}(T_p - t) \cdot \text{Процесс_идет};$$

$$\text{Стоп} := \text{not}(\text{Процесс_идет}) | \text{not}(\text{ed}(t)).$$

32. ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ МОДЕЛИРУЮЩИХ АЛГОРИТМОВ

При реализации моделирующих алгоритмов на цифровых машинах есть определенные общие закономерности, которые мы обсудим в данной лекции.

Основная проблема при составлении алгоритмов на машине с последовательной обработкой процессов состоит в том, что при моделировании необходимо отслеживать множество процессов, которые в реальном времени происходят параллельно. В связи с этим алгоритмы моделирования имеют свои особенности:

- продвижение системы во времени, отслеживание временной координаты;
- обеспечение синхронной работы объектов, из которых состоит моделируемая система.

В настоящий момент известны четыре основных принципа регламентации событий.

1. Принцип Δt («дельта-тэ»).
2. Принцип особых состояний.
3. Принцип последовательной проводки заявок.
4. Принцип параллельной работы объектов (объектный принцип моделирования).

Рассмотрим на примерах, как реализуется в моделирующих алгоритмах каждый принцип по отдельности.

Принцип Δt

Принцип состоит в том, что алгоритмом моделирования имитируется движение, то есть изменение состояния системы, в фиксированные моменты времени: $t, t + \Delta t, t + 2\Delta t, t + 3\Delta t, \dots$

Для этого заводится счетчик времени (часы), который на каждом цикле увеличивает свое значение t на величину шага во времени Δt , начиная с нуля (начало моделирования). Таким образом, изменения системы отслеживаются такт за тактом в заданные моменты: $t, t + \Delta t, t + 2\Delta t, t + 3\Delta t, \dots$

Особенности реализации принципа Δt

Это **наиболее универсальный** из рассматриваемых принципов, так как применяется для очень широкого класса систем. Он же является **наиболее простым** в реализации, поскольку принцип Δt совпадает с пониманием человека о времени, как о последовательном явлении, текущем с постоянным темпом.

Однако это **самый неэкономичный** принцип, поскольку вся система анализируется моделирующим алгоритмом на каждом такте, даже если в ней не происходит никаких изменений.

Другой недостаток состоит в том, что времена событий округляются до величины Δt , что ведет к **погрешностям** в определении переменных, характеризующих систему.

Рассмотрим **пример**. Моделируется склад изделий с максимальной емкостью G (см. рис. 32.1).



Рис. 32.1. Схема моделируемой производственной системы (пример)

Склад принимает изделия от трех поставщиков (обозначим номера входных потоков изделий 1, 2, 3) и выдает их трем потребителям (обозначим номера выходных потоков изделий 4, 5, 6). Примем в качестве характеристики каждого входа интенсивность i -го потока изделий (λ_i). Интенсивность характеризует в среднем расстояние между двумя моментами времени приходов (уходов) изделий на склад. Обозначим через P_i размер партии изделий, уходящих или приходящих на склад. То есть этими переменными мы определили, когда и сколько приходит или уходит изделий на тот или иной вход или выход склада (**блок 1** — здесь и далее см. блоки на рис. 32.2).

Обозначим переменной Z текущее количество изделий на складе. Если изделия приходят на склад (потоки 1, 2, 3), то Z увеличивается на P_1 , P_2 или P_3 . Если изделия изымаются со склада (потоки 4, 5, 6), то Z уменьшается на P_4 , P_5 или P_6 . То есть Z играет роль счетчика изделий, находящихся на складе в отдельный момент времени t . Начальное состояние склада на момент $t := 0$ примем $Z := Z_0$.

Алгоритм построим таким образом, чтобы вычислить вероятности событий возникновения дефицита P_d и переполнения P_n на складе.

Для накопления надежной статистики эксперимент повторяется KK раз. За количеством экспериментов следит счетчик экспериментов k (**блоки 2, 3, 8**). Каждый эксперимент длится от 0 до T_k момента времени (**блоки 5, 7**). Счетчик времени t отсчитывает время от 0 до T_k с дискретностью Δt (**блок 11**).

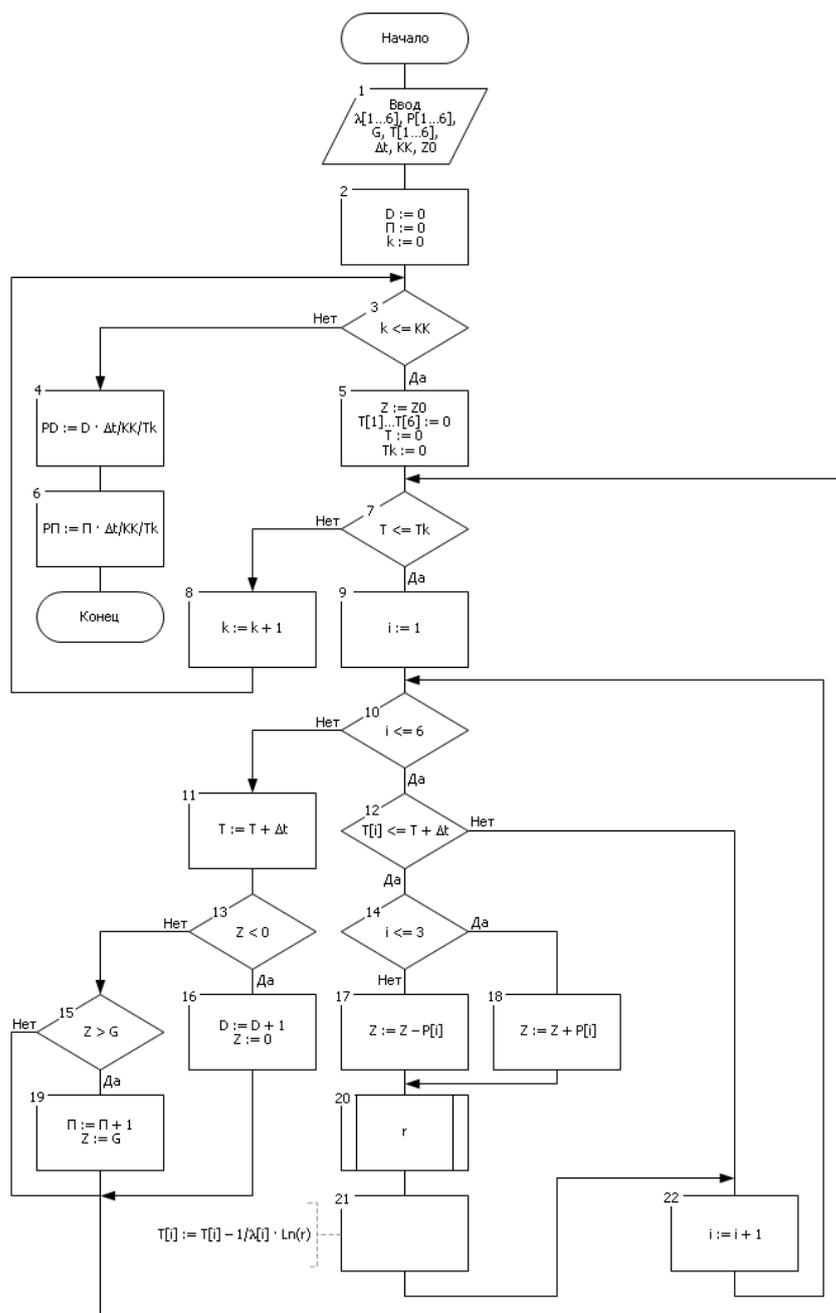
В каждом эксперименте подсчитывается, сколько раз на складе в результате действия входных сигналов (количество поставленных и изъятых изделий) возникает ситуация дефицита (**блок 13**). За этим следит счетчик D , если на складе возникает ситуация $Z < 0$ (**блок 13**), значит, возникает дефицит изделий для обработки и счетчик увеличивает свое значение на 1 (**блок 16**). Если дефицит не возникает $Z \geq 0$, то счетчик своего значения не меняет.

Поскольку всего может быть рассмотрено $N := T_k/\Delta t$ тактов, то вероятность возникновения дефицита может быть примерно оценена частотой возникновения этих событий как $P_d := D/N$ или с учетом того, что счетчик D

накапливался в течении KK экспериментов, то $P_d := D/N/KK$. Окончательно, подставляя N : $P_d := D \cdot \Delta t/T_k/KK$ (блоки 4, 6). Расчет вероятности переполнения склада моделируется аналогично (учитывается, что переполнение возникает, когда Z становится больше емкости склада G) (блоки 15, 19).

Поскольку на складе в реальности не может быть изделий меньше нуля, то значение Z в момент обнаружения этого факта должно быть возвращено на ближайшую границу, то есть: $Z := 0$ (блок 16). То же самое касается ситуации переполнения ($Z := G$) (блок 19).

Алгоритм представляет собой цикл по времени от 0 до T_k с шагом Δt — моделирование производится во времени. В каждом цикле (на каждом такте времени) проверяется, лежит ли сгенерированное T_i событие в интервале времени $[T, T + \Delta t]$: $T \leq T_i < T + \Delta t$ (блок 12). Если событие происходит в данный момент, то определяется, какой i -ый входной сигнал предопределил это событие (блоки 22, 10), обрабатывается ситуация (блок 17 или 18) и генерируется время следующего события (блоки 20, 21). Проверка на каждом такте происходит по всем возможным входным и выходным сигналам. Блок-схема алгоритма показана на рис. 32.2.



33.МОДЕЛИРОВАНИЕ МАРКОВСКИХ СЛУЧАЙНЫХ ПРОЦЕССОВ

Удобно описывать появление случайных событий в виде вероятностей переходов из одного состояния системы в другое, так как при этом считается, что, перейдя в одно из состояний, система не должна далее учитывать обстоятельства того, как она попала в это состояние.

Случайный процесс называется **марковским процессом** (или **процессом без последействия**), если для каждого момента времени t вероятность любого состояния системы в будущем зависит только от ее состояния в настоящем и не зависит от того, как система пришла в это состояние.

Марковский процесс удобно задавать графом переходов из состояния в состояние. Мы рассмотрим два варианта описания марковских процессов — с **дискретным и непрерывным временем**.

В первом случае переход из одного состояния в другое происходит в заранее известные моменты времени — такты (1, 2, 3, 4, ...). Переход осуществляется на каждом такте, то есть исследователя интересует только последовательность состояний, которую проходит случайный процесс в своем развитии, и не интересует, когда конкретно происходил каждый из переходов.

Во втором случае исследователя интересует и цепочка меняющихся друг друга состояний, и моменты времени, в которые происходили такие переходы.

И еще. Если вероятность перехода не зависит от времени, то марковскую цепь называют **однородной**.

Марковский процесс с дискретным временем

Итак, модель марковского процесса представим в виде графа, в котором состояния (вершины) связаны между собой связями (переходами из i -го состояния в j -е состояние), см. рис. 33.1.

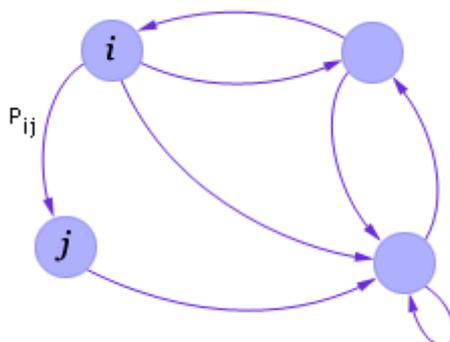


Рис. 33.1. Пример графа переходов

Каждый переход характеризуется **вероятностью перехода** P_{ij} . Вероятность P_{ij} показывает, как часто после попадания в i -е состояние осуществляется затем переход в j -е состояние. Конечно, такие переходы происходят случайно, но если измерить частоту переходов за достаточно большое время, то окажется, что эта частота будет совпадать с заданной вероятностью перехода.

Ясно, что у каждого состояния сумма вероятностей всех переходов (исходящих стрелок) из него в другие состояния должна быть всегда равна 1 (см. рис. 33.2).

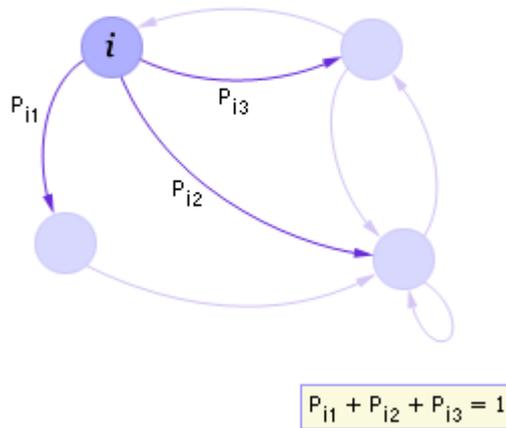


Рис. 33.2. Фрагмент графа переходов (переходы из i -го состояния являются полной группой случайных событий)

Например, полностью граф может выглядеть так, как показано на рис. 33.3.

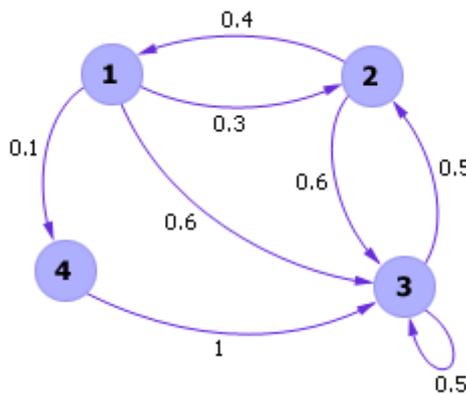


Рис. 33.3. Пример марковского графа переходов

Реализация марковского процесса (процесс его моделирования) представляет собой вычисление последовательности (цепи) переходов из состояния в состояние (см. рис. 33.4). Цепь на рис. 33.4 является случайной последовательностью и может иметь также и другие варианты реализации [9].



Рис. 33.4. Пример марковской цепи, смоделированной по марковскому графу, изображенному на рис. 33.3

Чтобы определить, в какое новое состояние перейдет процесс из текущего i -го состояния, достаточно разбить интервал $[0; 1]$ на подынтервалы величиной $P_{i1}, P_{i2}, P_{i3}, \dots$ ($P_{i1} + P_{i2} + P_{i3} + \dots = 1$), см. рис. 33.5. Далее с помощью ГСЧ надо получить очередное равномерно распределенное в интервале $[0; 1]$ случайное число r_{pp} и определить, в какой из интервалов оно попадает (см. лекцию 23).

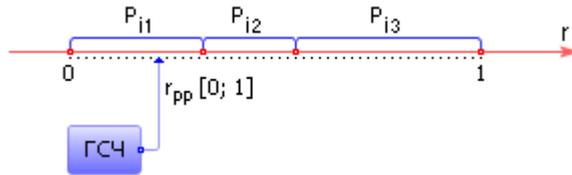


Рис. 33.5. Процесс моделирования перехода из i -го состояния марковской цепи в j -е с использованием генератора случайных чисел

После этого осуществляется переход в состояние, определенное ГСЧ, и повтор описанной процедуры для нового состояния. Результатом работы модели является марковская цепь (см. рис. 33.4).

Пример. Имитация стрельбы из пушки по цели. Для того, чтобы проимитировать стрельбу из пушки по цели, построим модель марковского случайного процесса.

Определим следующие три состояния: S_0 — цель не повреждена; S_1 — цель повреждена; S_2 — цель разрушена. Зададим вектор начальных вероятностей:

	S_0	S_1	S_2
P_0	0.8	0.2	0

Значение P_0 для каждого из состояний показывает, какова вероятность каждого из состояний объекта до начала стрельбы.

Зададим матрицу перехода состояний (см. табл. 33.1).

Таблица 33.1.
Матрица вероятностей перехода дискретного марковского процесса

	В S_0	В S_1	В S_2	Сумма вероятностей переходов
Из S_0	0.45	0.40	0.15	$0.45 + 0.40 + 0.15 = 1$
Из S_1	0	0.45	0.55	$0 + 0.45 + 0.55 = 1$
Из S_2	0	0	1	$0 + 0 + 1 = 1$

Матрица задает вероятность перехода из каждого состояния в каждое. Заметим, что вероятности заданы так, что сумма вероятностей перехода из некоторого состояния в остальные всегда равна единице (куда-то система должна перейти обязательно).

Наглядно модель марковского процесса можно представить себе в виде следующего графа (см. рис. 33.6).

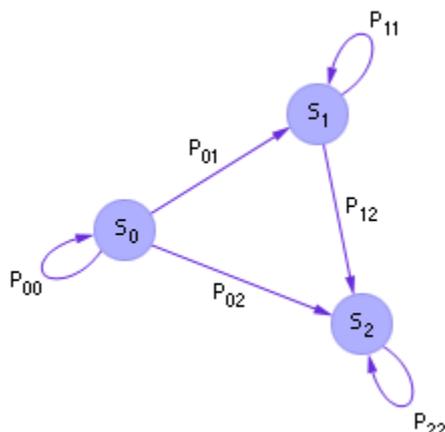


Рис. 33.6. Граф марковского процесса, моделирующий стрельбу из пушки по цели

Используя модель и метод статистического моделирования, попытаемся решить следующую задачу: определить среднее количество снарядов, необходимое для полного разрушения цели.

Проимитируем, используя таблицу случайных чисел, процесс стрельбы. Пусть начальное состояние будет S_0 . Возьмем последовательность из таблицы случайных чисел: 0.31, 0.53, 0.23, 0.42, 0.63, 0.21, ... (случайные числа можно взять, например, из этой таблицы).

0.31: цель находится в состоянии S_0 и остается в состоянии S_0 , так как $0 < \mathbf{0.31} < 0.45$;

0.53: цель находится в состоянии S_0 и переходит в состояние S_1 , так как $0.45 < \mathbf{0.53} < 0.45 + 0.40$;

0.23: цель находится в состоянии S_1 и остается в состоянии S_1 , так как $0 < \mathbf{0.23} < 0.45$;

0.42: цель находится в состоянии S_1 и остается в состоянии S_1 , так как $0 < \mathbf{0.42} < 0.45$;

0.63: цель находится в состоянии S_1 и переходит в состояние S_2 , так как $0.45 < \mathbf{0.63} < 0.45 + 0.55$.

Так как достигнуто состояние S_2 (далее цель переходит из S_2 в состояние S_2 с вероятностью 1), то цель поражена. Для этого в данном эксперименте потребовалось 5 снарядов.

На рис. 33.7 приведена временная диаграмма, которая получается во время описанного процесса моделирования. Диаграмма показывает, как во времени происходит процесс изменения состояний. Такт моделирования для данного случая имеет фиксированную величину. Нам важен сам факт перехода (в какое состояние переходит система) и не важно, когда это происходит.

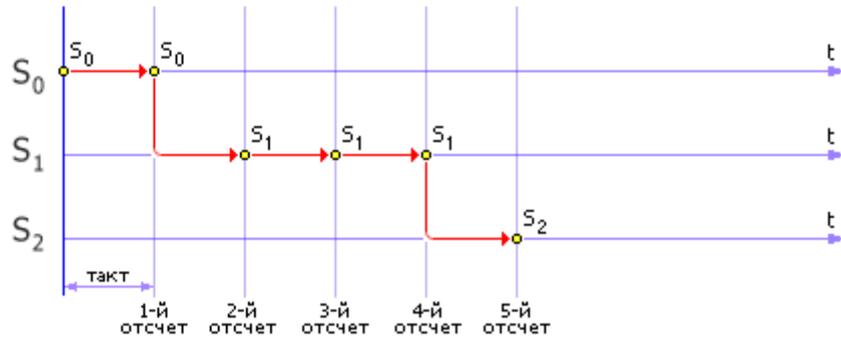


Рис. 33.7. Временная диаграмма переходов в марковском графе (пример имитации)

Процедура уничтожения цели совершена за 5 тактов, то есть марковская цепь этой реализации выглядит следующим образом: $S_0 \rightarrow S_0 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2$. Конечно, ответом задачи это число быть не может, так как в разных реализациях получатся разные ответы. А ответ у задачи может быть только один.

Повторяя данную имитацию, можно получить, например, еще такие реализации (это зависит от того, какие конкретно случайные числа выпадут): 4 ($S_0 \rightarrow S_0 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2$); 11 ($S_0 \rightarrow S_0 \rightarrow S_0 \rightarrow S_0 \rightarrow S_0 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2$); 5 ($S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2$); 6 ($S_0 \rightarrow S_0 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2$); 4 ($S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2$); 6 ($S_0 \rightarrow S_0 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2$); 5 ($S_0 \rightarrow S_0 \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2$). Всего уничтожено 8 целей. Среднее число циклов в процедуре стрельбы составило: $(5 + 4 + 11 + 5 + 6 + 4 + 6 + 5)/8 = 5.75$ или, округляя, 6. Именно столько снарядов, в среднем, рекомендуется иметь в боевом запасе пушки для уничтожения цели при таких вероятностях попаданий.

Теперь следует определить точность. Именно точность может нам показать, насколько следует доверять данному ответу. Для этого проследим, как сходится последовательность случайных (приближенных) ответов к правильному (точному) результату. Напомним, что, согласно центральной предельной теореме (см. лекцию 25, [лекцию 21](#)), сумма случайных величин есть величина неслучайная, поэтому для получения статистически достоверного ответа необходимо следить за средним числом снарядов, получаемых в ряде случайных реализаций.

На первом этапе вычислений средний ответ составил 5 снарядов, на втором этапе средний ответ составил $(5 + 4)/2 = 4.5$ снаряда, на третьем — $(5 + 4 + 11)/3 = 6.7$. Далее ряд средних величин, по мере накопления статистики, выглядит следующим образом: 6.3, 6.2, 5.8, 5.9, 5.8. Если изобразить этот ряд в виде графика средней величины выпущенных снарядов, необходимых для поражения цели, в зависимости от номера эксперимента, то обнаружится, что данный ряд сходится к некоторой величине, которая и является ответом (см. рис. 33.8).

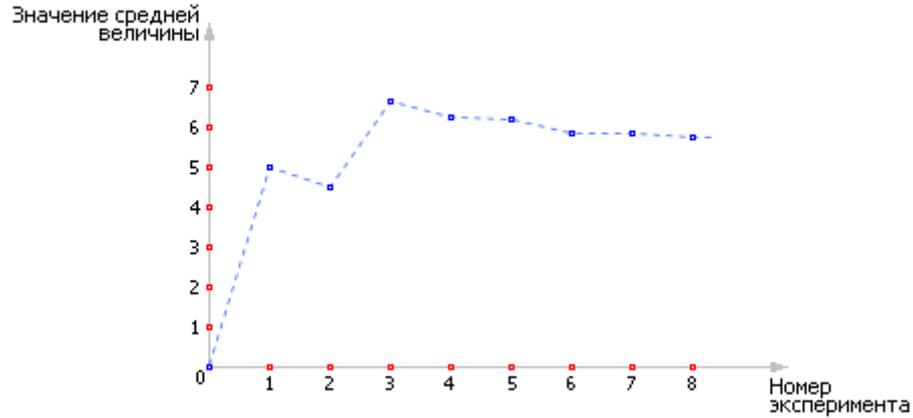


Рис. 33.8. Изменение средней величины в зависимости от номера эксперимента

Визуально мы можем наблюдать, что график «успокаивается», разброс между вычисляемой текущей величиной и ее теоретическим значением со временем уменьшается, стремясь к статистически точному результату. То есть в некоторый момент график входит в некоторую «трубку», размер которой и определяет точность ответа.

Алгоритм имитации будет иметь следующий вид (см. рис. 33.9).

Еще раз заметим, что в вышерассмотренном случае нам безразлично, в какие моменты времени будет происходить переход. Переходы идут такт за тактом. Если важно указать, в какой именно момент времени произойдет переход, сколько времени система пробудет в каждом из состояний, требуется применить модель с непрерывным временем.

Марковские случайные процессы с непрерывным временем

Модель марковского процесса представим в виде графа, в котором состояния (вершины) связаны между собой связями (переходами из i -го состояния в j -е состояние), см. рис. 33.10.

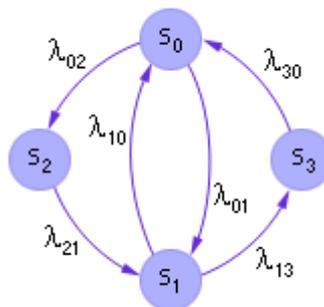


Рис. 33.10. Пример графа марковского процесса с непрерывным временем

Теперь каждый переход характеризуется плотностью вероятности перехода λ_{ij} . По определению:

$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(\Delta t)}{\Delta t}$$

При этом плотность понимают как распределение вероятности во времени.

Переход из i -го состояния в j -е происходит в случайные моменты времени, которые определяются интенсивностью перехода λ_{ij} .

К интенсивности переходов (здесь это понятие совпадает по смыслу с распределением плотности вероятности по времени t) переходят, когда процесс непрерывный, то есть, распределен во времени.

С интенсивностью потока (а переходы — это поток событий) мы уже научились работать в лекции 28. Зная интенсивность λ_{ij} появления событий, порождаемых потоком, можно симитировать случайный интервал между двумя событиями в этом потоке.

$$\tau_{ij} = -\frac{1}{\lambda_{ij}} \cdot \text{Ln}(R)$$

где τ_{ij} — интервал времени между нахождением системы в i -ом и j -ом состоянии.

Система из любого i -го состояния может перейти в одно из нескольких состояний $j, j + 1, j + 2, \dots$, связанных с ним переходами $\lambda_{ij}, \lambda_{ij+1}, \lambda_{ij+2}, \dots$

В j -е состояние она перейдет через τ_{ij} ; в $(j + 1)$ -е состояние она перейдет через τ_{ij+1} ; в $(j + 2)$ -е состояние она перейдет через τ_{ij+2} и т. д.

Система может перейти из i -го состояния только в одно из этих состояний, причем в то, переход в которое наступит раньше.

Из последовательности времен: $\tau_{ij}, \tau_{ij+1}, \tau_{ij+2}$ и т. д. надо выбрать минимальное и определить индекс j , указывающий, в какое именно состояние произойдет переход.

Пример. Моделирование работы станка. Промоделируем работу станка (см. рис. 33.10), который может находиться в следующих состояниях: S_0 — станок исправен, свободен (простой); S_1 — станок исправен, занят (обработка); S_2 — станок исправен, замена инструмента (переналадка) $\lambda_{02} < \lambda_{21}$; S_3 — станок неисправен, идет ремонт $\lambda_{13} < \lambda_{30}$.

Зададим значения параметров λ , используя экспериментальные данные, получаемые в производственных условиях: λ_{01} — поток на обработку (без переналадки); λ_{10} — поток обслуживания; λ_{13} — поток отказов оборудования; λ_{30} — поток восстановлений.

Реализация будет иметь следующий вид (см. рис. 33.11).

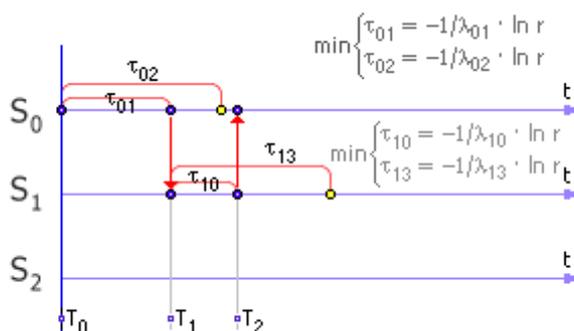


Рис. 33.11. Пример моделирования непрерывного марковского процесса с визуализацией на временной диаграмме (желтым цветом указаны запрещенные, синим — реализовавшиеся состояния)

В частности, из рис. 33.11 видно, что реализовавшаяся цепь выглядит так: $S_0—S_1—S_0—\dots$. Переходы произошли в следующие моменты времени: $T_0—T_1—T_2—T_3—\dots$, где $T_0 = 0$, $T_1 = \tau_{01}$, $T_2 = \tau_{01} + \tau_{10}$.

Задача. Поскольку модель строят для того, чтобы на ней можно было решить задачу, ответ которой до этого был для нас совсем не очевиден (см. лекцию 1), то сформулируем такую задачу к данному примеру. Определить долю времени в течение суток, которую занимает простой станка (посчитать по рисунку) $T_{cp} = (T + T + T + T)/N$.

Алгоритм имитации будет иметь следующий вид (см. рис. 33.12).

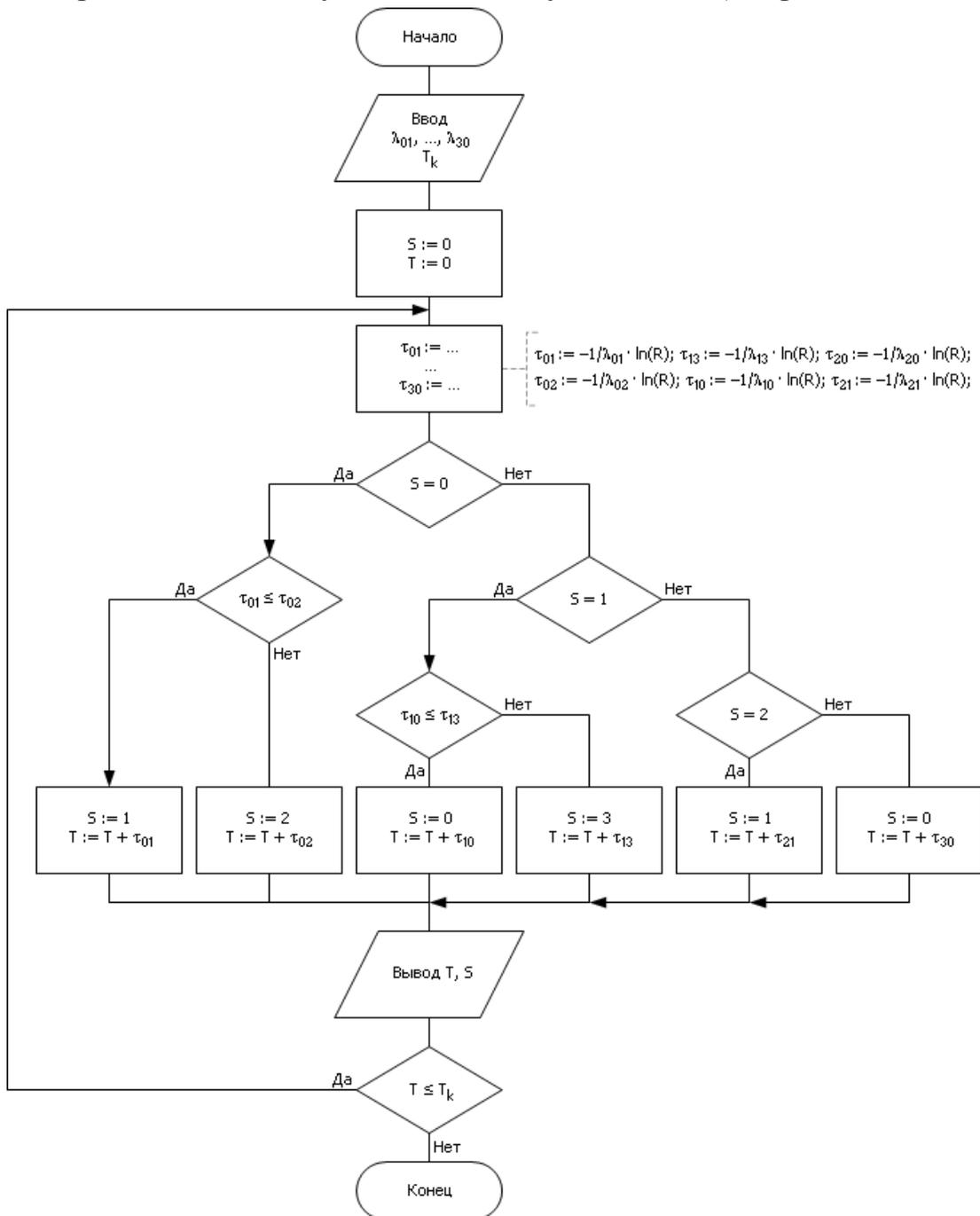


Рис. 33.12. Блок-схема алгоритма моделирования непрерывного марковского процесса на примере имитации работы станка

Часто аппарат марковских процессов используется при моделировании компьютерных игр, действий компьютерных героев.

34. ФИКСАЦИЯ И ОБРАБОТКА СТАТИСТИЧЕСКИХ РЕЗУЛЬТАТОВ

В лекции 21 мы познакомились со схемой статистического компьютерного эксперимента. В лекциях 21-26 мы рассмотрели практическую реализацию всех основных блоков (см. рис. 21.3) этой схемы. Сейчас важно научиться организовывать работу последних двух блоков - блок вычисления статистических характеристик (БВСХ) и блок оценки достоверности статистических результатов (БОД).

Рассмотрим, как следует фиксировать статистические величины в результате эксперимента, чтобы получить надежную информацию о свойствах моделируемого объекта. Напомним, что обобщенными характеристиками случайного процесса или явления являются средние величины.

Вычисление средних

Вычисление средних величин во время эксперимента, который многократно повторяется, а результат его усредняется, может быть организовано несколькими способами:

- вся статистика вычисляется в конце;
- вся статистика вычисляется в процессе вычисления (по рекурсивным соотношениям);
- вся статистика вычисляется в классовых интервалах (этот метод совмещает универсальность первого метода и экономичность второго).

Способ 1. Вычисление всей статистики в конце. Для этого в процессе эксперимента значения X_i выходной (изучаемой) случайной величины X накапливаются в массиве данных. После окончания эксперимента подсчитывается математическое ожидание (среднее) X и дисперсия D (характерный разброс величин относительно этого математического ожидания).

$$X = \frac{1}{n} \cdot \sum_{i=1}^n X_i \quad (1)$$

$$D = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - X)^2 \quad (2)$$

Часто используют среднеквадратичное отклонение $\sigma = \sqrt{D}$.

Заметим, что недостатком метода является неэффективное использование памяти, так как приходится накапливать и сохранять большое количество значений выходной величины в течение всего эксперимента, который может быть весьма продолжительным.

Второй минус заключается в том, что приходится дважды считывать массив X_i , так как воспользоваться формулой (2) в том виде, как она здесь записана, мы можем, только просчитав формулу (1) (от 1 до n), а потом еще раз прогнав для формулы (2) массив X_i .

Положительным моментом является сохранение всего массива данных, что дает возможность более подробного его изучения в дальнейшем при необходимости расследования тех или иных эффектов и результатов.

Способ 2. Вычисление всей статистики в процессе вычисления (по рекурсивным соотношениям). Этот способ предусматривает возможность

хранить только текущее значение математического ожидания X_i и дисперсии D_i , подправляемое на каждой итерации. Это избавляет нас от необходимости постоянного хранения всего массива экспериментальных данных. Каждое новое данное X_i учитывается в сумме с весовым коэффициентом — чем более слагаемых i накоплено в сумме X_i , тем более ее значение важно по отношению к очередной поправке X_i , поэтому соотношение весовых коэффициентов $i/(i+1) : 1/(i+1)$.

$$X_{i+1} = \frac{X_i \cdot i + X_{i+1}}{i+1} = X_i \cdot \frac{i}{i+1} + X_{i+1} \cdot \frac{1}{i+1}$$

$$D_{i+1} = \frac{(X_{i+1} - X_i) + i \cdot D_i}{i+1} = \frac{(X_{i+1} - X_i)}{i+1} + \frac{i}{i+1} \cdot D_i$$

где X_i — очередное значение экспериментальной выходной величины.

Способ 3. Вычисление всей статистики в классовых интервалах.

Этот способ предполагает, что в массив будут накапливать не все значения X_i , а только по значимым интервалам, в которых распределена случайная выходная величина X_i . Общий интервал изменения X_i разбивается на m подинтервалов, в каждом из которых фиксируется количество n_i , которое показывает, сколько раз X_i приняло значение из i -го интервала. При небольшом количестве интервалов ($m \approx 1$) мы получаем способ 1, при количестве интервалов $m = n$ мы получаем способ 2. В случае $1 < m < n$ получаем среднее решение — компромисс между занимаемой памятью и информативностью массива выходных данных.

$$X = \frac{n_1 \cdot X_1 + n_2 \cdot X_2 + \dots + n_m \cdot X_m}{n_1 + n_2 + \dots + n_m} = \frac{1}{N} \cdot \sum_{i=1}^n n_i X_i$$

$$D = \frac{1}{n} \cdot \sum_{i=1}^n (n_i \cdot X_i - X)^2$$

Вычисление геометрии распределения

Еще более информативным является вычисление геометрии распределения случайной величины. Оно необходимо для того, чтобы представить себе более точно характер распределения. Известно, что по значению статистического момента можно приблизительно судить о геометрическом виде распределения.

Первый момент (или среднее арифметическое) вычисляется так:

$$m_1 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - A)^1, \text{ где } A = \begin{cases} 0, (\text{начальный момент}) \\ X, (\text{центральный момент}) \end{cases}$$

Если A принимает значение 0, то первый момент называется **начальным моментом**, если A принимает значение X , то первый момент называется **центральным**. (В принципе A может быть любым числом, задаваемым исследователем.)

На практике принято использовать не сам первый момент, а нормированную его величину $R_1 = m_1/\sigma^1$.

Первый момент указывает на центр тяжести в геометрии распределения, см. рис. 34.1.

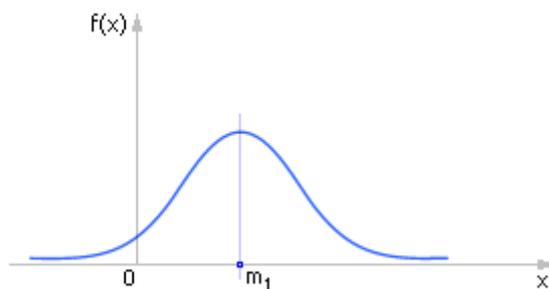


Рис. 34.1. Положение первого момента на графике распределения статистической величины

Второй момент (или дисперсия, разброс) вычисляется так:

$$m_2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$$

Понятие **среднеквадратичного отклонения**, связанным со вторым моментом:

$$\sigma = \sqrt{m_2}$$

Принято использовать не сам второй момент, а нормированную его величину $R_2 = m_2/\sigma^2$.

Дисперсия характеризует величину разброса экспериментальных данных относительно центра тяжести m_1 . По величине m_2 можно судить о втором параметре геометрии распределения (см. рис. 34.2).

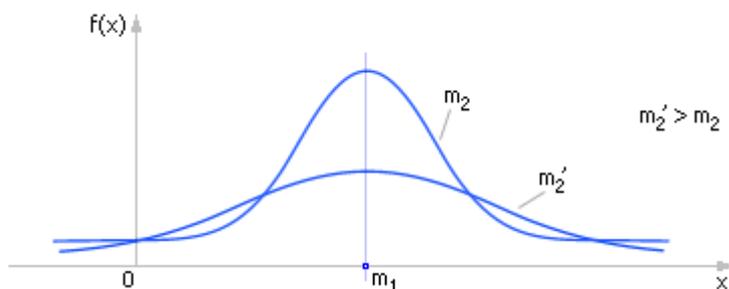


Рис. 34.2. Характерное изменение вида распределения статистической величины в зависимости от величины второго момента

Третий момент характеризует асимметрию (или скошенность) (см. рис. 34.3) вычисляется так:

$$m_3 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^3$$

Принято использовать не сам второй момент, а нормированную его величину $R_3 = m_3/\sigma^3$.

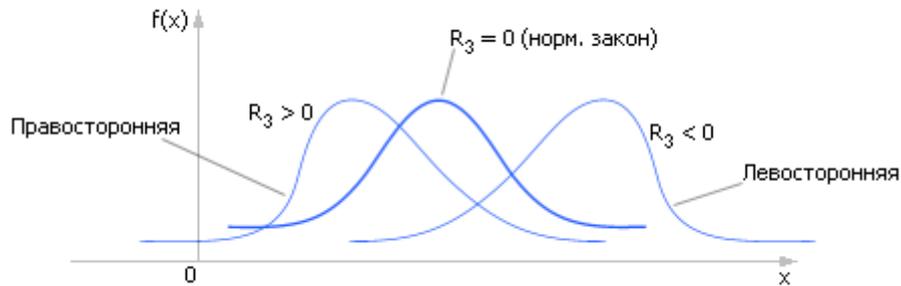


Рис. 34.3. Характерное изменение вида распределения статистической величины в зависимости от величины третьего момента

Определяя знак R_3 , можно определить, есть ли асимметрия у распределения (см. рис. 34.3), а если есть ($R_3 \neq 0$), то в какую сторону.

Четвертый момент (см. рис. 34.4) характеризует эксцесс (или остроконечность) и вычисляется так:

$$m_4 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^4$$

Нормированный момент равен: $R_4 = m_4/\sigma^4$.

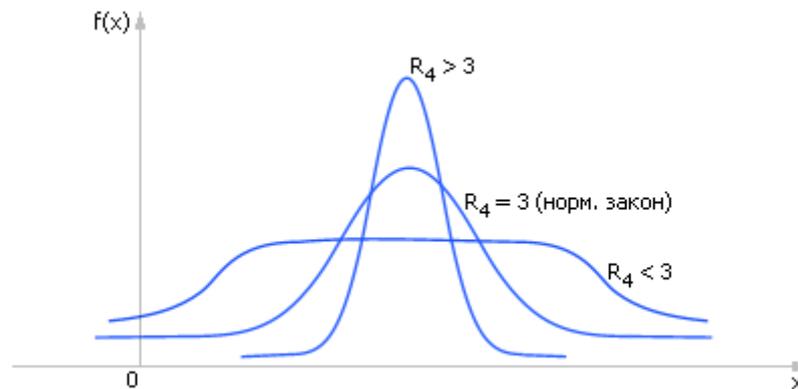


Рис. 34.4. Характерное изменение вида распределения статистической величины в зависимости от величины четвертого момента

Важным является выяснение того, на какое распределение более всего походит полученное экспериментальное распределение случайной величины. Оценка степени совпадения эмпирического закона распределения с теоретическим проводится в два этапа: определяют параметры экспериментального распределения и далее производят оценку по Колмогорову соответствия экспериментального распределения выбранному теоретическому.

Оценка (по Колмогорову) совпадения эмпирического закона распределения с теоретическим

1. Вычисляем моменты m_1, m_2, m_3, \dots . Число моментов равно числу неизвестных в теоретическом законе распределения.
2. Прежде всего, так как оценка касается непрерывного распределения, а мы имеем дело с дискретным распределением, снятым экспериментально, то надо решить, на сколько интервалов надо разбить при дискретизации и то, и другое распределение.

Для этого рекомендуется пользоваться **правилом Стерджеса**, хорошо зарекомендовавшим себя на практике: $K = 1 + \log_2 n = 1 + 3.322 \cdot \log_{10} n$, где n — количество случайных значений (опытов), k — количество интервалов распределения.

Строится интегральный (см. рис. 34.5) закон для эмпирического распределения $F(x) = P(x \leq x_i)$.

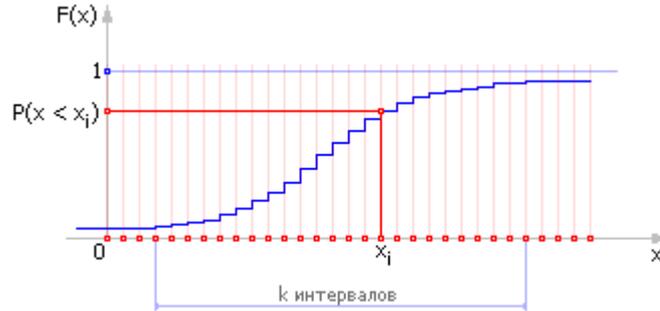


Рис. 34.5. Интегральный закон эмпирического распределения, дискретный вариант

3. В зависимости от числа экспериментов n и количества интервалов $1 \leq i \leq k$ можно посчитать число исходов в каждом из интервалов: $N_i = P_i \cdot n$.
4. Далее рассчитаем теоретическое распределение частоты: $N_i^{\text{ТЕОР.}} = P_i \cdot n$. Если в качестве теоретического принять нормальный закон распределения, то можно сделать так:

$$P_i(x < x_i) = F\left(\frac{x - a}{\sigma}\right) + \frac{1}{2}$$

где F — функция Лапласа, а параметры a и σ закона вычислены в п. 1.

5. Сравним полученные частоты: $N_i^{\text{ТЕОР.}}$ и N_i во всех k интервалах (см. рис. 34.6) и выберем наибольшее отклонение экспериментального распределения от проверяемого теоретического:

$$D = \max_i \left| N_i - N_i^{\text{ТЕОР.}} \right|$$

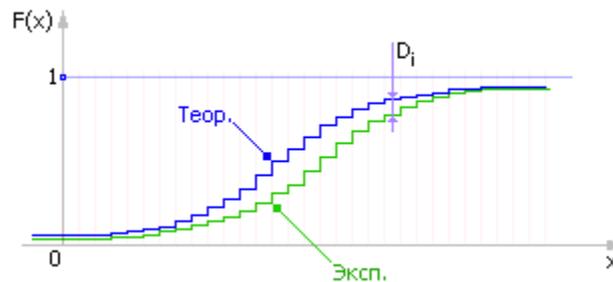


Рис. 34.6. Сравнение теоретического и эмпирического интегральных распределений случайной величины

6. Параметр Колмогорова λ характеризует отклонение теоретического распределения от экспериментального:

$$\lambda = \left(\sqrt{k} + 0.12 + \frac{0.11}{\sqrt{k}} \right) \cdot D$$

Используя табл. 34.1 Колмогорова, следует принять или отвергнуть гипотезу о том, является ли эмпирическое распределение с заданной нами вероятностью Q теоретическим или нет. Для принятия гипотезы должно быть: $\lambda < \lambda_{\text{табл.}}$.

Таблица 34.1.
Таблица критерия Колмогорова

Q	0.85	0.90	0.95	0.99
λ	1.14	1.22	1.36	1.63

Примечание. Критерий Колмогорова не единственный возможный к применению при оценивании; можно использовать критерий Хи-квадрат, критерий Андерсона-Дарлинга и другие.

Оценка точности статических характеристик

Крайне важным является вопрос, сколько экспериментов следует сделать, чтобы можно было доверять снятым характеристикам. Если экспериментов не достаточно, то характеристика недостоверна. Обычно исследователь задает доверительную вероятность, то есть вероятность, с которой он готов доверять снятым характеристикам. Чем больше будет задана доверительная вероятность, тем больше экспериментов потребуется сделать. Оценка будет основываться на центральной предельной теореме (см. лекцию 25, утверждающей, что сумма (или среднее) случайных величин есть величина неслучайная. ЦПТ утверждает, что значения вычисленной нами статистической характеристики будут распределены по нормальному закону, n_i — число i -ых исходов значения статистической характеристики в n экспериментах, $p_i = n_i/n$ — частота i -го исхода.

Если $n \rightarrow \infty$, то $p \rightarrow P$ (частота p стремится к теоретической вероятности P) и эмпирические характеристики будут стремиться к теоретическим (см. **рис. 34.7**). Итак, согласно ЦПТ p будет распределена по нормальному закону с математическим ожиданием m и среднеквадратичным отклонением σ .

При этом $m = P$, $\sigma = \sqrt{p \cdot (1 - p)/n}$.

Обозначим как Q **доверительную вероятность**, то есть вероятность того, что частота p отличается от вероятности P не более, чем на ε . Тогда по теореме Бернулли:

$$Q(|p - P| \leq \varepsilon) = F\left(\frac{\varepsilon}{\sigma}\right) = F\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

Величина ε называется **доверительным интервалом**. Смысл ε состоит в том, что в серии (каждая выборкой n) в среднем $\varepsilon \cdot 100\%$ доверительных интервалов содержат истинное значение статистической характеристики p . Как и ранее (см. лекцию 25), F — интеграл от функции нормального закона распределения, интегральная функция Лапласа.

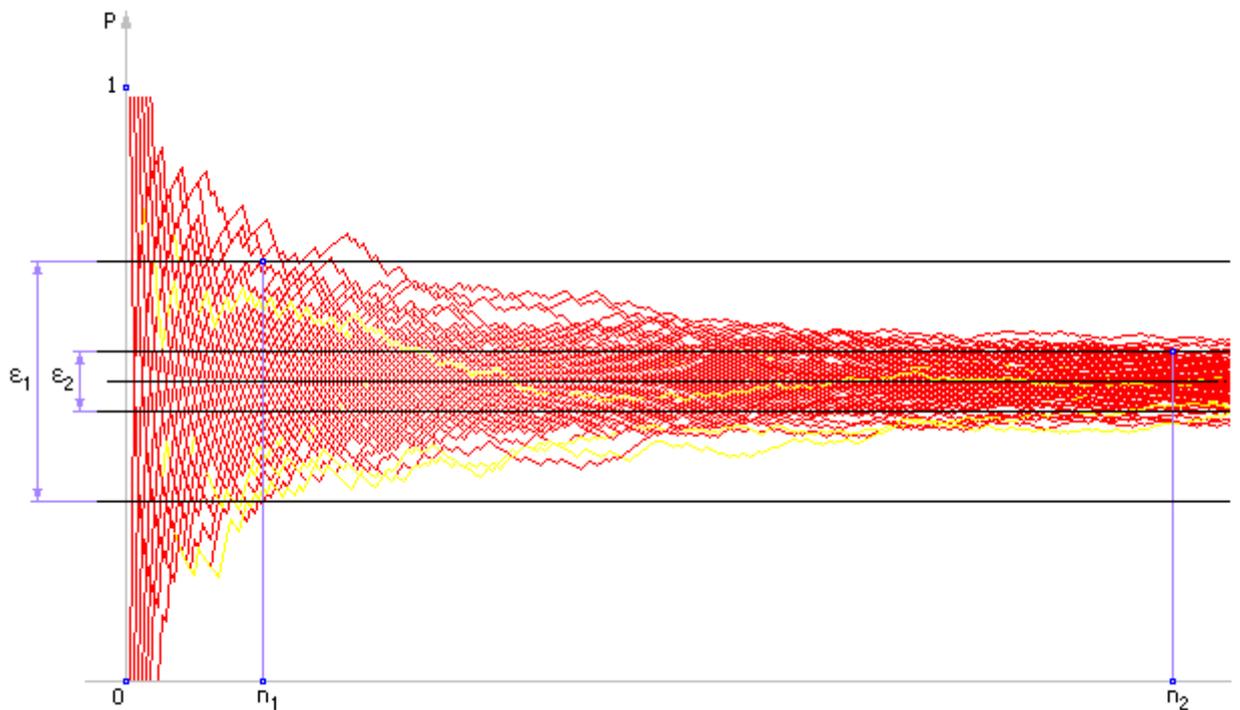


Рис. 34.7. Иллюстрация к вычислению количества экспериментов по величине доверительного интервала согласно центральной предельной теореме

Отсюда можно выразить требуемое для доверительной вероятности количество экспериментов (F^{-1} — обратная функция Лапласа):

$$n = \frac{p \cdot (1 - p)}{\varepsilon^2} \cdot \left(F^{-1}(Q) \right)^2$$

Пример. При моделировании выпускаемой продукции предприятием в результате имитации его работы в течение 50 дней были получены следующие выходные данные (см. табл. 34.2).

Таблица 34.2.
Экспериментальные статистические данные моделирования

Качество изделия в баллах (случайное событие i)	1	2	3	4
Количество исходов (n_i)	15	10	5	20
Частость исхода ($p_i = n_i/n$)	0.3	0.2	0.1	0.4

Было проведено: $15 + 10 + 5 + 20 = 50$ экспериментов ($n = 50$). Из таблицы экспериментов следует ответ задачи, что частость (вероятность) выпуска изделий 1 сорта равна $15/50$, частость (вероятность) выпуска изделий 2 сорта равна $10/50$, частость (вероятность) выпуска изделий 3 сорта равна $5/50$, частость (вероятность) выпуска изделий 4 сорта равна $20/50$.

Зададимся доверительной вероятностью к ответам модели $Q = 0.9$ и доверительным интервалом $\varepsilon = 0.05$.

Теперь надо ответить на вопрос: можно ли доверять с вероятностью Q вычисленному ответу?

Будем оценивать результат статистических экспериментов по наихудшей вероятности, таковой в нашей задаче является $p = 0.4$, так как вероятность, например, 0.1 определена намного лучше.

Очень важное примечание. Вообще вероятности (частоты) близкие к 0 или 1 весьма привлекательны в качестве ответа, так как вполне определяют решение. Вероятности близкие к 0.5 говорят о том, что ответ весьма неопределен, событие случится «50 на 50». Такой ответ удовлетворительным назвать сложно, он мало информативен.

Формула

$$n = \frac{p \cdot (1 - p)}{\varepsilon^2} \cdot \left(F^{-1}(Q) \right)^2$$

после подстановки значений $F^{-1}(0.9) = 1.65$ (см. [таблицу Лапласа](#)), далее $(F^{-1}(0.9))^2 = 2.7$, $p = 0.4$, $\varepsilon = 0.05$ дает $N = 0.4 \cdot 0.6 \cdot 2.7 / 0.05^2$ или окончательно $N = 250$.

Наш эксперимент и его ответ недостоверен относительно заданных Q и ε : 50 экспериментов недостаточно для ответа, требуется 250. То есть надо продолжать эксперименты и еще провести 200 экспериментов, чтобы достичь требуемой точности.

Формула использует себя рекуррентно. Сразу вычислить с ее помощью количество экспериментов n не удастся. Чтобы вычислить n , надо провести пробную серию экспериментов, оценить значение искомой статистической характеристики p , подставить это значение в формулу и определить необходимой число экспериментов.

Для уверенности данную процедуру следует провести несколько раз при разных получаемых последовательно значениях n .

В блоке оценки достоверности (БОД) (см. [лекцию 21](#)) анализируют степень достоверности статистических экспериментальных данных, снятых с модели (принимая во внимание точность результата Q и ε , заданные пользователем) и определяют необходимое для этого количество статистических испытаний n .

При большом количестве опытов n частота появления события p , полученная экспериментальным путем, стремится к значению теоретической вероятности появления события P . Если колебания значений частоты появления событий относительно теоретической вероятности меньше заданной точности, то экспериментальную частоту принимают в качестве ответа, иначе генерацию случайных входных воздействий продолжают, и процесс моделирования повторяется. При малом числе испытаний результат может оказаться недостоверным. Чем более испытаний, тем точнее ответ, согласно центральной предельной теореме. Количество требуемых экспериментов n даны для сравнения в табл. 34.3 и табл. 34.4 при различных комбинациях p и ε .

Таблица 34.3.

Количество экспериментов n , необходимых для вычисления достоверного ответа с доверительной вероятностью $Q = 0.95$, $(F^{-1}(0.95))^2 = 3.84$, $p = 0.1$

ε	0.001	0.005	0.010	0.050	0.100
Критическое количество экспериментов n	345600	13824	3456	138	35

Таблица 34.4.

Количество экспериментов n , необходимых для вычисления достоверного ответа с доверительной вероятностью $Q = 0.95$, $(F^{-1}(0.95))^2 = 3.84$, $p = 0.5$

ε	0.001	0.005	0.010	0.050	0.100
Критическое количество экспериментов n	960000	38400	9600	384	96

На рис. 34.8 отображен график зависимости $n(\varepsilon)$ при $Q = 0.95$ и $p = 0.5$.

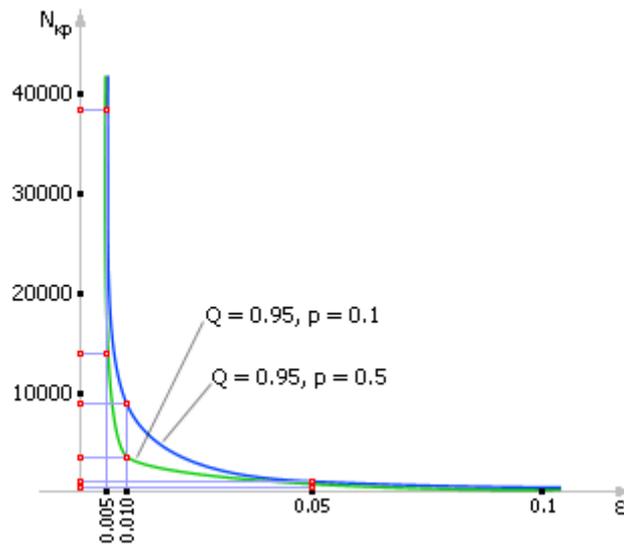


Рис. 34.8. Зависимость количества требуемых экспериментов от величины доверительной вероятности ε и доверительного интервала Q для случая частоты выпадения случайного события $p = 0.5$

Оценивание ведут по худшей из частот. Это обеспечивает достоверный результат сразу по всем снимаемым характеристикам модели.

Данная оценка количества экспериментов по ЦПТ не единственная из существующих. Известны аналогичные близкие по смыслу оценки Бернулли, Муавра-Лапласа, Чебышева.

Как объяснить, почему так странно ведет себя кривая снятой экспериментально статистической характеристики (см. рис. 34.7 и рис. 34.8)? При большом n кривая крайне медленно подходит к истинному значению, хотя сначала (при малых n) процесс идет с большой скоростью — мы быстро входим в область приближенного ответа (большие ε), но медленно приближаемся к точному ответу (малые ε).

Допустим, что мы провели N испытаний. Выпадений события в этих испытаниях составило число N_1 . Пусть вероятность выпадения события близка к $N_1/N = 0.5$ или $N = N_1 \cdot 2$.

Допустим, что мы хотим провести еще одно испытание $(N + 1)$ -е. Взяв ответ (частость N_1/N) при N за 100%, оценим, насколько процентов изменится ответ после следующего опыта? Составим пропорцию:

$$N_1/N — 100\% \quad (N_1 + 1)/(N + 1) — X\%$$

Отсюда имеем: $X = (N_1 + 1) \cdot 100 \cdot N / (N_1 \cdot (N + 1))$, при $N_1 = N/2$ (вероятность 0.5) получаем, что $X = 100 \cdot (N + 2)/(N + 1)$.

И величина X образует ряд: 150%, 133%, 125%, 120%, ..., 100.1%, ..., ...
→ 100%. Значит, сначала улучшение ответа на один дополнительный эксперимент составило 50%, на 2 — 33%, на 3 — 25%, на 4 — 20%, ..., на 100-м — всего на 0,1%.

Что улучшение точности на каждый новый эксперимент (значения X) сначала очень хорошее, а затем — незначительное, после 100 экспериментов эта величина меняется всего на доли процента в расчете на один дополнительный эксперимент! Итог: изменение оценки, основанной на сумме, после серии опытов перестает сильно меняться!!!

Итоги. Важно.

1. В качестве ответа статистического эксперимента принимается частость p появления некоторого выходного события, которая является оценкой вероятности. Чем больше экспериментов n , тем ближе частость p к вероятности P , а экспериментальный ответ к теоретическому.
2. Частости p , близкие по значению к 0 или 1, более предпочтительны в смысле информативности, чем частости близкие к 0.5, которые мало информативны и дают максимально неопределенный ответ.
3. В моделировании важной целью является понижение дисперсии ответа, разброса выходной величины модели относительно частости. Действительно, если разброс случайной величины m_2 мал, то вычисленный ответ достаточно достоверен. Если в ряду случайной величины встречаются значения достаточно удаленные друг от друга (см. рис. 34.2), то m_2' велика, и ответ недостаточно определен.
4. Статистический ответ оценивается не только значениями частости и разброса, но и точностью, роль которой играет **доверительная вероятность** Q и заданный доверительный интервал ε . Эти величины связаны с разбросом m_2 .

Необходимое количество статистических экспериментов n зависит от заданной точности (Q, ε) и характеристик процесса (частости p и разброса m_2). Повышение требований по точности, плохие характеристики существенно повышают затраты на исследование модели, увеличивая число экспериментов.

35. НЕФОРМАЛЬНЫЙ СИНТЕЗ

Если обратиться к технологической схеме моделирования (см. рис. 1.14), то мы увидим, что раздел «Неформальный синтез» предшествует непосредственному построению модели.

Данный этап представляет собой наиболее ответственную часть технологии (исправление просчетов в проектировании моделей, построенных на изначально ошибочной основе, обходится наиболее дорого) и одновременно довольно сложную, хотя на первый и неискушенный взгляд может показаться наоборот. Дело в том, что к моменту, когда человек приступает к проектированию, количество вариантов, возможных к развитию и дальнейшей реализации, довольно велико (можно сделать и так, и иначе, и еще десятком других способов), и аргументировано выбрать один из них достаточно сложно. Поэтому зачастую принимают первый попавшийся вариант, не особо аргументируя, почему был выбран именно он. Наиболее часто в такой ситуации оказываются программисты, которые в технологической цепочке разработчиков системы находятся ближе всех к ее непосредственной реализации и, соответственно, — дальше всех от постановки истинной проблемы, системно-технического исследования объекта [9].

Сложность этого этапа состоит в неформальном характере используемых им методов. Спецификой этапа является переход от неформального представления бесконечно сложного объекта, проблемы, не имеющей границ, от нечетких представлений, к формализованной постановке, четкому описанию системы. Переход от бесконечно сложного к представленному внутри неких рамок — это процесс перехода одного качества в другое, и при современном уровне развития науки он не может быть четко регламентирован. Данный переход можно осуществить только благодаря интеллектуальным усилиям и большому методологическому опыту, присущему человеку-постановщику, а также тщательно организованной работе группы экспертов, хорошо представляющих круг обсуждаемых проблем, но не владеющих специальными методиками формализации.

Процедура неформального синтеза строится на работе постановщика (аналитика) с экспертами и своей целью имеет формальное описание объекта, проблемы, задачи, идеи, которое далее развивается формальными методами в рамках формальной части технологии моделирования систем.

Следует отметить, что полагаться на плохо прогнозируемый опыт и знания аналитика трудно. Возможно, хорошая идея аналитику придет через пару дней, а возможно, и через год! Вторая трудность в его работе заключается в том, что сложно определить, какие факты являются важными для их учета в модели, а какие окажутся несущественными. Возможно, что аналитик не сможет учесть некоторый фактор, и впоследствии из-за этого придется переосмыслить всю концепцию в целом.

С целью снижения такой неопределенности были созданы (и продолжают создаваться) технологии, сопровождающие этапы неформального синтеза. Технология сама по себе, по определению, подразумевает получение при ее

использовании продукта гарантированного качества в заранее известные сроки и с заранее известными затратами.

Неформальный синтез — это предпроектная стадия исследования объекта. Основные разделы этой стадии:

- интервьюирование экспертов (раздел соответствует накоплению информации о проблеме, объекте);
- генерация идей (раздел соответствует генерации вариантов будущих решений на основе отобранной информации, сгущение информации до нескольких вариантов, порождение веера вариантов);
- экспертиза идей (раздел соответствует принятию решения, выбору из многих вариантов одного-двух самых перспективных, первоочередных, отбор из веера вариантов);
- процедура формализации (раздел соответствует осмыслению, оформлению, приведению к каноническому виду для дальнейшего синтеза модели проекта формальными методами).

Интервью с экспертом

Ни одно аналитическое средство не даст возможности аналитику определить, что имеет в виду эксперт, если последний не скажет об этом сам. Только при этом условии аналитик в состоянии извлечь из эксперта необходимую информацию. Отметим, что есть технологии силового извлечения информации, разведки, подавления, дезинформации, изучаемые в отдельных специальных курсах.

Задача аналитика — правильно построить интервью, чтобы эксперт дал постановщику максимально полезную и нужную для проекта информацию.

Еще до момента начала интервью аналитик должен четко и ясно отдавать себе отчет о его теме и рамках; аналитик должен сфокусироваться на определенной проблеме, настроиться на нее. Для этого надо составить план интервью. До начала интервью аналитик должен иметь план опроса, подобный оглавлению книги (пункты, подпункты и т. п.). Приветствуется иерархическое построение плана — это говорит о его глубокой проработке. Необходимо заранее представлять себе, сколь долго будет продолжаться интервью.

Логическая единица интервью — это тема и степень ее детализации. При проведении интервью необходимо соблюдать следующие рекомендации (которые имеют название «воронка Гэллапа»):

- выяснить, думал ли об этой проблеме эксперт;
- как относится к проблеме эксперт;
- получить ответы по основным аспектам;
- выяснить причины, по которым эксперт имеет такие взгляды;
- оценить интенсивность его взглядов (силу убежденности).

Следует отдавать себе отчет, что сведения, получаемые в ходе интервью, бывают трех типов:

- повествовательные;
- описательные;
- в виде рассуждений (они наиболее предпочтительны).

Некоторые сведения об интервью.

Вопросы могут быть «закрытыми» и «открытыми». Закрытые вопросы предполагают, что аналитик дает на выбор эксперту варианты ответов. Обычно интервью начинается с закрытых вопросов. В дальнейшем, когда эксперт входит в режим рассуждений, лучше перейти к открытым вопросам, хотя они могут значительно увести от темы интервью.

Вопросы могут быть общими и частными. Общие вопросы обычно не искажают мнения эксперта. Во время интервью следует продвигаться от общих вопросов к более частным.

Вопрос должен быть уравновешенным, то есть он не должен заранее содержать оценки (отвечающий может согласиться, не согласиться, воздержаться). Будет ошибкой строить вопрос так, что при любом ответе эксперта результат, тем не менее, одинаков. В этом случае аналитик программирует эксперта, что опасно для результата исследования.

Вопросы не должны содержать неявные утверждения, оказывающие на эксперта давление, и должны быть нейтральными.

Пример давления на эксперта: в вопросе «Потеряли ли вы рога?» содержится *неявное утверждение*. Так как при ответе «Да» подразумевается, что рога у вас были. При ответе «Нет» подразумевается, что они до сих пор при вас. То есть, как бы эксперт ни ответил, его ответ как бы запрограммирован заранее.

Не нейтральное утверждение. В зависимости от формулировки вопроса — «Как вы относитесь к захвату страны неграми?» или «Каково ваше отношение к тому, что в вашем микрорайоне поселится негр?» — в массе будет получен различный ответ.

Дадим еще некоторые советы, которые могут быть полезны при подготовке и проведении интервью.

- 1) Проводить интервью следует при отсутствии третьих лиц.
- 2) Не должно быть ограничений по времени окончания интервью (например, не надо начинать интервью в пять часов вечера в пятницу летом).
- 3) Оповестите сразу, сколько вы займете времени у интервьюируемого.
- 4) У аналитика все необходимое должно быть с собой (бумага, ручка и т. д.).
- 5) Одежда интервьюера должна быть в стиле организации, в которой он собирается провести интервью.
- 6) Необходимо, чтобы опрашиваемый имел вашу визитную карточку, что избавит его от неловкости и необходимости вспоминать ваше имя.
- 7) С самого начала эксперту должна быть гарантирована конфиденциальность. Результаты исследования, хотя и будут переданы руководству, но будут иметь обобщенный вид, без персонализации.
- 8) В начале интервью необходимо сразу же поблагодарить эксперта за что-либо, настроить его благожелательно.
- 9) Необходимо показать эксперту его важность и важность тех сведений, которые вы хотите получить от него.
- 10) Убедите эксперта, что интервью не экзамен, а результаты исследования не пойдут лично во вред эксперту.

11) Не надо и нельзя бояться эксперта!

12) Рекомендуется уточнять мысли эксперта, использовать термины привычного ему языка. Не используйте собственные специфические термины. Это *вы* должны вжиться в предметную область эксперта, а не он в вашу. Рекомендуется помогать эксперту связывать мысли в логические последовательности (например, «как вы ранее сказали...», «правильно ли я понял, что...», «если ..., то ..., не правда ли?»).

13) Определите зону компетентности эксперта, обратите внимание на то, какие выражения предпочитает использовать эксперт («я слышал», «я знаю», «я делал»). Сопоставляйте логичность его заявлений, их общезначимость.

14) Отличайте сведения субъективные, искаженные, запрограммированные от истинных.

15) В интервью должны присутствовать паузы, общее одобрение. Полезно подводить промежуточные итоги («итак, как вы сказали...»).

16) Вопросы надо ставить просто и прямо, соблюдая связность мысли: «Как ранее вы сказали...»

17) Дайте возможность эксперту уяснить установки интервью. Различайте установки организации и личные мотивы эксперта.

18) Следует учитывать невербальные знаки, исходящие от эксперта.

19) Следует самому влиять на движение интервью невербальными методами (поза, ладони, взгляд, движения, зоны).

20) Определите тип эксперта («застенчивый», «агрессивный», «самоуверенный», «неразговорчивый», «интеллектуал») и компенсируйте его недостатки (см. ниже).

21) Динамика опроса: разогрев, «медовый месяц», самоутверждение, адаптация, сотрудничество, эпилог, релаксация (см. ниже).

Для аналитика важной составляющей интервью является хорошее владение невербальной коммуникацией.

Невербальная коммуникация

Интервьюеру рекомендуется не только слушать эксперта, но и наблюдать за ним, за его жестами, мимикой. Около 55% информации передается невербальными средствами при помощи мимики, жестов, кивков, тона и силы голоса, наличия юмора и т. д. С методикой невербальной коммуникации вы можете обстоятельно познакомиться в книге Аллана Пиза «Язык телодвижений», но некоторые сведения из этой книги мы приведем далее.

Жесты совершаются обычными людьми неосознанно (исключая профессиональных актеров, политиков, преподавателей) и могут многое сказать наблюдателю об истинном положении вещей. Но следует иметь в виду:

- жест с возрастом меняется, приглушается (например, прикрывание ладонью рта во время лживого утверждения с возрастом превращается в едва уловимое взглядом касание рта пальцем);
- жест зависит от природных условий (потирание ладоней в условиях компании или на морозе означает разное);

- жест зависит от истории и культуры народа (например, жест «о'кей» в разных культурах соответствует разным информационным сообщениям);
- жест зависит от контекста разговора.

Необходимо научиться читать «язык» тела: следить, в каком положении находятся руки (скрещены, за спиной, «лодочкой»...), ладони (открыты, в карманах, сжаты в кулаки, в «замке», в «шпиле»). О многом говорят: положение глаз, наклон головы, положение (угол) ступней, разворот тела и т. д.

К примеру, когда человек искренен, то ладони открыты, при обратном — ладони спрятаны. Когда человек критично относится к вам, ладони могут быть в карманах, руки скрещены или совсем за спиной. Можно сказать, что жесты упорядочены по силе психологического воздействия.

Следует обращать внимание на взаиморасположение и расстояние между людьми. Вокруг человека есть пространственная зона собственности. До 15 см. располагается свержинтимная зона, область физического контакта; в диапазоне от 15 до 50 см. располагается интимная зона, в которую допускаются только супруг, дети, и вторжение в которую постороннего вызывает непроизвольную реакцию тела, физиологические реакции (выброс в кровь адреналина), ведущие к повышению боеготовности. В личную зону (от 50 до 120 см.) допускаются люди при определенных условиях, например, на вечеринке, при разговоре. В социальной зоне (от 1.2 до 3.6 м.) допустимо располагаться слесарю, почтальону... Общественная зона расположена на расстоянии от 3.6 метра и далее.

Примеры: эффект толпы (агрессия возрастает при сокращении зоны собственности и падает при расчленении толпы), протыкание зоны (указательным пальцем, подрезание траектории движения автомобиля).

Если нарушение размеров зоны неизбежно (например, в автобусе, лифте), восстановление зоны происходит другими средствами (избегание прямого взгляда в глаза, беспристрастность взгляда).

Нейтральное расстояние между людьми — около 90 см., меньше — человек пытается что-то доказать, больше — человек хочет уйти.

Используя подобные сведения, можно получать дополнительную информацию во время интервью. Невербальная коммуникация подразумевает не только фиксацию дополнительной информации, но и управление ситуацией, воздействие на человека. Такими приемами можно подавить человека, усмирить агрессора, расположить его к себе, разговорить.

Инструменты изменения положения («статус-кво») (примеры).

- Использование глаз. Попробуйте пристально на кого-нибудь посмотреть; человек смутится, обратит на вас внимание, замолчит, насторожится. Различие реакций зависит от разности психологических потенциалов двух людей.
- Рассмотрите разницу в рукопожатиях: поданная открытая ладонь, перевернутая ладонь, перехват руки на запястье и энергичное встряхивание, захват руки двумя руками («перчатка»), подача кончиков пальцев, рукопожатие на отлете длинной рукой, пожатие локтя.

- Перемещение по комнате. Зайдите человеку за спину, встаньте напротив глаза в глаза, встаньте сбоку, под 45 градусов.
- Отзеркаливание эксперта (повторение его жестов).
- Изменение расстояния между людьми (вход в личную зону), положение за столом.
- Проведите беседу с интервьюируемым в кресле, на стуле, на табурете. Используйте свет. Взаимное расположение тел. Разница очевидна.

В *вербальной коммуникации* примеры приемов для управления:

- наводящие вопросы;
- паузы (я жду...);
- повтор вопросов, настойчивость;
- эхо последних слов;
- уточнение («...что вы под этим понимаете...»);
- парафраз («...правильно ли я понял...»).

В последнее время получило развитие так называемое нейролингвистическое программирование (сокращенно НЛП), технология, позволяющая манипулировать человеком, используя средства его подсознания.

Следует отдавать себе отчет в проблемах интервью. Осознавая эти проблемы, вы поможете себе в их решении. Неформальные фразы могут носить двусмысленный характер, сложность фразы вуалирует ее смысл, отсутствие разъяснений, разный уровень очевидности терминов, логики, фактов, различная интерпретация ответов, телеграфный стиль общения, директивный стиль общения, неадаптированность терминов, нераспознавание мотивов, нюансов, оттенков, несоблюдение временных рамок...

Приведем ниже некоторые рекомендации по ведению интервью в зависимости от типа эксперта.

Тип эксперта:

- застенчивый — разговаривать следует неспеша, но без больших пауз, вопросы должны быть косвенные, рекомендуются поощрения; старайтесь не давить своим авторитетом, знаниями; попытайтесь выяснить его отношение к уже готовой точке зрения, помогайте ему;
- агрессивный — надо быть настойчивым, выяснять допускаемые им противоречия, указывая на них, можно провоцировать его к более динамичному развитию событий;
- самоуверенный — рекомендуется играть роль ученика;
- неразговорчивый — личная доброжелательность, «просто расскажите мне о...»;
- интеллеktуал — лучше сразу сузить тему, необходимо чаще прерывать вопросами, следует сразу «раскрыть карты».

Этапы интервью (динамика опроса):

- разогрев — определяет нормы интервью, в каком духе оно будет происходить;
- «медовый месяц» — общие фразы с обязательными поощрениями; цель — разговорить собеседника, войти к нему в доверие;

- самоутверждение — необходимо определить роль, которую вы будете играть во время интервью: «ученик», «спорщик»;
- адаптация — идет поиск пути к согласию;
- сотрудничество — нормальный диалог, нормы выровнены; основной и самый продуктивный этап работы;
- эпилог — признаком служит выход на мелкий уровень информации, включение защиты, плавный выход из интервью;
- релаксация — обязательно скажите эксперту о том, что если он позже вспомнит еще какие-либо важные факты, чтобы он обязательно сообщил вам об этом.

Генерация идей

Этапы стадии генерации идей:

- осознание беспокойства по поводу некоторой проблемы, формулировка проблемы;
- накопление материала;
- озарение;
- техническое решение;
- борьба за идею.

Ускорителем процесса является технология генерации идей. Технология необходима для обеспечения гарантированного результата процесса генерации идей. На сегодняшний день используются следующие технологии:

- мозговой штурм (автор А. Осборн);
- технология АРИЗ (автор Г. С. Альтшуллер);
- формализация поиска (авторы Коллер, Половинкин и др.).

На практике часто используют несколько технологий одновременно.

Правила проведения мозгового штурма

Технология предполагает наличие коллектива экспертов и свободного обмена идеями внутри него. Для того, чтобы задача была решена, для эффективной работы технологии от вас требуется сформировать у коллектива сильный побуждающий мотив («жизнь или смерть», серьезное поощрение, известность). Критика идей во время мозгового штурма запрещается (анализ этих идей осуществляется только по окончании штурма), поощряется развитие идеи, цепочка мыслей, приветствуется появление даже безумных на первый взгляд идей. Обязательно фиксируйте на носителе все выдвинутые идеи. Смысл мозгового штурма заключается в наведении новой идеи одного участника от идеи другого участника — в отличие от ограниченности мышления каждого из членов коллектива в отдельности. Тем самым существенно расширяется поле поиска идей для рассмотрения.

На рис. 35.1 показана схема штурма (для трех участников А, Б, В). Специальности у штурмующих разные (условно это показано тремя разными окружностями), поэтому пробы не так привязаны к вектору инерции ВИ, как обычно. К тому же правила штурма стимулируют «генерирование» смелых и даже фантастических идей: штурмующие выходят за пределы узкой специальности — а именно там, за этими пределами, и лежат решения высших уровней.

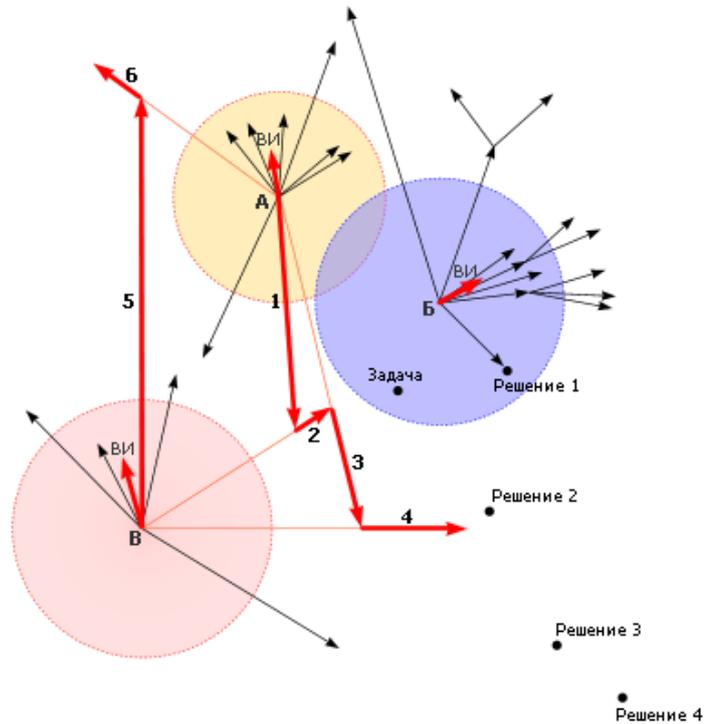


Рис. 35.1. Схема взаимодействия участников в методе мозгового штурма и процесс поиска решений

На схеме отражен еще один важный механизм штурма — взаимодействие и развитие идей. Участник А высказал идею 1, ее тут же видоизменил В — возникла идея 2. Теперь А иначе видит свою идею, это позволяет продолжить ее развитие (стрелка 3). Образуется цепь идей 1-2-3-4, направленная к решению второго уровня. Правда, механизм подхватывания идей иногда столь же последовательно (цепь 5-6) ведет и в сторону от решения.

Технология АРИЗ (алгоритм решения изобретательских задач)

Технология АРИЗ — поэтапная методика постановки цепи проблем и вопросов, отвечая на которые, аналитик приходит к решению проблемы. В настоящее время методика дополнена фондом приемов решения проблем, наиболее часто встречающихся в практике, и фондом физических эффектов.

Фонд эвристических приемов (эвристик) содержит до 200 типовых приемов. Основная идея — устранение физических противоречий путем их обхода; нахождение аналогий в прошлом опыте человечества. Примеры: заранее подложить противодействие (введение противоядия до укуса, прививка); обратить недостаток в достоинство усилением силы недостатка (вредная обычно вибрация полезна для разрушения твердых пород); использование посредника; использование симметрии и т. д. См. также статью «Межотраслевой фонд эвристических приемов преобразования объекта» из электронного учебника «Модели и методы искусственного интеллекта».

Полезным для данной технологии является определение уровня сложности задачи.

По статистике, задач 1 уровня — 32%. Это задачи, решение которых находится в пределах профессии. Обычно решение достигается за перебор от 1 до 10 вариантов. Требуется небольшого изменения свойств объекта (напри-

мер, смена материала на более дешевый дала замечательные изобретения — одноразовая бритва, одноразовая посуда, одноразовый подгузник).

Задач 2 уровня — 45%. Решение задач 2 уровня находится в пределах одной отрасли. Обычно решение достигается перебором от 10 до 100 вариантов. Требуется выбора одного объекта из нескольких (например, выбор компоновки двигателя).

Задач 3 уровня — 19%. Решение находится в пределах одной науки. Обычно решение достигается перебором от 100 до 1000 вариантов. Требуется частичного изменения выбранного объекта (например, использование взрыва в жидкости для перекрытия луча света между пластинами).

Задач 4 уровня — 4%. Решение находится в смежных науках. Обычно решение достигается перебором от 1000 до 10 000 вариантов. Требуется создания нового объекта (например, запаховый метод контроля сварки).

Задач 5 уровня — 0.3%. Решение находится за пределами современной науки. Обычно решение достигается перебором от 10 000 до 1 000 000 вариантов. Требуется создания нового комплекса объектов (например, ракетостроение, электроискровая обработка металлов).

Решите самостоятельно следующие задачи различных уровней.

Задание 1 уровня. Требуется изготовить стеклянный (инертный в химическом отношении) куб со стороной 1 метр со множеством вертикальных каналов (капилляров) для пропускания жидкости. Противоречие: хрупкость, размеры отверстия, толщина стенок, длина канала, множественность каналов.

Задание 2 уровня. Предложите, в чем хранить универсальный растворитель. Противоречие: растворитель растворяет все. Но хранить его в чем-то надо.

Задание 3 уровня. Требуется наладить производство окрашенных деревянных досок. Противоречие: краска проникает неглубоко в древесину и трескается, требует ручной работы, неравномерно ложится.

Формализация поиска

Применение определенного языка (графического, геометрического, таблиц, деревьев) для описания проблемы и автоматическое оперирование описанием для нахождения решений. К данному разделу относятся методики информационного моделирования (объектно-ориентированного, функционального). См. также лекцию «Внутреннее представление данных и проблема проектирования» из учебника «Модели и методы искусственного интеллекта».

При генерации идей следует осознавать помехи, оказывающие отрицательное влияние на продуктивность работы, для того, чтобы далее компенсировать их. Перечислим некоторые из них: привычки, лень, излишняя напряженность, боязнь показаться смешным или подвергнуться критике в глазах других людей, слабая целеустремленность, боязнь начала («синдром понедельника»), отсутствие ресурсов, излишняя серьезность.

При реализации вышеизложенных процедур следует иметь в виду, что есть определенные правила формирования коллектива. Коллектив должен быть неоднороден, в нем должен быть лидер, генератор идей, фиксатор идей,

критик, исполнитель, библиотекарь. В коллективе должны быть сформулированы мотивы. При стимулировании коллектива следует иметь в виду закономерности, которые наблюдаются между результатом и силой стимула, для того, чтобы управлять ими (см. рис. 35.2).

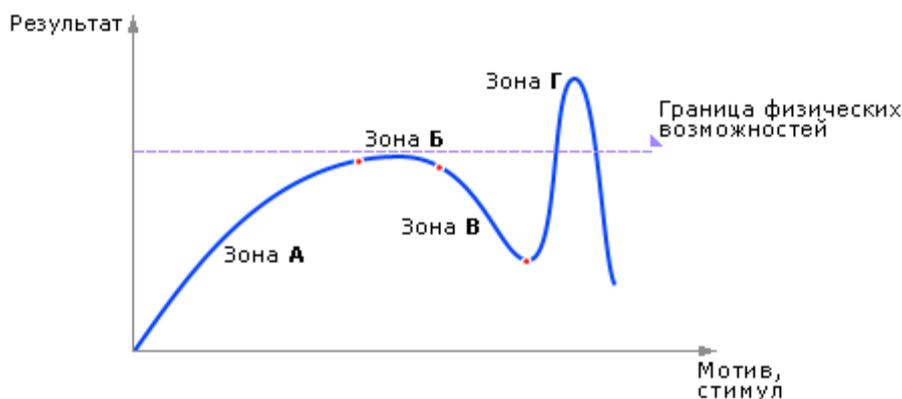


Рис. 35.2. Зависимость результатов индивидуальной деятельности человека от степени ее мотивации (зависимость снята коллективом психологов под руководством Леонтьева)

Рост стимулов, мотивированность деятельности (**зона А** на рис. 35.2) ведет к росту результатов. Дальнейшее стимулирование (**зона Б**) не ведет к росту результатов в силу естественных ограничений возможностей человека. Дальнейшее наращивание мощности стимула (**зона В**) сковывает возможности человека; человек, которому поставлены слишком сильные стимулы, может наоборот вести себя скованно и в результате не сумеет достичь даже тех результатов, которые бы он мог достигнуть в нормальной обстановке. И, наконец, в условиях стресса (**зона Г**) происходит резкое возрастание результатов деятельности человека. Стресс вызывает мобилизацию всех человеческих сил (в том числе и подсознательных) для решения проблемы. Следует иметь в виду, что возвращение организма (восстановление) к нормальному функционированию после стресса — процесс достаточно медленный. Цепь быстро чередующихся стрессов может вызвать необратимые последствия в организме человека, так как возможен переход через критический порог истощения ресурсов, и даже смерть.

36. ЭКСПЕРТИЗА

В процессе неформального синтеза после выдвижения множества идей необходимо окончательно принять к реализации какую-то одну из них, поскольку исследовать до конца детально все идеи довольно сложно. Чтобы отобрать лучшую (лучшие) из них, проводят их **экспертизу**.

Для этого коллектив экспертов анализирует идеи, и каждый эксперт выставляет каждой идее свою собственную оценку. Условиями проведения экспертизы является **коллектив экспертов и перечень оцениваемых идей**. Результатом первого этапа экспертизы является матрица оценок. Далее с помощью определенного выбранного метода учитывают совместно все эти оценки и выбирают лучшую идею.

К экспертизе прибегают, когда необходимо сделать выбор из множества возможных идей, решений, проектов. Экспертиза — целесообразный отбор одного решения из множества возможных, полученных на предыдущем этапе «Генерация идей» (см. лекцию 35. Неформальный синтез).

Существует много методов проведения экспертизы: голосование, ранжирование, попарное сравнение, разбиение на классы. Рассмотрим некоторые из них.

Голосование

Голосование можно применять только в том случае, если мнений, которые предложены к рассмотрению, меньше, чем собравшихся.

Примером, иллюстрирующим этот тезис, является решение профкома о выделении квартиры. Представитель каждого отдела предприятия, участвующий в заседании профкома, если он честен, должен голосовать только за своего представителя. Таким образом, никто и никогда не наберет большинство голосов. Задача попросту не решается. Или решается нечестными способами, например, сговором двух представителей, попеременно голосующих то за отдел А, то за отдел Б, и побеждающих всегда двумя голосами против остальных.

Простейший способ голосования — выбор большинством голосов. Он имеет свои достоинства (простота) и недостатки (решение принимается неквалифицированным большинством).

Данный способ можно существенно улучшить.

1. Голосование большинством 2/3. Для принятия решения требуется набрать 2/3 голосов. Такое правило применяется в особо ответственных случаях, повышая порог принятия решения. Но при этом метод часто приводит в тупик, так как необходимого количества голосов набрать какому-то одному из проектов достаточно трудно.
2. Голосование с правом вето. В этом случае любой эксперт имеет право заблокировать общее намечающееся решение одним своим голосом, если решение ему не выгодно или опасно. Это еще более жесткая процедура голосования. Способ применяется в сверхответственных ситуациях (например, принятие решений в ООН).

3. Талонное голосование. Голосовать можно только при наличии талона. Талонов меньше, чем голосований в течение оговоренного отрезка времени. Например, на 12 совещаний в год каждому эксперту выдается 8 талонов. Таким образом, если эксперт чувствует себя не вполне компетентным по какому-то вопросу, он не вмешивается в процесс голосования, сохраняя талон для более важного для него вопроса. Метод существенно улучшает процедуру голосования, заставляя ответственно оценивать не только проекты, но и себя.
4. Туровое голосование. В первом туре голосования отбирается как результат 2-3 проекта, которые набирают большинство голосов или проходят установленный порог. Во втором туре все эксперты голосуют снова, но только уже за эти 2-3 отобранных проекта. В качестве примера можно привести выборы президента в России или во Франции. Недостатком такого голосования является то, что эксперт во втором туре вынужден голосовать не за то решение, которое ему нравится, а за одно из тех, которые прошли во второй тур. Возникает иллюзия выбора.
5. Туровое голосование можно существенно улучшить, если дать возможность экспертам знакомиться с результатами голосования после каждого тура и не убирать проекты из матрицы от тура к туру. В результате такой процедуры эксперты, видя и оценивая общий результат, могут менять собственные предпочтения, исключая случайность и приходя к равносному мнению. Этот метод получил название «метод Дельфы».

Пример турового голосования. Пусть имеется семеро кандидатов на получение премии: Дмитриев (Д), Иванов (И), Кузнецов (К), Лукьянов (Л), Михайлов (М), Петров (П), Сидоров (С). Премию могут получить лишь трое лучших. В первом туре каждый *эксперт* (в примере их пятеро) оценивает каждого из кандидатов и ставит его на то или иное место (см. табл. 36.1).

Таблица 36.1.
Распределение мест проектов, данное коллективом экспертов (первый тур)

	1	2	3	4	5	6	7
Эксперт 1	И	П	С	Л	К	М	Д
Эксперт 2	Д	И	М	Л	К	С	П
Эксперт 3	И	М	К	Л	С	П	Д
Эксперт 4	Д	С	И	Л	П	М	К
Эксперт 5	И	К	П	Л	С	Д	М

Далее подсчитываются очки. Трое экспертов присвоили Иванову первое место, один эксперт дал ему второе место и еще один — третье. Зная это, подсчитываем общее число очков для Иванова: $3 \cdot 1 + 1 \cdot 2 + 1 \cdot 3 = 8$. Таким же образом подсчитываются очки для всех остальных кандидатов. Результаты сводятся в таблицу (см. табл. 36.2).

Таблица 36.2.
Итоги первого тура

Кандидат	Очки	Место
----------	------	-------

Иванов	$08 = 3 \cdot 1 + 1 \cdot 2 + 1 \cdot 3$	1
Лукьянов	$20 = 5 \cdot 4$	2
Сидоров	$21 = 1 \cdot 2 + 1 \cdot 3 + 2 \cdot 5 + 1 \cdot 6$	3
Дмитриев	$22 = 2 \cdot 1 + 1 \cdot 6 + 2 \cdot 7$	4
Кузнецов	$22 = 1 \cdot 2 + 1 \cdot 3 + 2 \cdot 5 + 1 \cdot 7$	4
Петров	$23 = 1 \cdot 2 + 1 \cdot 3 + 1 \cdot 5 + 1 \cdot 6 + 1 \cdot 7$	5
Михайлов	$24 = 1 \cdot 2 + 1 \cdot 3 + 2 \cdot 6 + 1 \cdot 7$	6

Чем меньше очков у кандидата, тем более высокое у него место. Это непосредственно следует из принципа подсчета очков. Как видим, после первого тура призовые места достаются Иванову, Лукьянову и Сидорову.

После опубликования результатов выявляется и недостаток голосования, так как часть экспертов не удовлетворена результатами. Например, кандидат Лукьянов (см. табл. 36.1), которому никто из экспертов не давал ни одного из призовых мест, ибо никто не видел Лукьянова как достойного кандидата на премию, получил второе место, а вместе с ним и премию. Второй эксперт недоволен тем, что Сидоров вошел в тройку призеров. И если бы он голосовал снова, то он бы поменял Сидорова и Петрова местами, заставив опуститься Сидорова в общем рейтинге. Так как есть эксперты, заинтересованные в изменении результатов (путем перестановки проектов), то имеет смысл провести второй тур.

Посмотрим, что будет делать первый эксперт во втором туре: скорее всего, он поменяет местами Петрова и Иванова, переставит Сидорова на второе место, так как ему «наступают на пятки» конкуренты Дмитриев и Кузнецов, а Иванова, которому ничего не угрожает, поставит на третье место.

Так же, но по-своему, поступят другие эксперты. Поэтому результаты голосования от тура к туру будут меняться.

Например, после второго тура матрица оценок может выглядеть так (см. табл. 36.3).

Таблица 36.3.
Распределение мест проектов, данное коллективом экспертов (второй тур)

	1	2	3	4	5	6	7
Эксперт 1	П	С	И	Л	К	М	Д
Эксперт 2	Д	И	М	Л	К	П	С
Эксперт 3	И	М	К	Л	С	П	Д
Эксперт 4	Д	С	И	Л	П	М	К
Эксперт 5	И	К	П	Л	С	Д	М

Кроме таблицы результатов удобно вычислять таблицу неудовлетворенности экспертов. Например, подсчитывая сумму разностей по каждому эксперту между желательным распределением проектов по местам и общим результатом тура. Если эксперт действует логично, показатель неудовлетворенности от тура к туру, в принципе, должен уменьшаться. Хотя нулевым он станет вряд ли, так как неудовлетворенность одного эксперта уменьшается за счет увеличения неудовлетворенности другого. Поэтому процедура должна

прийти к равновесному состоянию за счет действия всех экспертов одновременно, понижающих собственные показатели неудовлетворенности.

Туровая процедура продолжается до тех пор, пока ни один эксперт в одиночку не в силах будет изменить получившийся порядок (результат). Какие-то эксперты могут остаться неудовлетворенными, изменить результат голосования они не смогут. Технически это означает, что результат в нескольких последовательных турах перестанет меняться. Его и принимают в качестве ответа.

Рассмотрим второй класс методов экспертизы.

Ранжирование

Ранжирование есть упорядочение, разбиение множества на элементы с введением между ними некоторого порядка, в данном случае порядка «лучше-хуже».

Пусть четверо экспертов расставляют проекты А, В, С и D по рангам (местам), см. табл. 36.4.

Таблица 36.4.
Распределение мест проектов, данное коллективом экспертов

	А	В	С	Д
Эксперт 1	1	2	3	4
Эксперт 2	1	3	2	4
Эксперт 3	2	1	3	4
Эксперт 4	1	3	4	2

Ранжирование > Метод «Медиана Кемени»

Вычисляются значения r_{ij} — расстояние между мнениями i -го и j -го экспертов. Вычислять расстояние между двумя векторами (вектор — это мнение эксперта по кортежу проектов) можно по-разному. Расстояние в математике называется **нормой**. Есть пифагорова норма, манхэттенская, Минковского и т. д. Например, вычислим в матрице (см. табл. 36.4) расстояние между мнениями эксперта 1 и эксперта 2 по *манхэттенской норме*:

$$r_{12} = |1 - 1| + |2 - 3| + |3 - 2| + |4 - 4| = 0 + 1 + 1 + 0 = 2.$$

Расстояние мнений по манхэттенской норме — это сумма разностей (по абсолютной величине) мнений двух экспертов по каждому проекту.

Значения заносятся в таблицу. Затем вычисляются значения R_i — расстояние от i -го эксперта до всех остальных — как сумма расстояний между мнением этого эксперта и остальных экспертов (см. табл. 36.5).

Таблица 36.5.
Расчет расстояний между мнениями экспертов

	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	R_i
Эксперт 1	0	r_{12}	r_{13}	r_{14}	$0 + r_{12} + r_{13} + r_{14} = R_1$
Эксперт 2	r_{21}	0	r_{23}	r_{24}	$r_{21} + 0 + r_{23} + r_{24} = R_2$
Эксперт 3	r_{31}	r_{32}	0	r_{34}	$r_{31} + r_{32} + 0 + r_{34} = R_3$
Эксперт 4	r_{41}	r_{42}	r_{43}	0	$r_{41} + r_{42} + r_{43} + 0 = R_4$
					$\min(R_1, R_2, R_3, R_4)$

Среди вычисленных суммарных расстояний ищется минимум из R_i . Мнение соответствующего i -го эксперта (с минимальным R) является самым средним и объявляется результатом экспертизы. Геометрическая интерпретация показывает, что мнение данного эксперта находится примерно в центре многогранника, образованного вершинами, каждая из которых представляет собой мнение отдельного эксперта, а ребра — расстояния между мнениями экспертов. Как известно, центр — место, наиболее приближенное ко всем вершинам (мнениям) фигуры одновременно.

Система координат для построения многогранника образована проектами. Таким образом, 4 проекта образуют 4-хмерное пространство. По осям координат откладывают баллы, полученные проектом у данного эксперта. Тогда мнение эксперта — точка в пространстве проектов.

На рис. 36.1 показано четырехмерное пространство, образованное проектами ABCD. Чтобы найти точку — мнение эксперта — в четырехмерном пространстве, надо отложить вдоль оси А число баллов, поставленное экспертом проекту А. Далее провести через полученную точку линию, параллельную следующей оси В и отложить на этой линии число баллов, поставленное экспертом проекту В. Через полученную точку снова провести линию, параллельную оси С, отложив на ней число баллов, поставленное экспертом проекту С. И, наконец, через полученную точку провести линию, параллельную D, отложив на ней число баллов, поставленное экспертом проекту D. Точка, соответствующая мнению эксперта в четырехмерном пространстве, найдена. Так же следует найти точки — мнения экспертов 2, 3, 4. Соединив точки линиями, получим многогранник мнений экспертов.

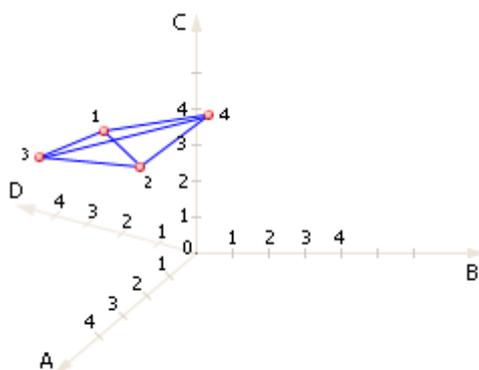


Рис. 36.1. Многогранник мнений экспертов 1, 2, 3, 4 в четырехмерном пространстве проектов ABCD

Вы можете построить любое количество осей и отразить на плоскости пространство любой размерности (см. лекцию 16. Понятие размерности пространства в дисциплине «Компьютерная графика»). В зависимости от того, насколько удачно вы расположите оси (точнее сказать, проекции осей) на плоскости, рисунок будет более или менее наглядным. Поворачивая оси, можно добиться наглядного изображения. Это называется найти удачный ракурс рассмотрения многомерной фигуры. Рассматривая рис. 36.1, обратите внимание на то, что точка 1 (символизирует мнение эксперта 1 по поводу

проектов A, B, C, D) находится ближе всех к мнениям остальных экспертов, в отличие от точек 2, 3, 4. Рассматривая многогранник с разных сторон (поворачивая оси) можно убедиться, что так точка выглядит с любого ракурса. Значит, мнение эксперта 1 — среднее и является решением.

Можно изображать эту фигуру и в трехмерном пространстве, и двухмерном, и даже одномерном или нульмерном (см. рис. 36.2), что является для вас более привычным. Такие рисунки являются сечениями четырехмерного пространства по одной или двум осям. Это вопрос удобства мышления. Обычно рисуют несколько сечений, желательно во взаимоперпендикулярных плоскостях, чтобы попытаться восстановить общий вид фигуры по нескольким частным видам в уме, разглядывая их одновременно. Так поступают при работе с чертежами.

На рис. 36.2 показано сечение четырехмерного пространства для трех осей (A, B, C) и различных пар двух осей. (Проекция на одномерные оси и на нульмерную ось постройте сами). Но и из этих рисунков видно, что мнение эксперта 1 является средним, так как всегда лежит внутри многогранника (1-2-3-4).

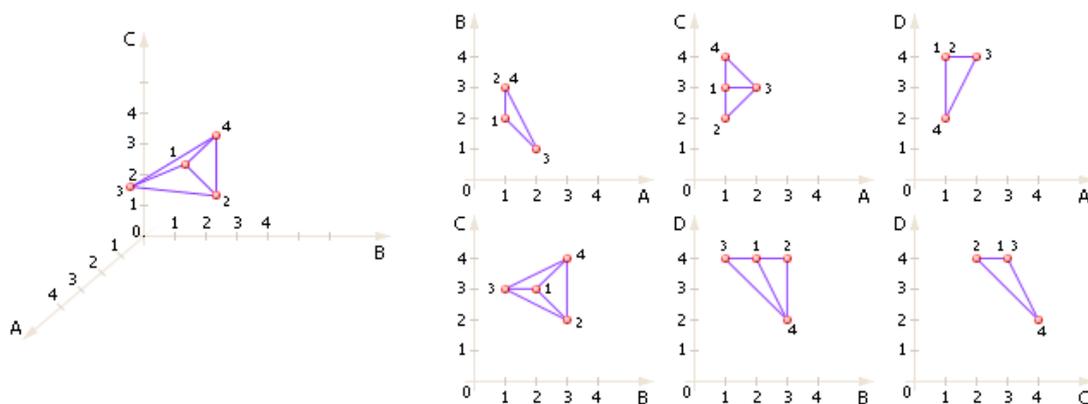


Рис. 36.2. Проекция многогранника мнений экспертов на оси проектов ABCD

Самая удаленная от центра вершина характеризует самое ненадежное мнение и может этим характеризовать самого эксперта.

В нашем примере матрица расстояний будет выглядеть так (см. табл. 36.6).

Таблица 36.6.

Расчет значений расстояний между мнениями экспертов

	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	R_i
Эксперт 1	0	2	2	4	8
Эксперт 2	2	0	4	4	10
Эксперт 3	2	4	0	6	12
Эксперт 4	4	4	6	0	14

Очевидно, что наименьшее расстояние от остальных мнений — у эксперта 1, поэтому его мнение объявляется результатом.

Ответ: проект А — 1 место, проект В — 2, проект С — 3, проект D — 4. Наилучшим проектом является проект А.

Ранжирование > Метод большинства

На первое место ставится тот проект, которому большинство экспертов присвоило первое место. На второе место ставится тот проект, которому большинство экспертов присвоило второе место и так далее.

В нашем примере (см. табл. 36.4) проект А займет в итоге первое место, так как трое экспертов отдали ему первое место. Проекты В и С поделят второе и третье места, так как двое экспертов отдали им третье место. Проект D займет четвертое место, так как три эксперта отдали ему четвертое место.

Ответ: проект А — 1 место, проект В и С делят 2 и 3 место, проект D — 4 место. Наилучшим проектом является проект А.

Можно уточнить решение задачи, разрешив спор за 2 и 3 место в пользу проекта В. Если посмотреть внимательно, сравнивая окончательно между собой проекты В и С, то проект В более предпочтителен, чем С, поскольку занимает дополнительно места 2 и 1, а проект С — только места 2 и 4.

Ответы задачи, полученные методом «Медиана Кемени» и методом большинства, совпали, хотя это происходит не всегда. Решения, полученные разными методами, могут слегка отличаться друг от друга. Но в целом решение устойчиво. В ряде случаев отдельные методы могут не сработать, поэтому экспертизу проводят несколькими методами, а затем выводят общий ответ.

Ранжирование > Альтернатива Кондорсе

Снова рассмотрим наш пример (табл. 36.7).

Таблица 36.7.

Распределение мест проектов, данное коллективом экспертов

	А	В	С	D
Эксперт 1	1	2	3	4
Эксперт 2	1	3	2	4
Эксперт 3	2	1	3	4
Эксперт 4	1	3	4	2

Для каждой пары проектов (i и j) подсчитаем, какое количество S_{ij} экспертов считает, что i -ый проект лучше, чем j -ый, и какое количество S_{ji} экспертов считает наоборот (см. табл. 36.8).

Таблица 36.8.

Расчет отношений «лучше-хуже» между проектами

S_{ij}	Отношение	S_{ji}	Пояснение
$S_{AB} = 3$	>	$S_{BA} = 1$	А лучше В
$S_{AC} = 4$	>	$S_{CA} = 0$	А лучше С
$S_{AD} = 4$	>	$S_{DA} = 0$	А лучше D
$S_{BC} = 3$	>	$S_{CB} = 1$	В лучше С
$S_{BD} = 3$	>	$S_{DB} = 1$	В лучше D
$S_{CD} = 3$	>	$S_{DC} = 1$	С лучше D

Далее на графе стрелками покажем отношения между парами проектов (см. рис. 36.3). Стрелка от i -го проекта в графе к j -му указывает, что i -му проекту отдали предпочтение больше экспертов, чем j -му. Например, стрелка от А направлена к В, так как проект А лучше проекта В (трое экспертов предпочли проекту В проект А, и лишь только один эксперт — наоборот).

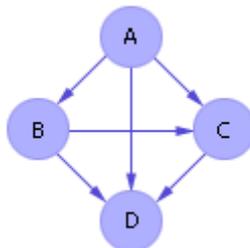


Рис. 36.3. Граф предпочтений проектов ABCD

Упорядочим граф, приведя его к ярусно-параллельной форме (ЯПФ) (см. рис. 36.4).

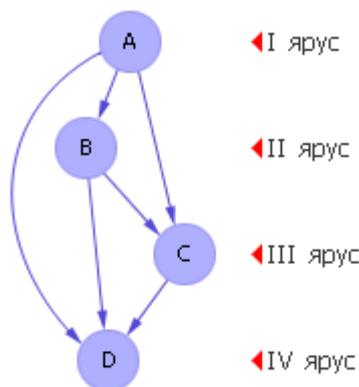


Рис. 36.4. Граф предпочтений ABCD в ярусно-параллельной форме

Приведение графа к ЯПФ есть приведение его к такой форме, где в верхних ярусах находятся вершины, в которые нет входящих стрелок из вершин, лежащих в нижних ярусах. Алгоритм построения ЯПФ графа разбирается в дисциплине «Дискретная математика». Итак, ЯПФ показывает, что вершиной, в которую нет стрелок, является вершина А, следовательно, абсолютным лидером среди проектов является проект А. Проект А лучше В, одновременно он лучше С, одновременно он лучше D. Далее, среди оставшихся проектов (BCD) проект В лучший, так как он одновременно лучше С и D. И, наконец, С лучше D.

Ответ: проект А — 1 место, проект В — 2 место, проект С — 3 место, проект D — 4 место. Наилучшим проектом является проект А.

Ранжирование>Принцип Борда

За первое место каждому проекту дадим 1 балл, за второе место — 2 балла, за третье — 3 балла и так далее. Подсчитаем количество баллов, которое получит каждый проект в сумме. Первое место в итоге получит проект, который наберет наименьшее число баллов, остальные места определятся сортировкой набранных баллов.

Итак, для нашего примера из табл. 34.4 получим следующее (см. табл. 36.9).

Таблица 36.9.
Распределение мест проектов методом Борда (пример)

	A	B	C	D
Эксперт 1	1	2	3	4
Эксперт 2	1	3	2	4
Эксперт 3	2	1	3	4
Эксперт 4	1	3	4	2
Итого:	5	9	12	14

Ответ: проект А — 1 место (5 баллов), проект В — 2 место (9 баллов), проект С — 3 место (12 баллов), проект D — 4 место (14 баллов). Наилучшим проектом является проект А.

Ранжирование > Метод «Электра»

Рассмотрим данный метод на примере решения о покупке автомобиля.

Сначала составляют таблицу критериев, по которым будут оценивать проекты (см. табл. 36.10).

Таблица 36.10.
Таблица критериев для оценки проектов

Критерии	Вес критерия	Шкала	Код	Стремление
Цена	5	Менее \$11 тыс.	10	Min
		\$11–16 тыс.	15	
		\$16–22 тыс.	20	
		\$22–29 тыс.	25	
		\$29–37 тыс.	30	
Комфорт	4	Высокая	Ca	Max
		Средняя	Cb	
		Низкая	Cc	
Скорость	3	Большая	Va	Max
		Средняя	Vb	
Дизайн	3	Изысканный	La	Max
		Обычный	Lb	

Далее эксперт составляет таблицу оценок проектов (автомобилей). Например, для 7-ми автомобилей эксперт заполняет таблицу так, как показано в табл. 36.11.

Таблица 36.11.
Таблица оценок проектов по критериям

Проект	Значение критерия			
	Цена	Комфортность	Скорость	Дизайн
1	30	Ca	Va	La
2	25	Ca	Vb	La
3	25	Cb	Va	La
4	20	Cb	Va	Lb
5	20	Cb	Vb	La
6	20	Cc	Va	La

7	10	Cc	Vb	Lb
Вес:	5	4	3	3

Рассматриваем все пары проектов i и j . Если по какому-либо критерию i -ый проект лучше, чем j -ый, то соответствующий критерию вес прибавляется к P_{ij} (эти баллы символизируют выбор «За»), в противном случае — к N_{ij} (эти баллы символизируют выбор «Против»). То же самое справедливо для j -го проекта: если j -ый проект оказывается лучше, чем i -ый, то соответствующий критерию вес прибавляется к P_{ji} , в противном случае — к N_{ji} (обратите внимание на порядок следования индексов j и i у P и N). Если повстречалось одинаковое для i -го и для j -го проектов значение критерия, то оно пропускается. Затем, когда по паре i и j рассмотрены все критерии, находятся отношения $D_{ij} = P_{ij}/N_{ij}$ и $D_{ji} = P_{ji}/N_{ji}$. Значения $D \leq 1$ отбрасываются. Заметим, что $D_{ji} = 1/D_{ij}$ (и наоборот), таким образом, вычисления можно несколько упростить.

Рассмотрим, для примера, проекты 2 и 4 ($i = 2, j = 4$). По критерию «Цена» (вес критерия — 5 баллов) проект 2 хуже проекта 4; по критерию «Комфортность» (вес — 4 балла) проект 2 лучше проекта 4; по критерию «Скорость» (вес — 3 балла) проект 2 хуже проекта 4; по критерию «Дизайн» (вес — 3 балла) проект 2 лучше проекта 4. Таким образом, имеем:

$$P_{24} = 0 + 4 + 0 + 3 = 7;$$

$$N_{24} = 5 + 0 + 3 + 0 = 8;$$

$$D_{24} = P_{24}/N_{24} = 7/8 = 0.875 < 1 \text{ — отбрасываем};$$

$$P_{42} = 5 + 0 + 3 + 0 = 8;$$

$$N_{42} = 0 + 4 + 0 + 3 = 7;$$

$$D_{42} = P_{42}/N_{42} = 8/7 = 1/0.875 = 1.14 > 1 \text{ — принимаем.}$$

Рассмотрим проекты 1 и 2 ($i = 1, j = 2$). По критерию «Цена» проект 1 хуже проекта 2; по критерию «Комфортность» проекты 1 и 2 одинаковы, поэтому ничего не делаем; по критерию «Скорость» проект 1 лучше проекта 2; по критерию «Дизайн» проекты 1 и 2 одинаковы, поэтому ничего не делаем.

Таким образом, имеем:

$$P_{12} = 0 + 0 + 3 + 0 = 3;$$

$$N_{12} = 5 + 0 + 0 + 0 = 5;$$

$$D_{12} = P_{12}/N_{12} = 3/5 = 0.6 < 1 \text{ — отбрасываем};$$

$$P_{21} = 5 + 0 + 0 + 0 = 5;$$

$$N_{21} = 0 + 0 + 3 + 0 = 3;$$

$$D_{21} = P_{21}/N_{21} = 5/3 = 1/0.6 = 1.67 > 1 \text{ — принимаем.}$$

Все остальные пары рассчитываются аналогично.

Составляем матрицу, внося вычисленные (и принятые) значения D . Матрица имеет смысл предпочтений проектов между собой. Для нашего примера матрица выглядит следующим образом (см. табл. 36.12).

Таблица 36.12.

Полная матрица предпочтений проектов, составленная методом «Электра»

	1	2	3	4	5	6	7
1				1.4	1.4		2
2	1.67		1.33				1.4
3	1.25						2
4		1.14	1.67			1.33	1.4
5		1.25	1.67			1.33	1.4
6	1.25	2	1.25				1.2
7							

Задаем порог принятия решения, например $C = 1.33$, и оставляем в матрице те числа, которые больше или равны значению порога C . Таким образом, матрица разрезается (см. табл. 36.13).

Таблица 36.13.
Матрица предпочтений проектов при пороге $C = 1.33$

	1	2	3	4	5	6	7
1				1.4	1.4		2
2	1.67		1.33				1.4
3							2
4			1.67			1.33	1.4
5			1.67			1.33	1.4
6		2					
7							

По матрице строится граф предпочтений (см. рис. 36.5). Из графа, построенного по табл. 36.13, видно, что проект 1 лучше проектов 4, 5, 7; проект 2 лучше проектов 1, 3, 7; проект 3 лучше проекта 7; проект 4 лучше проектов 3, 6, 7; проект 5 лучше проектов 3, 6, 7; проект 6 лучше проекта 2.

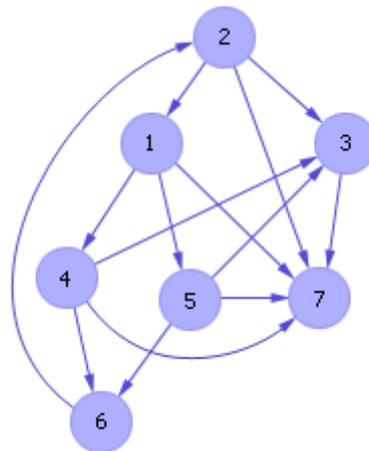


Рис. 36.5. Вид графа предпочтений для случая порога принятия решений $C = 1.33$

Очевидно, что решение не получено, так как в графе присутствуют петли. Например, 2 лучше 1, 1 лучше 5, 5 лучше 6, 6 лучше 2. Назначим порог отбора предпочтений $C = 1.4$ (это соответствует тому, что мы попробуем учесть только более сильные связи в графе, не отвлекаясь на малозначимые

расхождения в проектах). Таким образом, матрица еще разрезается. В ней остаются только самые сильные связи (см. табл. 36.14).

Таблица 36.14.

Матрица предпочтении проектов при пороге $C = 1.4$

	1	2	3	4	5	6	7
1				1.4	1.4		2
2	1.67						1.4
3							2
4			1.67				1.4
5			1.67				1.4
6		2					
7							

По матрице строится граф предпочтений (см. рис. 36.6). По графу и табл. 36.14 видно, что проект 1 лучше проектов 4, 5, 7; проект 2 лучше проектов 1, 7; проект 3 лучше проекта 7; проект 4 лучше проектов 3, 7; проект 5 лучше проектов 3, 7; проект 6 лучше проекта 2. Как видим из рис. 36.6, при $C = 1.4$ граф получился в таком виде, в котором он легко приводится к ЯПФ, следовательно, решение получено.

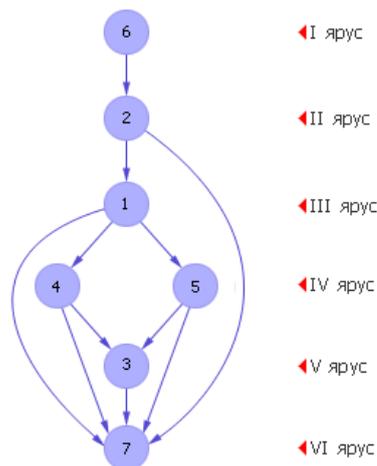


Рис. 36.6. Вид графа предпочтений для случая порога принятия решений $C = 1.4$

Петель в графе нет, при этом граф остался целостным. Решение говорит нам о том, что лучший проект — 6. На втором месте — проект 2, на третьем месте — проект 1, четвертое и пятое место делят проекты 4 и 5, на шестом месте — проект 3, на седьмом месте — проект 7.

1. Порог C подбирается эмпирически. При малом пороге C (в матрице много компонент), в графе содержится много петель. Увеличивая порог C , можно добиться уменьшения количества связей и устранения мало значащих связей и петель. Однако нужно иметь в виду, что при очень высоком значении порога C граф распадется на несколько отдельных графов из-за отсутствия необходимых связей; в качестве иллюстрации см. рис. 36.7: при пороге $C = 1.67$ проекты 4 и 5 остаются без входящих стрелок, и поэтому становится неясно, какой же проект — 6-ой, 4-ый или 5-ый — является наилучшим. А при пороге $C = 2$ граф распадается на четыре не связанных подграфа.

2. Если граф не будет найден ни при каких значениях порога C , то следует изменить веса критериев и всю процедуру провести снова.
3. Если разумное изменение весов не помогает, и проекты все еще путаются (петли в графе), то надо детализировать проблему, добавив критерии.

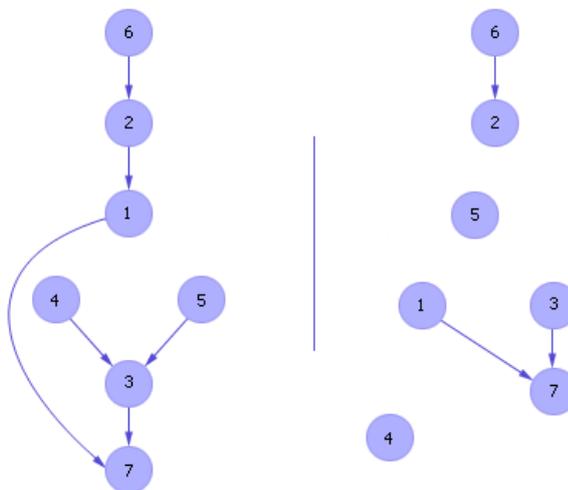


Рис. 36.7. Вид графа предпочтений для случая порога принятия решений $C = 1.67$ (слева) и $C = 2$ (справа)

Точность и доверие к результатам экспертизы. Оценка экспертов

Важным вопросом экспертизы является формирование группы экспертов. Если экспертов много, то в группу с высокой вероятностью попадают некомпетентные эксперты. Если экспертов мало, то результат экспертизы существенно зависит от конкретных лиц, попавших в число экспертов. Поэтому имеет смысл, опросив экспертов r_i и получив результат экспертизы R , оценить объективность каждого из членов экспертного коллектива по следующей формуле:

$$\alpha_i = 1 - \frac{d(r_i, R)}{\sum_{i=1}^n d(r_i, R)}$$

α_i — коэффициент объективности эксперта;

r_i — мнение i -го эксперта;

R — результат экспертизы (среднее мнение), вычисленное ранее изложенными способами;

$d(r_i, R)$ — расстояние от мнения i -го эксперта до среднего мнения R ;

n — количество экспертов;

$\sum_{i=1}^n d(r_i, R)$ — сумма расстояний от всех мнений экспертов до среднего мнения R .

Для примера можно просчитать объективность каждого эксперта. Рассмотрим уже полученное ранее решение (см. табл. 36.15).

Матрица мнений экспертов и результат экспертизы

	A	B	C	D
Эксперт 1	1	2	3	4
Эксперт 2	1	3	2	4
Эксперт 3	2	1	3	4
Эксперт 4	1	3	4	2
Результат экспертизы	1	2	3	4

Расстояние $d(r_i, R)$ от мнения i -го эксперта до результата (среднего мнения R):

$$d(r_1, R) = |1 - 1| + |2 - 2| + |3 - 3| + |4 - 4| = 0;$$

$$d(r_2, R) = |1 - 1| + |3 - 2| + |2 - 3| + |4 - 4| = 2;$$

$$d(r_3, R) = |2 - 1| + |1 - 2| + |3 - 3| + |4 - 4| = 2;$$

$$d(r_4, R) = |1 - 1| + |3 - 2| + |4 - 3| + |2 - 4| = 4.$$

$$\sum_{i=1}^4 d(r_i, R) = 0 + 2 + 2 + 4 = 8.$$

$$\alpha_1 = 1 - 0/8 = 1;$$

$$\alpha_2 = 1 - 2/8 = 0.75;$$

$$\alpha_3 = 1 - 2/8 = 0.75;$$

$$\alpha_4 = 1 - 4/8 = 0.5.$$

Первый эксперт имеет наибольшую объективность; наименьшую объективность имеет четвертый эксперт (см. рис. 36.8). Кстати, неслучайно в методе «Медиана Кемени» мнение первого эксперта было принято в качестве ответа экспертизы, а четвертый эксперт стоял на периферии многоугольника мнений экспертов ABCD (см. рис. 36.1 и рис. 36.2)...

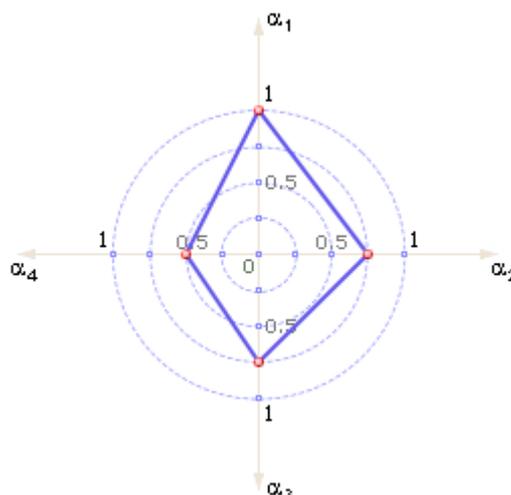


Рис. 36.8. Диаграмма объективности экспертов

Рекомендуется отсеять мнения экспертов с малым значением α и результат R пересчитать снова без их учета.

Вторая проверка экспертов касается согласованности их коллектива. Для расчета согласованности мнений рассчитаем дисперсию мнений (см.

табл. 36.16) и оценим ее величину статистической формулой $W = 12 \cdot D / (n^2 \cdot (m^3 - m))$, где n — число экспертов, m — число проектов.

Таблица 36.16.

Матрица расчета согласованности мнений экспертов для группы из четырех экспертов

	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	Баллы проекта	Отклонение балла проекта от среднего балла	Дисперсия D
Проект А	1	1	2	1	5	-5	25
Проект В	2	3	1	3	9	-1	1
Проект С	3	2	3	4	12	2	4
Проект D	4	4	4	2	14	4	16
Итого					40		46
Средний балл					10		

Для нашего случая имеем: $W = 12 \cdot 46 / (4^2 \cdot (4^3 - 4)) = 552 / 960 = 0.575$. При $W = 1$ эксперты находятся в полном согласии; при $W = 0$ наблюдается полная несогласованность. В нашем примере согласие экспертов слабое. Поэтому можно порекомендовать изменить число экспертов — либо удалить четвертого эксперта, как самого необъективного, либо увеличить число экспертов для большей статистической устойчивости результатов. Посмотрим, как возрастет согласованность экспертов, например, после удаления четвертого эксперта (см. табл. 36.17).

Таблица 36.17.

Матрица расчета согласованности мнений экспертов для группы из трех экспертов

	Эксперт 1	Эксперт 2	Эксперт 3	Баллы проекта	Отклонение балла проекта от среднего балла	Дисперсия D
Проект А	1	1	2	4	-2.5	6.25
Проект В	2	3	1	6	-1.5	2.25
Проект С	3	2	3	8	0.5	0.25
Проект D	4	4	4	12	4.5	20.25
Итого				30		29
Средний балл				7.5		

$W = 12 \cdot 29 / (3^2 \cdot (4^3 - 4)) = 0.64$ — как видим, согласие увеличилось.

Экспертиза помогает нам провести неформальный анализ в моделировании, отобрав из множества предложенных методами генерации идей наиболее перспективный вариант будущего проекта. В выбранный проект в дальнейшем будут вложены значительные средства для его детализации (процесс проектирования и моделирования — см. лекцию 01. Понятие моделирования. Способы представления моделей и Рис. 1.14. Схема процесса, способов и приемов моделирования (полный вариант)).

Для уверенного принятия решения, так как все методы экспертиз субъективны, используют сразу несколько методов и выводят общее решение.

Полученное решение обязательно следует проверить. Мы рассмотрели две проверки — на качество эксперта (объективность эксперта) и качество группы (согласованность группы).

Сама величина критерия, полученная в результате проверки, дает только относительную оценку. Рекомендуется наблюдать изменение величины в ре-

зультате изменения некоторого существенного фактора и добиваться наилучшего значения этой величины, сознательно управляя этим фактором (см., например, табл. 36.16 и табл. 36.17). Таким фактором чаще всего может служить величина группы, качественный состав группы экспертов, система критериев, состав проектов. Хорошие результаты дает применение итераций в экспертизе.

Следует иметь в виду, что никакое количество проверок, даже очень большое, не гарантирует абсолютной правильности выбора решения.

Задачи для студентов

Задача 1. Вычислить и изобразить графики в масштабе распределения температуры в стенке.

Алгоритм решения:

1. Плотность теплового потока:

$$q = \frac{\lambda_{\text{ср}}}{\delta} (t_{c_1} - t_{c_2}) \quad [\text{Вт/м}^2]$$

где $\lambda_{\text{ср}} = \lambda_0 \left(1 + \beta_\lambda \frac{t_{c_1} - t_{c_2}}{2}\right) \quad [^\circ\text{C}]$

2. Температура на любом расстоянии x от поверхности стенки

$$t_x = \sqrt{\left(\frac{1}{\beta_\lambda} + t_{c_1}\right)^2 - \frac{2qx}{\lambda_0 \beta_\lambda}} - \frac{1}{\beta_\lambda}$$

Задача 2. Определить потери теплоты с одного квадратного метра поверхности и температуры на внешних поверхностях стены печи из шамотного кирпича при известных температурах газа в печи и воздуха в помещении и коэффициента теплоотдачи. Для решения задачи использован метод последовательных приближений.

Алгоритм решения:

1. Задаем среднюю температурой стенки $\bar{t}_c = 650 \text{ } ^\circ\text{C}$;
2. Определяем значение коэффициентов теплопроводности шамотного кирпича:

$$\lambda_{\text{ср}} = 0,84(1 + 0,695 \cdot 10^{-3} \cdot \bar{t}_c)$$

3. Вычисляем коэффициент теплопередачи:

$$K = \frac{1}{\frac{1}{\alpha_1} + \frac{\delta}{\lambda} + \frac{1}{\alpha_2}} \quad [\text{Вт/м}^2 \text{ } ^\circ\text{C}]$$

4. Определяем плотность теплового потока:

$$q = K (t_{ж_1} - t_{ж_2}) \quad [\text{Вт/м}^2]$$

5. Вычисляем температуры на поверхностях стенки:

$$t_{c_1} = t_{ж_1} - q \frac{1}{\alpha_1} \quad [^{\circ}\text{C}]$$

$$t_{c_2} = t_{ж_2} - q \frac{1}{\alpha_2} \quad [^{\circ}\text{C}]$$

6. Определяем среднюю температуру стенки

$$t'_c = 0,5 (t_{c_1} + t_{c_2})$$

7. Уточняем значение коэффициента теплопроводности:

$$\lambda_{cp} = 0,84(1 + 0,695 \cdot 10^{-3} \cdot t'_c)$$

Далее расчет повторяется по той же схеме до тех пор, пока полученное среднее значение коэффициента теплопроводности практически совпадает с принятым ранее значением.

Задача 3. Определить необходимую площадь поверхности нагревателя, высоту труб в одном ходе и количество труб, расположенных вдоль и поперек потока воздуха в трубчатом воздухоподогревателе парового котла.

Алгоритм решения:

1. Определяем среднеарифметическую температуру воздуха в воздухоподогревателе

$$t_{ж_2} = 0,5 (t'_{ж_2} + t''_{ж_2}) \quad [^{\circ}\text{C}]$$

2. Вычисляем количество передаваемой теплоты:

$$Q = G_2 C_{pж_2} (t''_{ж_2} + t'_{ж_2}) \quad [\text{Вт}]$$

3. Определяем температуру газов на выходе из воздухоподогревателя:

$$t''_{ж_2} = t'_{ж_2} - \frac{Q}{GC_{pж_2}} \quad [^{\circ}\text{C}]$$

4. Находим среднеарифметическую температуру воздуха в воздухоподогревателе:

$$t_{ж_1} = 0,5 (t'_{ж_1} + t''_{ж_1}) \quad [^{\circ}\text{C}]$$

5. Число Рейнольдса для потока газов:

$$\text{Re}_{ж_1} = \frac{\omega_1 d_1}{\nu_{ж_1}}$$

6. Критерий Нуссельта:

$$\text{Nu}_{ж_1} = 0,021 \text{Re}_{ж_1}^{0,8} \text{Pr}_{ж_1}^{0,43}$$

7. Коэффициент теплоотдачи от газов к стенкам труб

$$\alpha_1 = \text{Nu}_{ж_1} \frac{\lambda_{ж_1}}{d_1} \quad [\text{Вт}/\text{м}^2 \text{ } ^{\circ}\text{C}]$$

8. Число Рейнольдса для потока воздуха

$$\text{Re}_{ж_2} = \frac{\omega_2 d_2}{\nu_{ж_2}}$$

9. Критерий Нуссельта:

$$\text{Nu}_{ж_2} = 0,41 \text{Re}_{ж_2}^{0,6} \text{Pr}_{ж_2}^{0,33} \varepsilon_s$$

10. Вычисляем коэффициент теплоотдачи от стенок труб к воздуху при поперечном потоке:

$$\alpha_2 = \text{Nu}_{ж_2} \frac{\lambda_{ж_2}}{d_2} \quad [\text{Вт/м}^2 \text{ } ^\circ\text{C}]$$

11. Определяем коэффициент теплопередачи:

$$K = \frac{1}{\frac{1}{\alpha_1} + \frac{\delta_c}{\lambda_c} + \frac{1}{\alpha_2}} \quad [\text{Вт/м}^2 \text{ } ^\circ\text{C}]$$

12. Определяем температурный напор:

$$\Delta t_{\text{прот}} = t_{ж_1} + t_{ж_2} \quad [^\circ\text{C}]$$

13. Площадь поверхности нагрева воздухоподогревателя:

$$F = \frac{Q}{k \Delta t} \quad [\text{м}^2]$$

14. Находим общее число труб:

$$n = \frac{4G_1}{\rho_{ж_1} \pi d_1^2 \omega_1} \quad [\text{шт}]$$

15. Вычисляем высоту труб в одном ходе:

$$\ell_1 = \frac{F}{2\pi d_1 n} \quad [\text{м}]$$

16. Определяем площадь живого сечения для прохода воздуха:

$$f = \frac{G_2}{\rho_{ж_2} \omega_2} \quad [\text{м}^2]$$

17. Вычисляем число труб, расположенных поперек потока

$$n = \frac{f}{\ell_1 (S_1 - d_2)} \quad [\text{шт}]$$

18. Вычисляем число труб, расположенных вдоль потока

$$n_2 = \frac{n}{n_1} \quad [\text{шт}]$$

Задача 4. На паропроводе перегретого пара установлена измерительная диафрагма, которая должна быть протарирована, т.е. найдена зависимость $\Delta P = f(G)$ перепада давлений от расхода пара. Тарировка выполнена на образце (модели) в 1/5 натуральной величины. Найти зависимость $\Delta P = f(G)$ для образца и указать границы ее применения при известных давлении и температуре пара.

Алгоритм решения:

Обработку опытных данных производим в критериях подобия и строим зависимость $Eu = f(Re)$ (Eu – критерий Эйлера, Re – число Рейнольдса).

Для определения зависимости $Eu = f(Re)$ вычисления значений критериев производятся в следующей последовательности:

1. Критерий Эйлера:

$$Eu = \frac{\Delta P_M}{\rho_M \omega_M^2},$$

учитывая, что скорость

$$\omega_M = \frac{4G_M}{\rho_M \pi d_M^2},$$

$$Eu = \frac{\Delta P_M}{\rho_M} \left(\frac{\rho_M \pi d_M^2}{4G_M} \right)^2.$$

2. Критерий Рейнольдса:

$$Re = \frac{\omega_M d_M}{\nu_M} = \frac{4G_M}{\rho_M \nu_M \pi d_M}$$

3. По данным Eu и Re строим график зависимости $Eu = f(Re)$ и определяем автомоделную область.
4. Определяем искомую зависимость

$$\Delta P = Eu \rho \omega^2,$$

причем скорость выражаем через расход:

$$\omega = \frac{G_v}{0,785d^2}$$

Задача 5. В закрытом сосуде объемом $V = 300$ л находится воздух при давлении $p_1 = 0,8$ МПа и температуре $t_1 = 20$ °С.

Какое количество теплоты необходимо подвести для того, чтобы температура воздуха поднялась до $t_2 = 120$ °С? Задачу решить, принимая теплоемкость воздуха постоянной, а также учитывая зависимость теплоемкости от температуры. Определить относительную ошибку, получаемую в первом случае.

Алгоритм решения:

1. Пользуясь уравнением состояния, определяем массу воздуха, находящегося в сосуде:

$$M = \frac{V p}{RT} \quad [\text{кг}].$$

2. Для двухатомных газов, считая теплоемкость величиной постоянной, имея $\mu c_v = 20,93$ кДж/(кмоль·К); вычисляем теплоемкость воздуха

$$c_v = \frac{\mu c_v}{\mu} \quad [\text{кДж}/(\text{кг} \cdot \text{К})].$$

3. Количество подведенной теплоты

$$Q = M c_v (t_2 - t_1) \quad [\text{кДж}].$$

4. Теплоемкость воздуха с учетом ее зависимости от температуры из табл. XII [3]. При необходимости пользуемся интерполяцией, находим c_v [кДж/(кг·К)].

5. Вычисляем относительную ошибку

$$\frac{c_v - c'_v}{c_v} \cdot 100 \quad .$$

Незначительная величина ошибки объясняется малым интервалом температур. При большой разности температур относительная ошибка может достигнуть весьма большой величины.

Задача 6. Какое количество теплоты необходимо затратить, чтобы нагреть 2 м³ воздуха при постоянном избыточном давлении $p=0,2$ МПа от $t_1 = 100$ °С до $t_2 = 500$ °С? Какую работу при этом совершит воздух? Давление атмосферы принять равным 101325 Па.

Алгоритм решения:

1. Определяем количество теплоты:

$$q_p = c_{pm2} t_2 - c_{pm1} t_1 \quad [\text{кДж/кг}]$$

Из табл. XII [3] находим

$$c_{pm1} = 1,0061 \text{ кДж/(кг·К)}; c_{pm2} = 1,0387 \text{ кДж/(кг·К)}.$$

2. Определяем массу воздуха из характеристического уравнения

$$M = \frac{pV}{RT} \quad . \quad [\text{кг}].$$

3. Определяем количество теплоты по объему.

$$q_p = c'_{pm2} t_2 - c'_{pm1} t_1 \quad . \quad [\text{кДж/кг}].$$

Из табл. XII [3] находим

$$c'_{pm1} = (c'_{pm})_0^{100} = 1,3004 \text{ кДж/(м}^3 \cdot \text{К)};$$

$$c'_{pm2} = (c'_{pm})_0^{500} = 1,3427 \text{ кДж/(м}^3 \cdot \text{К)},$$

4. Приводим объем воздуха к нормальным условиям.

$$V_H = \frac{pVT_H}{Tp_H}, \quad \text{м}^3, \quad ,$$

$$Q_p = q_p V_H \quad . \quad [\text{кДж}].$$

5. Вычисляем работу газа

$$L = MR (t_2 - t_1) \quad [\text{кДж}].$$

Задача 7. 1 кг воздуха при температуре $t_1 = 30^\circ\text{C}$, начальном давлении $p_1 = 0,1$ МПа сжимается изотермически до конечного давления $p_2 = 1$ МПа.

Определить конечный объем затрачиваемой работы и количество теплоты, отводимой от газа.

Алгоритм решения:

1. Определяем начальный объем воздуха из уравнения состояния:

$$v_1 = \frac{RT_1}{p_1}, \quad [\text{м}^3/\text{кг}]$$

Так как в изотермическом процессе

$$p_1 v_1 = p_2 v_2,$$

то конечный объем

$$v_2 = v_1 \frac{p_1}{p_2}. \quad [\text{м}^3/\text{кг}]$$

2. Вычисляем работу, затрачиваемую на сжатие 1 кг воздуха:

$$L = RT \ln \frac{p_1}{p_2} = 2,303RT \lg \frac{p_1}{p_2}. \quad [\text{кДж/кг}]$$

3. Количество теплоты, отводимой от газа, равно работе, затраченной на сжатие.

Задача 8. Сосуд емкостью 90 л содержит воздух при давлении 0,8 МПа и температуре 30°C . Определить количество теплоты, которое необходимо сообщить воздуху, чтобы повысить его давление при $V = \text{const}$ до 1,6 МПа. Принять зависимость $c = f(t)$ нелинейной.

Алгоритм решения:

1. Из соотношения параметров изохорного процесса получим

$$T_2 = T_1 \frac{p_2}{p_1}, \quad [^\circ\text{K}];$$

2. По уравнению $q_v = c_{vm2} t_2 - c_{vm1} t_1$

Из табл. XII [3] находим

$$c_{vm1} = 0,7173 \text{ кДж}/(\text{кг}\cdot\text{K}); \quad c_{vm2} = 0,7351 \text{ кДж}/(\text{кг}\cdot\text{K}).$$

3. Определяем массу воздуха, находящегося в резервуаре:

$$M = \frac{p_1 V_1}{RT_1}, \quad [\text{кг}].$$

4. Определяем сообщенное воздуху количество теплоты

$$Q_v = M \cdot q_v \quad [\text{кДж}].$$

Задача 9. 1 кг воздуха совершает цикл Карно в пределах температур $t_1 = 627^\circ\text{C}$ и $t_1 = 27^\circ\text{C}$, причем наивысшее давление составляет 6 МПа, а наинизшее – 0,1 МПа. Определить параметры состояния воздуха в характерных точках цикла, работу, термический к.п.д. цикла и количество подведенной и отведенной теплоты.

Алгоритм решения:

Точка 1.

$$p_1 = 6 \text{ МПа}; T_1 = 900 \text{ К}.$$

Удельный объем газа находим из характеристического уравнения

$$v_1 = \frac{RT_1}{p_1}, \quad [\text{м}^3/\text{кг}].$$

Точка 2.

$$T_2 = 900 \text{ К}.$$

Из уравнения адиабаты (линия 2-3) определяем p_2

$$\frac{p_2}{p_3} = \left(\frac{T_2}{T_3}\right)^{\frac{\kappa}{\kappa-1}};$$

Из уравнения изотермы (линия 1-2)

$$p_1 v_1 = p_2 v_2$$

получаем
$$v_2 = \frac{p_1 v_1}{p_2}, \quad [\text{м}^3/\text{кг}].$$

Точка 3.

$$p_3 = 0,1 \text{ МПа}; T_3 = 300 \text{ К};$$

$$v_3 = \frac{RT_3}{p_3}, \quad [\text{м}^3/\text{кг}].$$

Точка 4.

$$T_4 = 300 \text{ К}.$$

Из уравнения адиабаты (линия 4-1) имеем

$$\frac{p_1}{p_4} = \left(\frac{T_1}{T_4}\right)^{\frac{\kappa}{\kappa-1}} \text{ и находим } p_4$$

Из уравнения изотермы (линия 3-4) получаем

$$p_3 v_3 = p_4 v_4$$

$$v_4 = \frac{p_3 v_3}{p_4}, \quad [\text{м}^3/\text{кг}].$$

Термический к.п.д. цикла

$$\eta_t = \frac{T_1 - T_2}{T_1}.$$

Подведенное количество теплоты

$$q_1 = RT_1 \ln \frac{v_2}{v_1}, \quad [\text{кДж/кг}].$$

Отведенное количество теплоты

$$q_1 = RT_3 \ln \frac{v_3}{v_4}, \quad [\text{кДж/кг}].$$

Работа цикла

$$l_0 = q_1 - q_2, \quad [\text{кДж/кг}].$$

Для проверки можно воспользоваться формулой

$$\eta_t = \frac{q_1 - q_2}{q_1} = \frac{l_0}{q_1}.$$

Варианты задач для выполнения контрольной работы

№ варианта	Задача 1	Задача 2	Задача 3	Задача 4	Задача 5	Задача 6	Задача 7	Задача 8	Задача 9
1	Для всей группы	$\bar{t}_c = 640^\circ\text{C}$	Для всей группы	Для всей группы	V=300л P ₁ =0,7МПа t ₁ =21 ⁰ С	t ₁ =110 ⁰ С t ₂ =500 ⁰ С P=0,3МПа	t ₁ =28 ⁰ С P ₁ =0,3МПа P ₂ =1,2МПа	V=80л P=0,9МПа T=28 ⁰ С	T ₁ =606 ⁰ С T ₂ =25 ⁰ С P ₁ =5,3МПа P ₂ =0,1МПа
2	Для всей группы	$\bar{t}_c = 655^\circ\text{C}$	Для всей группы	Для всей группы	V=305л P ₁ =0,7МПа t ₁ =21 ⁰ С	t ₁ =110 ⁰ С t ₂ =500 ⁰ С P=0,3МПа	t ₁ =29 ⁰ С P ₁ =0,2МПа P ₂ =1,2МПа	V=82л P=0,8МПа T=25 ⁰ С	T ₁ =620 ⁰ С T ₂ =20 ⁰ С P ₁ =5,0МПа P ₂ =0,4МПа
3	Для всей группы	$\bar{t}_c = 653^\circ\text{C}$	Для всей группы	Для всей группы	V=306л P ₁ =0,6МПа t ₁ =21 ⁰ С	t ₁ =120 ⁰ С t ₂ =510 ⁰ С P=0,2МПа	t ₁ =26 ⁰ С P ₁ =0,2МПа P ₂ =1,1МПа	V=81л P=0,7МПа T=23 ⁰ С	T ₁ =612 ⁰ С T ₂ =20 ⁰ С P ₁ =5,9МПа P ₂ =0,4МПа
4	Для всей группы	$\bar{t}_c = 660^\circ\text{C}$	Для всей группы	Для всей группы	V=301л P ₁ =0,5МПа t ₁ =20 ⁰ С	t ₁ =115 ⁰ С t ₂ =505 ⁰ С P=0,4МПа	t ₁ =29 ⁰ С P ₁ =0,4МПа P ₂ =1,3МПа	V=85л P=0,6МПа T=25 ⁰ С	T ₁ =590 ⁰ С T ₂ =19 ⁰ С P ₁ =5,1МПа P ₂ =0,3МПа
5	Для всей группы	$\bar{t}_c = 662^\circ\text{C}$	Для всей группы	Для всей группы	V=310л P ₁ =0,5МПа t ₁ =19 ⁰ С	t ₁ =113 ⁰ С t ₂ =505 ⁰ С P=0,2МПа	t ₁ =30 ⁰ С P ₁ =0,2МПа P ₂ =1,1МПа	V=86л P=0,8МПа T=25 ⁰ С	T ₁ =600 ⁰ С T ₂ =20 ⁰ С P ₁ =6,3МПа P ₂ =1,1МПа
6	Для всей группы	$\bar{t}_c = 656^\circ\text{C}$	Для всей группы	Для всей группы	V=295л P ₁ =0,6МПа t ₁ =22 ⁰ С	t ₁ =100 ⁰ С t ₂ =501 ⁰ С P=0,2МПа	t ₁ =29 ⁰ С P ₁ =0,2МПа P ₂ =1,0МПа	V=89л P=0,9МПа T=26 ⁰ С	T ₁ =610 ⁰ С T ₂ =20 ⁰ С P ₁ =5,1МПа P ₂ =0,3МПа
7	Для всей группы	$\bar{t}_c = 658^\circ\text{C}$	Для всей группы	Для всей группы	V=300л P ₁ =0,6МПа t ₁ =18 ⁰ С	t ₁ =111 ⁰ С t ₂ =504 ⁰ С P=0,3МПа	t ₁ =28 ⁰ С P ₁ =0,2МПа P ₂ =1,0МПа	V=80л P=0,8МПа T=26 ⁰ С	T ₁ =580 ⁰ С T ₂ =19 ⁰ С P ₁ =5,1МПа P ₂ =0,2МПа
8	Для всей группы	$\bar{t}_c = 659^\circ\text{C}$	Для всей группы	Для всей группы	V=307л P ₁ =0,5МПа t ₁ =21 ⁰ С	t ₁ =114 ⁰ С t ₂ =503 ⁰ С P=0,1МПа	t ₁ =30 ⁰ С P ₁ =0,4МПа P ₂ =1,3МПа	V=82л P=0,7МПа T=25 ⁰ С	T ₁ =601 ⁰ С T ₂ =23 ⁰ С P ₁ =5,0МПа P ₂ =0,3МПа
9	Для всей группы	$\bar{t}_c = 656^\circ\text{C}$	Для всей группы	Для всей группы	V=300л P ₁ =0,7МПа t ₁ =23 ⁰ С	t ₁ =110 ⁰ С t ₂ =505 ⁰ С P=0,4МПа	t ₁ =28 ⁰ С P ₁ =0,2МПа P ₂ =1,0МПа	V=82л P=0,6МПа T=24 ⁰ С	T ₁ =603 ⁰ С T ₂ =21 ⁰ С P ₁ =5,0МПа P ₂ =0,2МПа
10	Для всей группы	$\bar{t}_c = 655^\circ\text{C}$	Для всей группы	Для всей группы	V=305л P ₁ =0,7МПа t ₁ =23 ⁰ С	t ₁ =112 ⁰ С t ₂ =506 ⁰ С P=0,3МПа	t ₁ =25 ⁰ С P ₁ =0,1МПа P ₂ =1,1МПа	V=81л P=0,8МПа T=27 ⁰ С	T ₁ =600 ⁰ С T ₂ =23 ⁰ С P ₁ =5,1МПа P ₂ =0,2МПа
11	Для всей группы	$\bar{t}_c = 663^\circ\text{C}$	Для всей группы	Для всей группы	V=301л P ₁ =0,8МПа t ₁ =22 ⁰ С	t ₁ =110 ⁰ С t ₂ =504 ⁰ С P=0,3МПа	t ₁ =29 ⁰ С P ₁ =0,2МПа P ₂ =1,1МПа	V=84л P=0,6МПа T=21 ⁰ С	T ₁ =616 ⁰ С T ₂ =20 ⁰ С P ₁ =6,3МПа P ₂ =1,1МПа
12	Для всей группы	$\bar{t}_c = 665^\circ\text{C}$	Для всей группы	Для всей группы	V=300л P ₁ =0,4МПа t ₁ =23 ⁰ С	t ₁ =109 ⁰ С t ₂ =500 ⁰ С P=0,4МПа	t ₁ =29 ⁰ С P ₁ =0,2МПа P ₂ =1,0МПа	V=80л P=0,6МПа T=28 ⁰ С	T ₁ =600 ⁰ С T ₂ =22 ⁰ С P ₁ =5,1МПа P ₂ =0,1МПа
13	Для всей группы	$\bar{t}_c = 667^\circ\text{C}$	Для всей группы	Для всей группы	V=303л P ₁ =0,7МПа t ₁ =24 ⁰ С	t ₁ =115 ⁰ С t ₂ =501 ⁰ С P=0,2МПа	t ₁ =27 ⁰ С P ₁ =0,2МПа P ₂ =1,0МПа	V=80л P=0,9МПа T=24 ⁰ С	T ₁ =620 ⁰ С T ₂ =26 ⁰ С P ₁ =5,0МПа P ₂ =0,3МПа

14	Для всей группы	$\bar{t}_c = 661^{\circ}\text{C}$	Для всей группы	Для всей группы	V=303л P ₁ =0,6МПа t ₁ =23 ⁰ С	t ₁ =103 ⁰ С t ₂ =509 ⁰ С P=0,1МПа	t ₁ =30 ⁰ С P ₁ =0,1МПа P ₂ =1,2МПа	V=82л P=0,8МПа T=26 ⁰ С	T ₁ =626 ⁰ С T ₂ =28 ⁰ С P ₁ =5,0МПа P ₂ =0,2МПа
15	Для всей группы	$\bar{t}_c = 663^{\circ}\text{C}$	Для всей группы	Для всей группы	V=311л P ₁ =0,7МПа t ₁ =23 ⁰ С	t ₁ =116 ⁰ С t ₂ =507 ⁰ С P=0,1МПа	t ₁ =31 ⁰ С P ₁ =0,4МПа P ₂ =1,3МПа	V=85л P=0,7МПа T=24 ⁰ С	T ₁ =612 ⁰ С T ₂ =22 ⁰ С P ₁ =5,5МПа P ₂ =0,3МПа
16	Для всей группы	$\bar{t}_c = 662^{\circ}\text{C}$	Для всей группы	Для всей группы	V=293л P ₁ =0,6МПа t ₁ =21 ⁰ С	t ₁ =112 ⁰ С t ₂ =506 ⁰ С P=0,2МПа	t ₁ =30 ⁰ С P ₁ =0,2МПа P ₂ =1,3МПа	V=82л P=0,5МПа T=23 ⁰ С	T ₁ =600 ⁰ С T ₂ =20 ⁰ С P ₁ =5,0МПа P ₂ =0,1МПа
17	Для всей группы	$\bar{t}_c = 659^{\circ}\text{C}$	Для всей группы	Для всей группы	V=297л P ₁ =0,6МПа t ₁ =23 ⁰ С	t ₁ =110 ⁰ С t ₂ =505 ⁰ С P=0,4МПа	t ₁ =29 ⁰ С P ₁ =0,1МПа P ₂ =1,3МПа	V=83л P=0,6МПа T=22 ⁰ С	T ₁ =603 ⁰ С T ₂ =22 ⁰ С P ₁ =5,3МПа P ₂ =0,1МПа
18	Для всей группы	$\bar{t}_c = 658^{\circ}\text{C}$	Для всей группы	Для всей группы	V=299л P ₁ =0,5МПа t ₁ =21 ⁰ С	t ₁ =117 ⁰ С t ₂ =503 ⁰ С P=0,3МПа	t ₁ =28 ⁰ С P ₁ =0,3МПа P ₂ =1,2МПа	V=89л P=0,7МПа T=25 ⁰ С	T ₁ =610 ⁰ С T ₂ =21 ⁰ С P ₁ =6,3МПа P ₂ =0,5МПа
19	Для всей группы	$\bar{t}_c = 653^{\circ}\text{C}$	Для всей группы	Для всей группы	V=301л P ₁ =0,7МПа t ₁ =24 ⁰ С	t ₁ =116 ⁰ С t ₂ =505 ⁰ С P=0,4МПа	t ₁ =31 ⁰ С P ₁ =0,4МПа P ₂ =1,3МПа	V=86л P=0,8МПа T=23 ⁰ С	T ₁ =608 ⁰ С T ₂ =24 ⁰ С P ₁ =5,2МПа P ₂ =0,2МПа
20	Для всей группы	$\bar{t}_c = 656^{\circ}\text{C}$	Для всей группы	Для всей группы	V=298л P ₁ =0,8МПа t ₁ =23 ⁰ С	t ₁ =113 ⁰ С t ₂ =509 ⁰ С P=0,5МПа	t ₁ =31 ⁰ С P ₁ =0,2МПа P ₂ =1,1МПа	V=87л P=0,8МПа T=26 ⁰ С	T ₁ =603 ⁰ С T ₂ =20 ⁰ С P ₁ =5,0МПа P ₂ =0,1МПа

ЛИТЕРАТУРА

1. Бадалов Ф.Б. и др. Моделирование и управление технологическим процессом. Сборник научных трудов Математические модели и методы решения инженерных задач ЭВМ - Т.: Фан, 1991.
2. Балашевич В.А. Основы математического моделирования. Уч. пособие – Минск: Высшая школа, 1985.
3. Рабинович О.М. Сборник задач по технической термодинамике - М.: Машиностроение, 1973.
4. Краснощеков В.А., Сукомел А.С. Задачник по теплопередаче- М.: Энергия, 1980.
5. http://dhes.ime.mrsu.ru/studies/tot/tot_lit.htm
6. http://rbip.bookchamber.ru/description.aspx?product_no=854
7. Аверченков В.И., Федоров В.П., Хейфец М.Л. Основы математического моделирования технических систем. Учебное пособие. - М.: Изд-во «ФЛИНТА», 2011.-271с. Режим доступа: <http://biblioclub.ru/index.php?page=book&id=93344>
8. Попырин Л.С. Математическое моделирование и оптимизация теплоэнергетических установок. – М. Энергия, 1978-416с., ил.
9. Мухин, О. И. Моделирование систем: электронный учебник. — Пермь: Изд-во ПГТУ, 2001. URL: stratum.pstu.ac.ru.

Составители

Кудряшев Геннадий Сергеевич
Третьяков Александр Николаевич

**Математическое моделирование и алгоритмизация задач
теплоэнергетики**

Учебное пособие по дисциплине «Математическое моделирование и алгоритмизация задач теплоэнергетики»

для студентов очной и заочной форм обучения направлений подготовки

13.04.01 Теплоэнергетика и теплотехника (уровень магистратура),

13.04.02 Электроэнергетика и электротехника (уровень магистратуры)
(методические указания и задания для выполнения контрольных работ)

Лицензия на издательскую деятельность
ЛР №070444 от 11.03.1998 г.

Подписано в печать 19.01.2021 г.
Формат 60×86/16. Печ. л. 11,55
Тираж 50 экз.

Издательство Иркутского государственного
аграрного университета им. А.А. Ежевского
664038, Иркутская обл., Иркутский р-н,
пос. Молодежный