

Учебно- исследовательская работа студента

Кафедра «Радиоэлектроника»

Лекционный курс

Авторы

Звездина М.Ю., Шокова Ю.А.

Аннотация

Лекционный курс предназначен для обучения студентов направлений 210700 и 210400. Раскрывает основные моменты обработки экспериментальных данных.

Авторы



**Звездина Марина Юрьевна – зав.
кафедрой «Радиоэлектроника»**

**ДОКТОР ФИЗИКО-МАТЕМАТИЧЕСКИХ НАУК,
ДОЦЕНТ**

Сфера научных интересов – устройства СВЧ и антенны



**Шокова Юлия Александровна –
старший преподаватель,**

**КАНДИДАТ ФИЗИКО-МАТЕМАТИЧЕСКИХ
НАУК**

Сфера научных интересов – антенны, цифровая обработка сигналов

ОГЛАВЛЕНИЕ

Лекция 1. Введение. Понятие моделирования. Способы представления моделей.....	5
Введение. Цель и место дисциплины.....	5
Понятие моделирования	5
Способы представления моделей	10
Лекция 2. Основы планирования эксперимента.....	12
Классификация видов экспериментальных исследований	12
Планирование эксперимента	14
Типы измеряемых физических величин	15
Типы погрешностей измерений.....	17
Лекция 3. Фиксация и обработка статистических данных	18
Принципы подбора выборки.....	18
Определение параметров экспериментального распределения	20
Оценка размера выборки.....	24
Оценки погрешности измерений	25
Лекция 4. Основные законы распределения, применяемые при обработке данных научного эксперимента.....	27
Нормальный закон распределения и его основные свойства	27
Логнормальный закон распределения и распределение Вейбулла	31
Оценка согласия эмпирических и теоретических распределений.....	33
Лекция 5. Регрессионный, корреляционный и дисперсионный виды анализа	36
Линейные регрессионные модели	36
Корреляционный анализ.....	41
Дисперсионный анализ.....	43
Лекция 6. Модель динамической системы в виде Фурье представления	45
Общие сведения о динамических системах.....	45
Модель объекта.....	48
Модель сигнала.....	51
Лекция 7. Статистическое моделирование	54

Учебно-исследовательская работа студента

Общие сведения о статистическом моделировании	54
Методы генерирования случайной величины	56
Марковские цепи	57
Лекция 8. Процесс экспертизы.....	63
Методы получения оценок экспертизы. Голосование	63
Методы ранжирования.....	64
Лекция 8. Построение отчета о выполненной учебно-исследовательской работе	70
План отчета об учебно-исследовательской работе	70
Содержание работы.....	71
Оформление графической части отчета	73
Структура выступления на защите научной работы.....	74

Лекция 1. Введение. Понятие моделирования. Способы представления моделей

Введение. Цель и место дисциплины

Основная цель дисциплины «Учебно-исследовательская работа студента» - развитие практических навыков у студентов в обработке экспериментальных данных, проведении анализа полученных результатов и формулировке рекомендаций по их использованию, развитию умения излагать полученные результаты на семинарах, конференциях или других формах представления результатов работы.

Учебный план по данной дисциплине предполагает 18 часов лекций и 18 часов лабораторных работ. Завершается изучение дисциплины сдачей зачета в зимний семестр. Зачет по дисциплине выставляется за успешно выполненные и защищенные лабораторные работы (5 работ).

Лабораторные работы, выполняемые на компьютере, включают в себя следующие темы:

1. Статистическая обработка результатов исследований в Excel (4 час).
2. Статистическая обработка результатов исследований в языковой среде MathCad (4 час).
3. Статистическая обработка результатов исследований в языковой среде MatLab (4 час).
4. Построение схемы установки в пакете Visio (4 час).
5. Построение текста доклада по результатам исследований (2 час).

Понятие моделирования

Модель - способ замещения реального объекта, используемый для его изучения.

Модель используется в случаях, когда эксперимент опасен, дорог, происходит в неудобном масштабе пространства и времени (долговременен, слишком кратковременен, протяжен...), невозможен, неповторим, ненагляден и т. д. Например,

- «эксперимент опасен». При деятельности в агрессивной среде вместо человека лучше использовать его макет (луноход);

- «дорог». Прежде чем использовать идею в реальной экономике страны, лучше опробовать её на математической или имитационной модели экономики и получив представление о возможных последствиях;

- «долговременен». Изучить старение электрокабелей выгоднее и быстрее на модели;

- «кратковременен». (Например, взрыв);

- «протяжен в пространстве» (для изучения дальнего космоса);

- «микроскопичен» (изучения взаимодействия атомов);

- «невозможен» (объект 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.

Учебно-исследовательская работа студента

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» (эксперимент повторить нельзя. Пример — исторические процессы);

- «ненагляден» (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 (рисунок 1.1). Если требуется уточнение, эти этапы повторяются по спирали вновь и вновь: формализация (переход от реального объекта к модели), моделирование (исследование и преобразование модели), интерпретация (перевод результатов моделирования в область реальности).

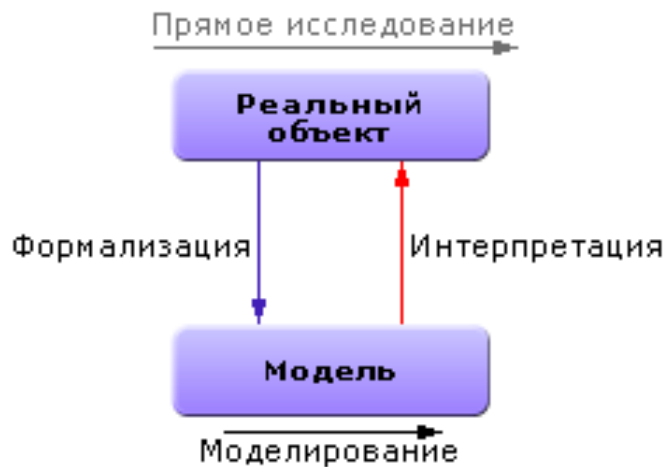


Рисунок 1.1 – Процесс моделирования (базовый вариант)

Поскольку моделирование — способ замещения реального объекта его аналогом, то возникает вопрос: насколько аналог должен соответствовать исходному объекту? Двумя крайними вариантами являются следующие.

Вариант 1: соответствие — 100%. Очевидно, что точность решения в этом случае максимальна, а ущерб от применения модели минимален. Однако затраты на построение такой модели бесконечно велики, поскольку объект повторяется во всех своих деталях. Фактически создаётся точно такой же объект путём копирования его до атомов, что само по себе не имеет смысла.

Учебно-исследовательская работа студента

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.

Обычно 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 (см. рисунок 1.2).

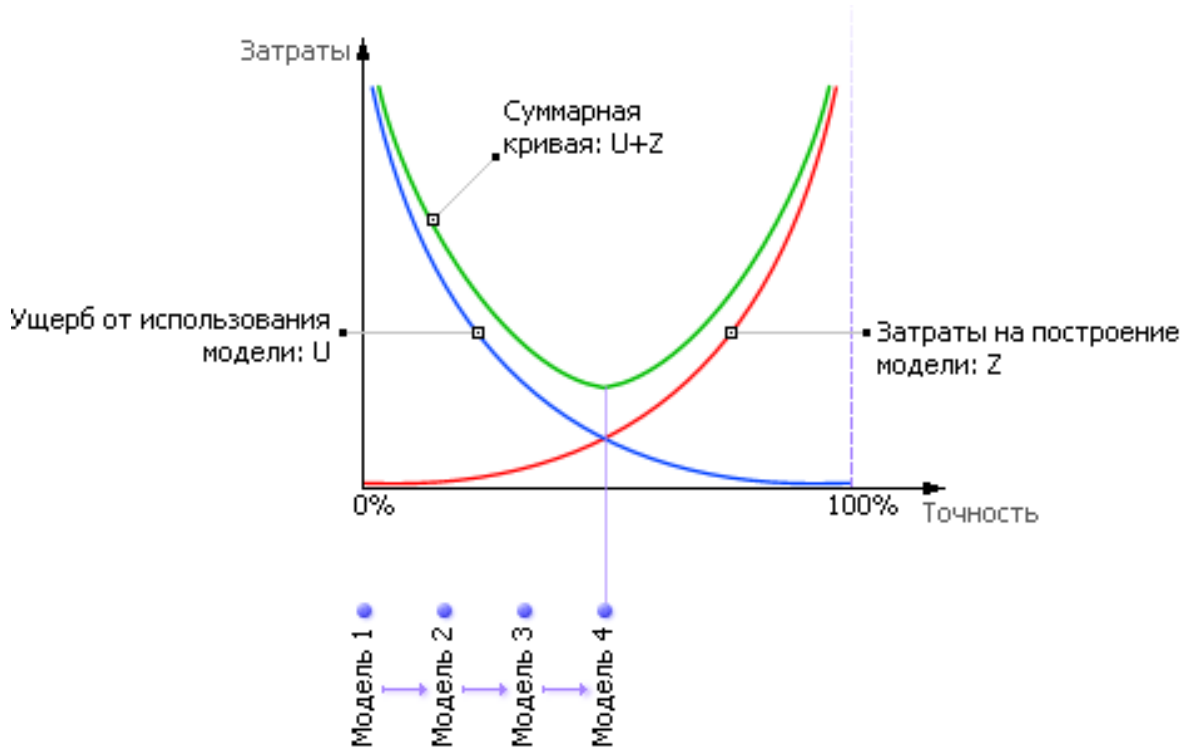


Рисунок 1.2 – Соотношение суммарных затрат и точности для различных вариантов детализации прикладной модели

Просуммировав две кривые затрат, можно получить кривую общих затрат $U+Z$. Оптимальное значение расположено на суммарной кривой между этими крайними вариантами. Таким образом, можно сделать вывод, что как неточные модели не нужны, так и абсолютная точность тоже не нужна, да и невозможна.

Что следует отбросить, что оставить в модели? Обычно в модель включаются только существенные аспекты, представляющие объект, и отбрасываются *все остальные* (бесконечное большинство). Существенный или несущественный аспект описания определяют согласно цели исследования. Таким образом, каждая модель составляется с какой-то целью. Начиная моделирование, исследователь должен определить цель, отделив её от всех возможных других целей, число которых, по-видимому, бесконечно.

К сожалению, указанная на рисунке 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»...) к все более сложным

Учебно-исследовательская работа студента

(«Модель 3», «Модель 4»...). Процесс моделирования при этом имеет циклический спиралевидный характер: если построенная модель не удовлетворяет требованиям точности, то её детализируют, дорабатывают на следующем цикле.

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.

Из всего сказанного следует, что моделей может быть несколько: приближенная, более точная, ещё точнее и так далее. Модели как бы образуют ряд. 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 (имиджи, стандартные библиотеки, мастера, линейки, резинки...).

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.

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

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. Однако при этом сохранена структура модели, связи, элементы, подсистемы. Всегда можно вернуться назад и что-то переделать. Модель, которая сдаётся заказчику, может быть оформлена в виде специализированного автоматизированного рабочего места (АРМа), написанного на языке программирования, в котором можно уделить внимание интерфейсу, скоростным параметрам и другим потребительским свойствам.

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

Учебно-исследовательская работа студента

Моделирование является инженерной наукой, технологией решения задач. Смежными моделированию предметами являются: программирование, математика, исследование операций.

Программирование есть способ изложения алгоритма в языковой форме. **Алгоритм** - один из способов представления (отражения) мысли, процесса, явления в искусственной вычислительной среде, которой является компьютер. **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.

Основные подсистемы при проектировании комплексных моделей представлены на рисунке 1.3. Входящие в ее состав компьютерная графика организует интерфейс пользователя с моделью, система искусственного интеллекта применяется для построения высших моделей, например, адаптивных, которые умеют самонастраиваться.

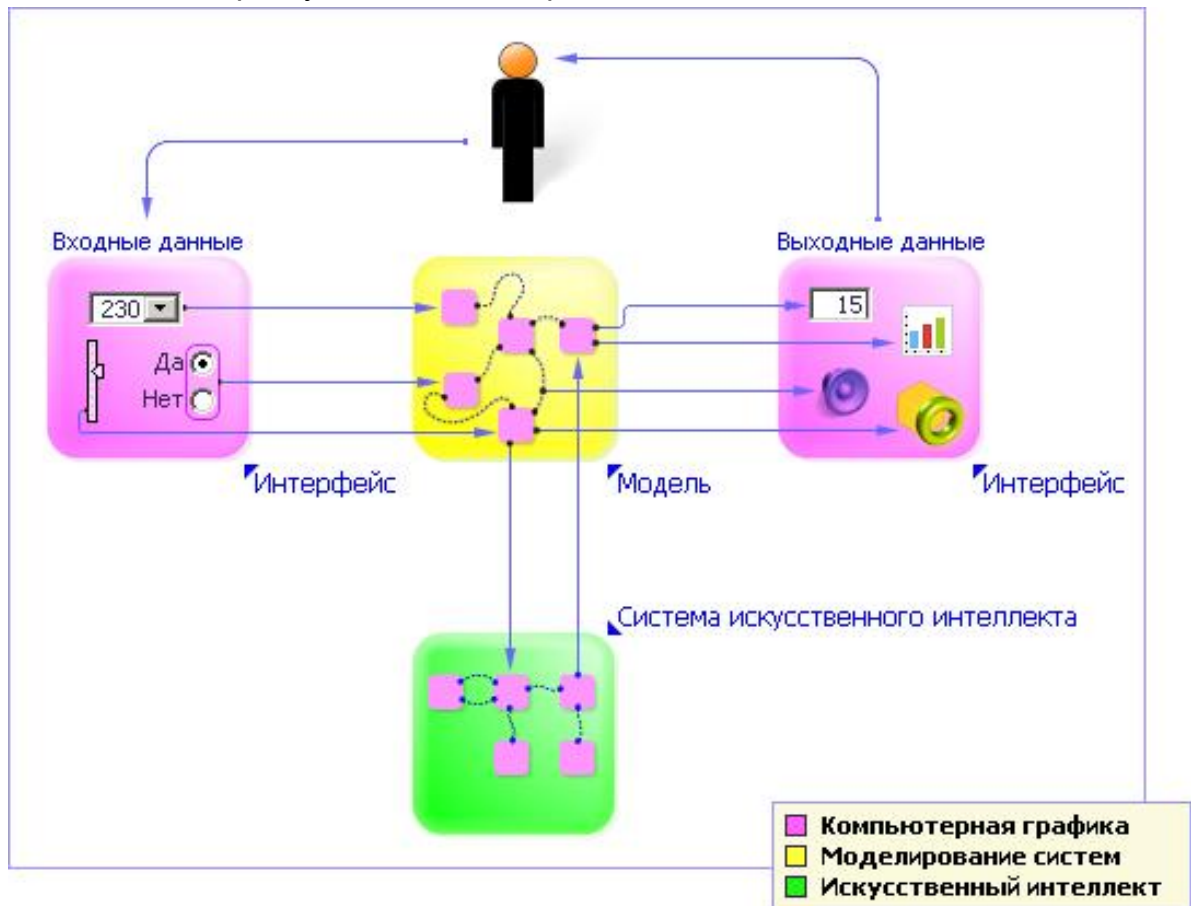


Рисунок 1.3 – Основные подсистемы при проектировании комплексных моделей

Способы представления моделей

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. Данные 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.)

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 его затылок. Приближение в виде трёхмерной модели позволяет решить данную задачу, однако с её помощью нельзя определить, когда, например, у виртуального человека вырастут волосы длиной 50 см? Имитационная модель ещё более информативна. Однако наибольшей ценностью обладают модели, обладающие прогностическими свойствами, умеющие отвечать на вопросы.

Следует различать два понятия - «модель» и «задача». **Модель** связывает переменные между собой законами. Данные 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. Человек, как существо субъективное, имеющее собственные цели, часто меняющиеся желания, ставит, в зависимости от своих потребностей, каждый раз

Учебно-исследовательская работа студента

новые задачи, требует решить возникающие у него проблемы. **Задача** - это совокупность вопроса и модели. Модели можно задавать все новые и новые вопросы и при этом не менять модель, но менять задачу.

Таким образом, **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 (алгебра есть правила преобразования). Процедура, которая помогает применить такие правила к модели, называется **методом**.

Способность модели преобразовываться с помощью алгебры даёт возможность в дальнейшем использовать её многократно для решения различных задач, делать на ней прогнозы. В связи с этим, создавая модель, следует либо создавать параллельно алгебру преобразования, либо использовать уже готовую алгебру и не отходить при построении модели от её правил.

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. Однако 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**.

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

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

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.

Использование модели. Построив модель, исследователь может:

Учебно-исследовательская работа студента

- 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).

Лекция 2. Основы планирования эксперимента

Классификация видов экспериментальных исследований

В предыдущей лекции было показано, что модель представляет собой фактически «черный» ящик, связывающий между собой входные и выходные параметры. Для установления функции преобразования входных данных в выходные применяются различные способы, зависящие от типа данных. Данные для установления зависимостей получают на основе наблюдений или эксперимента. *Наблюдение* – способ получения данных, при котором воздействие наблюдателя на объект сведено к минимуму. *Эксперимент* - наблюдение с воздействием на наблюдаемый объект.

Экспериментальные исследования заканчиваются представлением его результатов в виде сформулированных выводов и выдачи рекомендаций. Данная информация может быть выражена в виде графиков, чертежей, таблиц, статистических данных или словесных описаний.

Эксперимент предполагает проведение тех или иных опытов.

Опыт – это воспроизведение исследуемого явления в определенных условиях проведения эксперимента при возможности регистрации его результатов.

По цели проведения и форме представления полученных результатов эксперимент делят на качественный и количественный.

Качественный эксперимент устанавливает только сам факт существования какого-либо явления, но при этом не дает никаких количественных характеристик объекта исследований, т.е. качественный эксперимент предусматривает только словесное описание его результатов.

Для анализа свойств объекта в иных условиях, а также формулировке количественных рекомендаций необходимо использовать данные *количественного эксперимента*, который не только фиксирует факт существования того или иного явления, но и позволяет установить соотношение между количественными характеристиками явления и количественными характеристиками способов внешнего воздействия на объект исследований.

Учебно-исследовательская работа студента

Количественный эксперимент предполагает количественное определение всех тех способов внешнего воздействия на объект исследования, от которых зависит его поведение – количественное описание всех факторов.

Фактор – переменная величина, по предположению влияющая на результаты эксперимента.

Например, в качестве факторов эксперимента при исследовании удельного сопротивления проводника можно выбрать величину температуры окружающей среды, его качественный состав.

В отдельном конкретном опыте каждый фактор может принимать одно из возможных значений – *уровень фактора*, т.е. фиксированное значение фактора относительно начала отсчета. Фиксированный набор уровней всех факторов в каждом конкретном опыте как раз и определяет одно из возможных состояний объекта исследований.

При проведении опытов очень многое зависит от того, насколько активно экспериментатор может вмешиваться в исследуемое явление, имеет ли он или нет возможность устанавливать те уровня факторов, которые представляют для него интерес. С этой точки зрения все факторы можно разбить на три группы:

- *контролируемые и управляемые* – это факторы, для которых можно не только зарегистрировать их уровень, но еще и задать в каждом конкретном опыте любое его возможное значение;

- *контролируемые, но неуправляемые факторы* – это факторы, уровни которых можно только регистрировать, а задать в каждом опыте их определенное значение практически невозможно;

- *неконтролируемые* – это факторы, уровни которых не регистрируются экспериментатором и о существовании которых он даже может и не подозревать.

В количественном эксперименте необходимо не только регистрировать уровни всех контролируемых факторов, но и иметь возможность устанавливать количественное описание того свойства (отклика) исследуемого явления, которое изучает (наблюдает) экспериментатор. Причем поскольку на объект исследования в процессе эксперимента всегда влияет огромное количество неконтролируемых факторов, что вносит в получаемые результаты некоторый элемент неопределенности, значение отклика в каждом конкретном опыте невозможно предсказать заранее. В связи с этим воспроизведение исследуемого явления при одном и том же фиксированном наборе уровней всех контролируемых факторов будет приводить к различным значениям отклика, т.е. отклик – это всегда случайная величины.

Отклик – наблюдаемая случайная величина, по предположению зависящая от факторов.

В результате количественного эксперимента необходимо найти зависимость между откликом и факторами – *функцию отклика*. Данная функция также будет случайной величиной и ее можно задать одним из параметров своего распределения, например, математическим ожиданием.

С учетом приведенного выше деления факторов на три группы функцию отклика можно в самом общем случае записать в виде:

$$M_y = f(x_i, h_i) + \varepsilon_\delta, \quad (2.1)$$

Учебно-исследовательская работа студента

где M_y - математическое ожидание отклика; x_i - контролируемые и управляемые факторы; h_i - контролируемые, но неуправляемые факторы; ε_δ - ошибка эксперимента, учитывающая влияние неконтролируемых факторов.

В зависимости от того, какой группой факторов располагает исследователь количественный эксперимент можно разделить еще на два вида:

- *пассивный*, когда в распоряжении исследователя нет управляемых факторов, уровни факторов регистрируются, но не задаются;
- *активный*, когда экспериментатор имеет возможность не только контролировать факторы, но и управлять ими.

Планирование эксперимента

Поскольку в случае активного эксперимента исследователь имеет возможность вмешиваться в исследуемое явление, то предполагается план его проведения, а сам эксперимент планируется.

План эксперимента – совокупность данных, определяющих число, условия и порядок реализации опытов.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего поставленным требованиям.

К требованиям, предъявляемым при планировании активного эксперимента, можно отнести:

- степень точности и надежности результатов, полученных после проведения эксперимента;
- сроки и средства, имеющиеся в распоряжении исследователя и т.д.

Целью активного эксперимента может быть либо определение функции отклика в виде

$$M_y = f(x_i) + \varepsilon_\delta, \quad (2.2)$$

либо поиск такого сочетания уровней управляемых факторов x_i , при котором достигается оптимальное (экстремальное – минимальное или максимальное) значение функции отклика. В последнем случае эксперимент носит название *поискового (экстремального) эксперимента*.

По условиям проведения различают лабораторный и промышленный эксперименты.

Лабораторный эксперимент применяется для обеспечения большой «стерильности» условий проведения опытов с более тщательной подготовкой опытов.

Промышленный эксперимент по сравнению с лабораторным более грубый по получаемым результатам, однако более приближенный к реальным условиям. В промышленном эксперименте усложняются измерения и сбор информации, значительное влияние на измерительные приборы оказывают различного рода помехи (резко возрастает число неконтролируемых факторов). В связи с этим для обработки результатов промышленного эксперимента требуется использовать специальные статистические методы обработки результатов.

Учебно-исследовательская работа студента

Одним из важнейших документов, имеющих юридическую силу, является *лабораторный журнал*, в котором в последовательном хронологическом порядке указываются условия проведения экспериментов и результаты измерений. Аккуратное ведение лабораторного журнала позволяет исследователю создать адекватный и поддающийся проверке отчет, защитить свой приоритет относительно сделанного им открытия.

Лабораторный журнал представляет собой тетрадь (журнал) с пронумерованными страницами, прошитый толстой ниткой, концы которой скреплены на последней странице официальной печатью учреждения. Данные следует вписывать ручкой, а не карандашом. Если в процессе занесения в журнал результатов эксперимента были позже обнаружены опечатки или фактические ошибки, они исправляются ручкой другого цвета, например, красного, ставится дата и фамилия исправляющего.

Каждый рабочий день в журнале выделяется отдельно: дата в начале рабочего дня и заполнение (Z) до начала следующего (чтобы нельзя было в дальнейшем сделать записи этой датой). Если журнал общий для всей лаборатории, то для каждого эксперимента указывают фамилии его участников. Также для эксперимента необходимо указывать цель, используемые материалы, условия проведения (температура, давление, напряженность электрического поля, частота вращения и т.п.). Это делается как для того, чтобы опыт мог воспроизвести любой другой экспериментатор, так и для самого экспериментатора, чтобы впоследствии можно было проанализировать ход эксперимента, наметить пути повышения точности измерений, продумать следующие эксперименты, учесть все факторы при оформлении научных отчетов и статей.

Перед проведением эксперимента исследователь должен заранее продумать роль различных факторов, стоимость используемых в эксперименте ресурсов, учесть возможные риски для экспериментатора и окружающих, принять необходимые меры безопасности. Все это надо заранее записать в лабораторный журнал, подготовить таблицы для записи однотипных данных.

Типы измеряемых физических величин

Предметом количественного эксперимента являются количественные величины. Для определения абсолютного значения некоторой физической величины ее сравнивают с эталоном, который считается единицей величины. Например, единицей длины является метр, времени – секунда, частоты – Герц.

Различают прямое и косвенное измерения. Наиболее простым является *прямое измерение*, при котором искомое значение величины находят непосредственно с помощью измерительного прибора. Например, длина измеряется линейкой, напряжение – вольтметром и т.п.

Если прямые измерения невозможны, то используют *косвенные измерения*. В них искомое значение величины находят на основании известной зависимости этой величины от других, допускающих прямое измерение. Например,

Учебно-исследовательская работа студента

электрическое сопротивление резистора определяют по падению на нем напряжения и току через него.

Измерения могут быть выполнены как *однократные и многократные*. Однократное измерение дает единственный результат, который принимают за окончательный результат измерения значения искомой величины. Многократное измерение проводят путем повторения однократных измерений одной и той же постоянной физической величины, что приводит к получению набора данных. Окончательный результат многократного измерения, как правило, находят из набора данных в виде среднего арифметического результатов всех отдельных измерений.

Физические величины, встречающиеся в эксперименте, относят к следующим основным *типам*:

- случайная величина;
- постоянная величина;
- изменяющаяся (переменная) величина;
- нестабильная величина.

Случайная величина связана со случайными процессами, поэтому результат отдельного измерения не может быть однозначно предсказан заранее. В то же время проведение достаточно большого числа измерений случайной величины позволяет установить, что результаты измерений отвечают определенным статистически закономерностям. Их выявление, изучение и учет составляют неотъемлемую часть любого эксперимента. В качестве случайных величин можно рассматривать, например, отклонение значения амплитуды сетевого напряжения от номинальной величины.

К *постоянным величинам* можно отнести физические постоянные, например, скорость света в вакууме, заряд электрона и т.п. Кроме того, постоянными величинами можно считать некоторые характеристики конкретного объекта, находящегося при фиксированных условиях. Этот тип физических величин чаще всего встречается в экспериментах, например, при определении теплоемкости. Однако многократные измерения постоянной величины могут дать неодинаковые результаты. Дело в том, что результаты измерений подвержены неконтролируемым влияниям многочисленных воздействий внешней среды, включая и неконтролируемые процессы в исследуемых объектах и используемых измерительных приборах. Вследствие этого постоянная величина зачастую проявляет себя как случайная величина, а результат ее измерений отражает случайную природу воздействия и отвечает определенным статистическим закономерностям. В связи с этим для обработки результатов измерения постоянной величины естественно использовать методы, характерные для обработки результатов измерения случайной величины.

Изменяющаяся (переменная) величина закономерно меняется с течением времени вследствие процессов, проходящих в исследуемом объекте. Примером может служить затухание собственных колебаний в резонаторе. Измерения, проводимые в различные моменты времени, фиксируют величину в новых условиях. Набор результатов однократных измерений представляет собой результаты принципиально неповторимых измерений, так как время нельзя

Учебно-исследовательская работа студента

повернуть вспять, а измерение в целом не может расцениваться как многократное.

Особого внимания заслуживает *нестабильная величина*, которая меняется с течением времени без каких бы то ни было статистических закономерностей. К основной характеристике нестабильной величины следует отнести отсутствие у экспериментатора информации о ее зависимости от времени. Измерения такой величины дают набор данных, не несущих сколько-нибудь полезных сведений. Вместе с тем нестабильная величина может быть переведена в разряд изменяющихся величин, если экспериментально или теоретически установлена закономерность в зависимости ее от времени.

Типы погрешностей измерений

Погрешность – количественная характеристика неоднозначности результата измерений. Ее оценивают исходя из всей информации, накопленной при подготовке и выполнении измерений. Данную информацию обрабатывают для совместного определения окончательного результата измерения и его погрешности.

Погрешность может быть выражена в единицах измеряемой величины x , в таком случае она обозначается Δx и носит название *абсолютной погрешности*. Однако абсолютная погрешность не отражает качества измерений. Например, абсолютная погрешность 1 мм при измерении размеров помещения свидетельствует о высоком качестве измерения. Однако та же погрешность совершенно неприемлема при измерении диаметра тонкой проволоки.

Критерием качества измерения является отношение абсолютной погрешности к окончательному результату измерения:

$$\delta x = \frac{\Delta x}{x}. \quad (2.3)$$

Данное отношение является безразмерным и определяет *относительную погрешность*. Измеряется данная величина как в абсолютном, так и в процентном отношении. Высокой точности измерения соответствует малое значение относительной погрешности.

Выделяют следующие *основные типы погрешностей*:

- *промахи или грубые погрешности*, возникающие вследствие неисправности измерительных приборов или ошибок в эксперименте, сделанных по невнимательности. Обычно легко обнаруживаются и выбрасываются из совокупности результатов, и эксперимент повторяется;

- *приборная погрешность* – систематическая погрешность, присутствующая в результатах измерений, выполненных с помощью любого измерительного прибора. Обычно данная погрешность неизвестна, и ее можно оценить только путем сравнения показаний прибора с показаниями другого, более точного прибора. Иногда результаты специально проведенного сравнения приводят в паспорте прибора, однако чаще указывают максимально возможную погрешность для приборов данного типа;

Учебно-исследовательская работа студента

- *модельная погрешность*. Как уже отмечалось ранее, модель с некоторой степенью точности описывает свойства моделируемого объекта. В связи с этим измеряемые в эксперименте величины, вычисляемые по полученным из модели рабочим формулам, содержат погрешности, которые носят название модельных;

- *случайные погрешности*, возникающие при повторных измерениях и отражающие случайную природу явления. Возникают вследствие множества причин, совместное воздействие которых на каждое отдельное измерение невозможно учесть или заранее установить. Единственно возможный способ объективного учета их состоит в определении их статистических закономерностей, проявляющихся в результате многократных измерений. Рассчитанные статистические оценки вносят в окончательный результат измерения.

Лекция 3. Фиксация и обработка статистических данных

Принципы подбора выборки

На прошлой лекции было показано, что результатом проведения экспериментальных исследований является совокупность измерений. **Результат измерений**, т.е. совокупности операций, выполняемых с помощью технического средства, хранящего единицу величины, позволяющего сопоставить измеряемую величину с ее единицей и получить значение величины, должен сопровождаться указанием погрешности, с которой он получен.

Для оценки получаемых в результате эксперимента измерений проводится обработка результатов с помощью вероятностно-статистических методов теории вероятностей и математической статистики.

Сначала выполняется предварительная обработка, состоящая в отсеивании грубых погрешностей и оценке достоверности результатов измерений.

Поскольку однократные измерения допустимы только в порядке исключения, так как они по существу не позволяют судить о достоверности измерительной информации, то, в основном, проводятся многократные измерения. Однако при этом возникает вопрос об их количестве. Полный набор всех возможных значений, которые может принимать случайная величина при бесконечном числе испытаний, называется **генеральной совокупностью**. В распоряжении исследователя никогда нет генеральной совокупности, и он может изучать только ее часть – выборку, причем всегда ограниченного объема.

Выборка – набор n значений величин x_i , полученный из генеральной совокупности в результате конечного числа испытаний N , называют выборкой объема N . **Объем** – количество единиц в выборке.

По отдельной выборке с помощью статистической обработки набора величин x_i пытаются как можно точнее описать характеристики генеральной совокупности.

Учебно-исследовательская работа студента

Следует отметить, что сам процесс выборки может являться источником ошибок. Их принято называть **ошибками репрезентативности**. Однако правильная организация выборки позволяет их избежать.

Подбор выборки оказывает основное влияние на качество получаемых данных. Известны два основных принципа составления выборки: повторности и рандомизация. Повторности необходимы для уверенности в полученных результатах, а рандомизация – для избегания отклонений, вызванных посторонними причинами.

Принцип повторности выражается в использовании многократного эксперимента. Главное правило – повторения должны быть независимы друг от друга. Это значит, что нельзя в качестве повторений рассматривать данные, полученные в последовательные промежутки времени с одного и того же объекта или с одного и того же места. Например, при снятии показаний прибора необходимо осуществлять их как по мере возрастания показателя, так и в обратном порядке. Длина выборки зависит от требуемой точности получаемого результата, как $\text{точность} = \sqrt{\text{длина выборки}}$.

Рандомизация – еще одно условие создания выборки, подразумевающие выбор элементов исследований случайным образом. Важным условием рандомизации является то, что каждый объект должен иметь абсолютно те же самые шансы быть выбранным, что и все прочие объекты. Требование рандомизации приводит к серьезным затратам на проведение исследования. В связи с этим рандомизацию осуществляют лишь частично. Ограничение вводится с учетом характера генеральной выборки. Так, если определяют время старения кабелей определенного типа, то данным типом и ограничиваются.

Статистический анализ полученных данных выборки позволяет:

- дать для больших выборок общие характеристики, отражающие так называемую центральную тенденцию, т.е. число (или ряд чисел), вокруг которых «рассыпаны» данные, а также степень их разброса;
- проводить сравнение нескольких выборок и определять вероятность того, что их различия вызваны случайными причинами, а также оценить общие характеристики;
- получить сведения о взаимосвязях элементов в выборке (соответствия, при которых два явления чаще встречаются вместе, чем порознь, корреляция, позволяющая численно оценить силу взаимосвязи, зависимости, для которых можно определить и силу, и направление, и оценить, насколько вероятно, что они – результат случайных причин);
- установить с помощью многомерной статистики структуру. При этом создается и проверяется качество классификации объектов;
- применение результатов анализа для предсказания и описания. Методы предсказания позволяют выяснить, с какой вероятностью может быть верным или неверным наш вывод. Описательные методы сообщают информацию о данных без подтверждения какими-либо вероятностными методами.

Определение параметров экспериментального распределения

2.1 Предварительная обработка данных начинается с определения, какими типами переменных (признаков) представлены данные. Выделяют **три типа переменных**:

- *непрерывные* – представлены действительными числами (например, длина или вес);

- *дискретные* – представлены целыми, как правило, положительными числами;

- *категориальные* (например, марка кабеля, тип материала, географический регион). Значения категориальных данных не могут быть положены на числовую прямую.

2.2 Очень важным при проведении статистического анализа является выяснение того, на какое распределение более всего подходит полученное экспериментальное распределение случайной величины. Оценка степени совпадения эмпирического закона распределения с теоретическим проводится в два этапа:

- определяют параметры экспериментального распределения;

- производят оценку по Колмогорову соответствия экспериментального распределения выбранному теоретическому.

Определение параметров экспериментального распределения производят с помощью моментов распределения. Для пояснения данного процесса вернемся к построению гистограммы и определению понятия распределения.

Построение гистограммы осуществляется следующим образом. Пусть выборка представляет собой серию из N измерений, содержащая некоторый набор значений x_1, x_2, \dots, x_N . Для построения гистограммы по оси абсцисс откладывают полученные в отдельных измерениях значения x_i , как показано на рисунке 3.1,а. Причем ось Ox разбивается на равные интервалы с каким-то дискретом Δx . В выборке подсчитывается число Δn_k , в результате которых получены значения x , лежащие в интервале $x_k \pm 0.5\Delta x$ (здесь x_k - координата центра интервала на оси Ox). На каждом интервале строится прямоугольник высотой Δn_k и шириной Δx (точки, лежащие точно на границе интервала, будем всегда относить, например, в левый столбик). В полученной гистограмме сумма высот прямоугольников равна полному числу экспериментов в данной серии.

Однако для сравнения различных по объему выборок удобнее строить нормированную гистограмму, когда по оси ординат откладывается величина $\frac{1}{N} \frac{\Delta n_k}{\Delta x_k}$ ($N = \sum n_k$), как показано на рисунке 3.1,б. В этом случае произведение

высоты на Δx , т.е. площадь столбика, имеет размерность вероятности попадания результата отдельного измерения в данный интервал Δx , а суммарная площадь под всей гистограммой равна 1:

$$\sum \frac{1}{N} \frac{\Delta n_k}{\Delta x_k} \Delta x_k = 1. \quad (3.1)$$

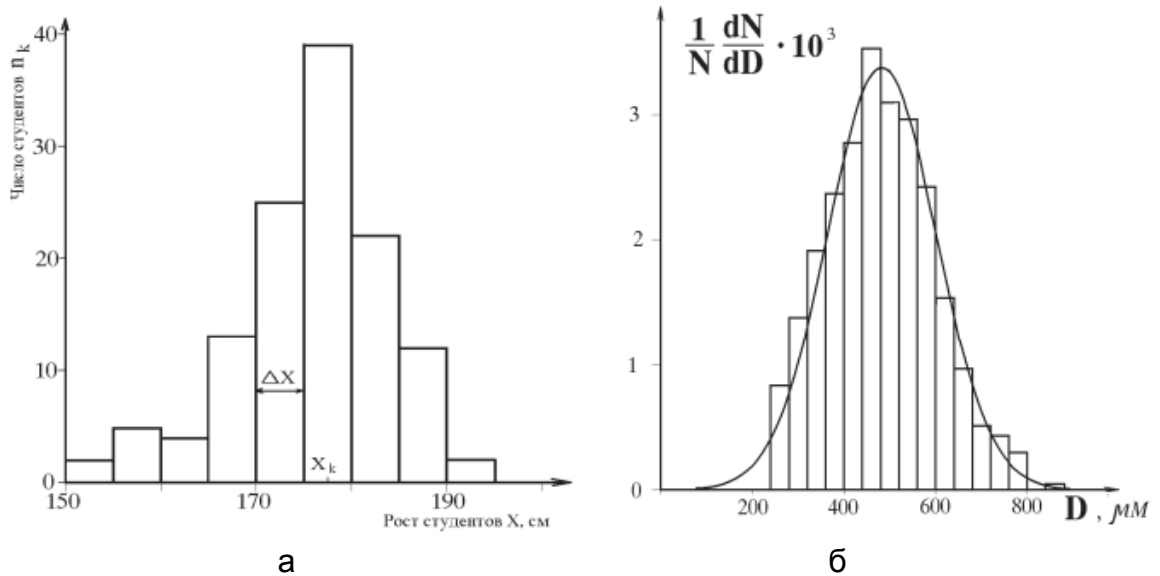


Рисунок 3.1 – Построение гистограммы

Различают нормальное распределение данных (чем больше значение признака отличается от его среднего по выборке значения, тем реже это значение встречается в выборке) и распределение данных, отличное от нормального. На рисунке 3.1 приведены примеры распределений. Причем рисунки с индексами «а» и «б» иллюстрируют данные, которые можно отнести к распределенным по нормальному закону распределения, а с индексами «в» и «г» - к тем, что нельзя отнести к нормальному закону распределения.

Если число измерений N достаточно велико, то ширину интервала можно сделать очень малой. Тогда в пределе вместо гистограммы получим график типа показанного на рисунке 3.1,б, на котором по оси ординат отложена величина $f(x)$, имеющая размерность $|x^{-1}|$ и пропорциональна доле числа отсчетов n_k/n , попадающих в каждый интервал. Такой график называют **кривой распределения**, а функцию $f(x)$ называют **плотностью вероятности**. Полагается, что

$$\int_{-\infty}^{\infty} f(x)dx = 1. \quad (3.2)$$

В случае дискретной величины вместо $f(x)dx$ используют вероятность p_i получить в результате измерения значение x_i .

Вероятность попадания измеряемой величины в интервал $(-\infty, x]$ называют **функцией распределения** или **интегральной функцией распределения**:

$$F(x) = \int_{-\infty}^x f(z)dz. \quad (3.2)$$

Учебно-исследовательская работа студента

Анализ соотношений (3.2) и (3.1) показывает, что плотность вероятности $f(x)$ равна производной функции распределения $F(x)$ в точке x . Если проинтегрировать плотность вероятности в пределах от x_1 до x_2 , то полученная величина будет представлять вероятность P того, что результат отдельного измерения будет лежать в интервале $[x_1, x_2]$:

$$P(x_1 < x < x_2) \equiv \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1). \quad (3.3)$$

Смысл введения функции (3.3) заключается в том, что при заданной вероятности распределения результат любого отдельного измерения с достоверностью P даст величину, лежащую при определении понятия **доверительного интервала** или **доверительной вероятности**.

Для определения характера распределения вычисляют статистические моменты, к которым относятся:

- **первый момент (среднее арифметическое, математическое ожидание):**

$$M_x = m_1 = \frac{1}{N} \sum_{i=1}^N (x_i - A)^1 = \int_{-\infty}^{\infty} (x - A)^1 f(x)dx, \quad (3.4)$$

где $A = \begin{cases} 0, & \text{начальный момент,} \\ X, & \text{центральный момент.} \end{cases}$

Первый момент указывает на центр тяжести в геометрии распределения, как показано на рисунке 3.2,а.

- **второй момент (или дисперсия, разброс)**, вычисляемая по формуле:

$$D_x = m_2 = \frac{1}{N} \sum_{i=1}^N (x_i - M_x)^2 = \int_{-\infty}^{\infty} (x - M_x)^2 f(x)dx. \quad (3.5)$$

Кроме того, используют понятие среднеквадратичного отклонения:

$$\sigma = \sqrt{D_x}. \quad (3.6a)$$

Дисперсия характеризует величину разброса экспериментальных данных относительно центра тяжести m_1 . Таким образом, по величине m_2 можно судить о втором параметре геометрии распределения.

Кроме этого используют понятие стандартного отклонения:

$$\sigma_{cm} = \sigma \begin{cases} 1/\sqrt{N}, & N > 30 \\ 1/\sqrt{N-1}, & N \leq 30 \end{cases}. \quad (3.6b)$$

- **третий момент** характеризует **асимметрию** (или **скошенность**). На практике обычно используется его нормированная величина, вычисляемая по формуле:

$$A_x = R_3 = \frac{1}{N} \frac{\sum_{i=1}^N (x_i - M_x)^3}{\sigma_{cm}^3} = \frac{1}{\sigma_{cm}^3} \int_{-\infty}^{\infty} (x - M_x)^3 f(x)dx. \quad (3.7)$$

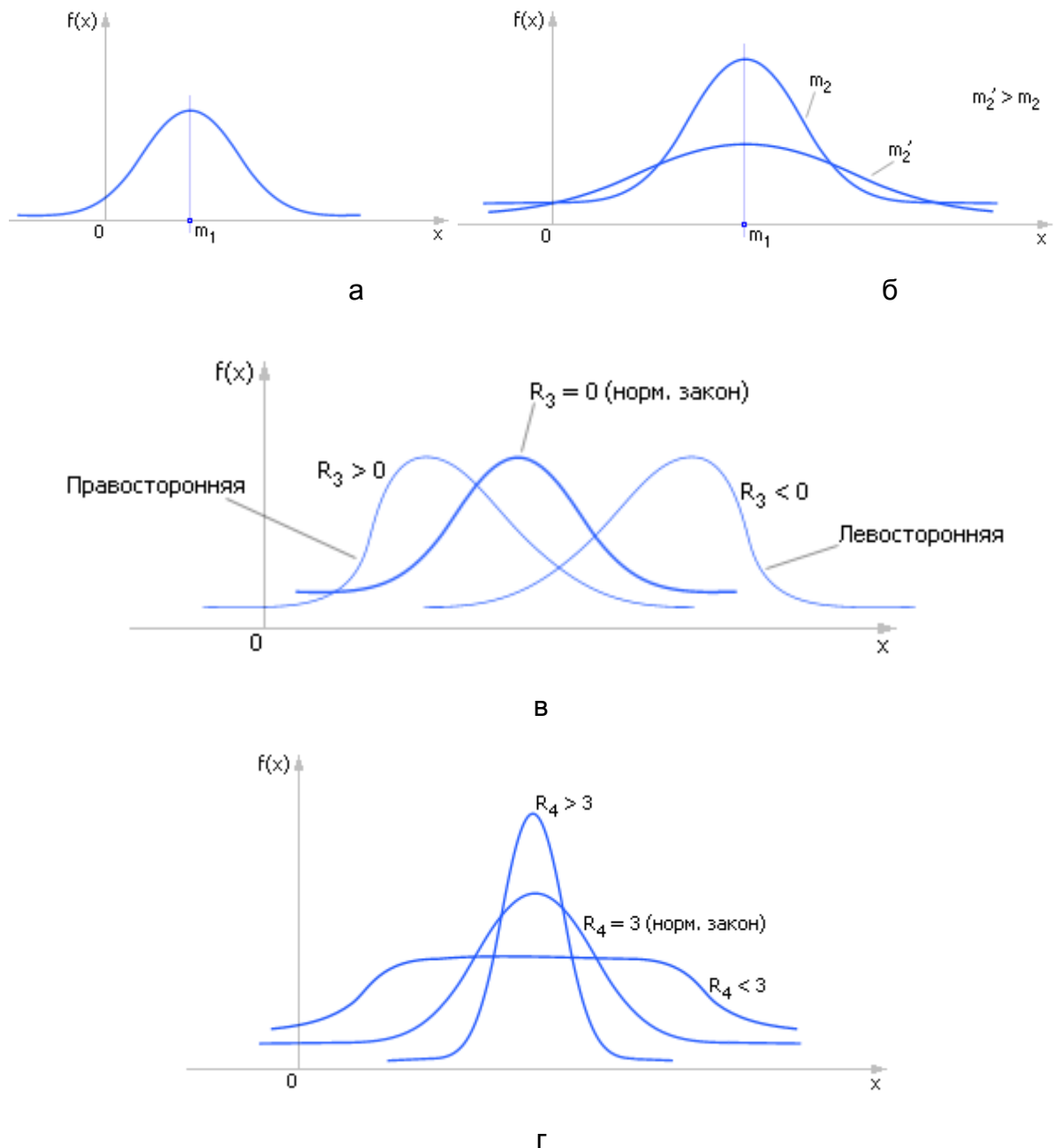


Рисунок 3.2 – Характерное положение моментов на графике распределения статистической величины

Определяя знак R_3 , можно определить, есть ли асимметрия у распределения и в какую сторону (см. рисунок 3.2,в).

- *четвертый момент характеризует эксцесс* (или *островершинность*) (см. рисунок 3.2,г). Так же, как и асимметрия на практике используется не сама величина эксцесса, а ее нормированное значение:

$$E_x = R_4 - 3 = \frac{1}{N} \frac{\sum_{i=1}^N (x_i - M_x)^4}{\sigma_{cm}^4} - 3 = \frac{1}{\sigma_{cm}^4} \int_{-\infty}^{\infty} (x - M_x)^4 f(x) dx - 3. \quad (3.7)$$

Учебно-исследовательская работа студента

Помимо перечисленных выше моментов функции распределения при анализе используется производная оценка от величины математического ожидания и дисперсии – **коэффициент вариации**, определяемый в процентах:

$$V = \frac{\sigma}{M_x} 100\% . \tag{3.8}$$

Оценка размера выборки

Крайне важным, как уже отмечалось выше, является вопрос, сколько экспериментов следует сделать, чтобы можно было доверять снятым характеристикам. Обычно исследователь задает **доверительную вероятность**, т.е. вероятность, с которой он готов доверять снятым характеристикам. Чем больше будет доверительная вероятность, тем больше экспериментов потребуется сделать.

Для оценки размера выборки используем центральную предельную теорему, утверждающую, что сумма (или среднее) случайных величин есть величина неслучайная. Значения вычисленной статистической характеристики будут распределены по нормальному закону.

Обозначим как Q **доверительную вероятность**, т.е. вероятность того, что частота p отличается от теоретической вероятности P не более чем на ε . Тогда по теореме Бернулли имеем:

$$Q(|p - P| \leq \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma}\right) = \Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right). \tag{3.9}$$

Величина ε называется **доверительным интервалом**, функция $\Phi(\cdot)$ называется функцией Лапласа или интегралом вероятности и будет рассмотрена на следующей лекции. Из соотношения (3.9) несложно выразить требуемое значение для доверительной вероятности количества экспериментов n :

$$n = \frac{P(1-P)}{\varepsilon^2} \left(\Phi^{-1}(Q)\right)^2, \tag{3.10}$$

где $\Phi^{-1}(\cdot)$ - обратная функция Лапласа.

Рассмотрим пример использования данной методики. Пусть при моделировании выпускаемой продукции предприятием в результате имитации его работы в течение 50 дней были получены выходные данные, приведенные в Таблице 3.1.

Т а б л и ц а 3.1 – Экспериментальные статистические данные

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

В соответствии с ней всего было проведено $15+10+5+20=50$ экспериментов ($n = 50$). Зададимся доверительной вероятностью к ответам модели $Q = 0.9$ и

Учебно-исследовательская работа студента

доверительным интервалом $\varepsilon = 0.05$. Для оценки будем оценивать результат статистических экспериментов по худшей из частот (по максимальной частоте. В нашем случае $p = 0.4$). С использованием формулы (3.10) получаем, что $n = 250$.

Последовательно вычислений следующая: $\Phi^{-1}(0.9) = 1.65$, $(\Phi^{-1}(0.9))^2 = 2.7$,

$$n = \frac{0.4 \cdot 0.6}{0.05^2} 2.7 = 250.$$

Таким образом, число проведенных экспериментов недостаточно для получения результатов с достоверностью 0.9. Необходимо продолжать эксперименты, чтобы достичь требуемой точности.

Приведенный пример показывает, что формула (3.10) является рекуррентной. Сразу вычислить с ее помощью длину выборки нельзя, требуется провести пробную серию экспериментов. Следует также иметь в виду, что оценивание ведут по худшей из частот. Это обеспечивает достоверный результат сразу по всем снимаемым характеристикам.

Графики зависимости числа требуемых экспериментов от величины требуемой вероятности и доверительного интервала для ряда частных значений приведены на рисунке 3.3.

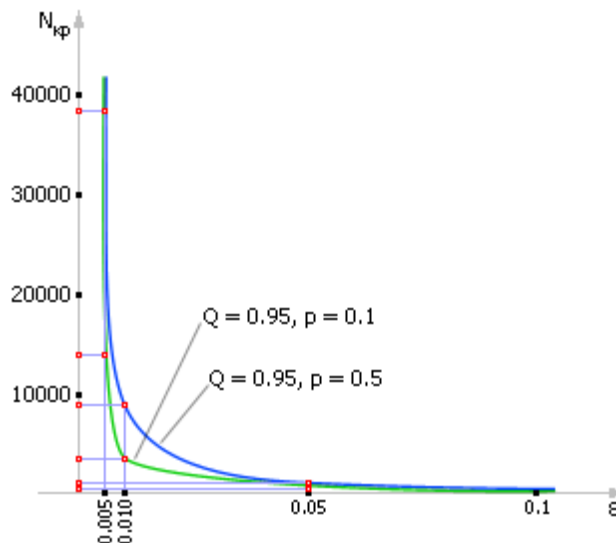


Рисунок 3.3 – Зависимость количества требуемых экспериментов от величины доверительной вероятности и доверительного интервала

Приведенная оценка количества экспериментов не является единственной. Известны аналогичные близкие по смыслу оценки Бернулли, Муавра-Лапласа, Чебышева.

Оценки погрешности измерений

К вычисляемым в результате эксперимента оценкам случайных величин предъявляются **три основных требования**:

- состоятельности,

Учебно-исследовательская работа студента

- несмещенности;
- эффективности.

Полагают, что оценка состоятельна, если с ростом объема выборки она стремится по вероятности к истинному значению, несмещена, если ее математическое ожидание стремится к истинному значению, и эффективна, когда оценка обладает наименьшим рассеянием по сравнению с другими оценками. Из двух оценок эффективнее та, которая обладает меньшей дисперсией.

С учетом сказанного выше выработаны следующие **требования прямого многократного измерения**:

1. Провести многократные измерения при одних и тех же условиях и записать их в таблицу.

2. Рассчитать среднее значение по формуле:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3.11)$$

3. Вычислить оценку дисперсии:

$$\bar{\sigma}^2 = \sigma_n^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}. \quad (3.12)$$

4. Вычислить среднеквадратическую ошибку среднего

$$\sigma_x = \frac{\sigma_n}{\sqrt{n}}. \quad (3.13)$$

5. Задавшись требуемым уровнем доверительной вероятности P , определить по соответствующей таблице коэффициент Стьюдента $t(P, n - 1)$ и модуль доверительного интервала

$$\Delta x = \sigma_x t(P, n - 1). \quad (3.14)$$

6. Округлив соответствующие результаты, записать ответ в виде

$$X = \bar{x} \pm \Delta x \quad \text{при доверительной вероятности } P. \quad (3.15)$$

Коэффициент Стьюдента описывает значение распределения Стьюдента (или t -распределения). Данное распределение отличается от нормального меньшей величиной максимума и более длинными «хвостами».

Оценка погрешности при косвенных измерениях. В большинстве экспериментов измеряется не непосредственная величины, подлежащая определению, а другая величина или ряд величин, зависящих от интересующей нас величины тем или иным образом. Например, сопротивление R может быть вычислено из закона Ома

$$R = \frac{U}{I} \quad (3.16)$$

по измеренным значениям напряжения и тока. При таких измерениях, являющихся косвенными, необходимо уметь вычислить ошибку результата. Когда интересующая величина y является функцией измеряемых величин x_1, x_2, \dots, x_k , т.е. $y = f(x_1, x_2, \dots, x_k)$, а для случайных ошибок, измеряемых непосредственно,

Учебно-исследовательская работа студента

справедливы описанные выше допущения, необходимо действовать следующим образом:

1. Оценка математического ожидания \bar{y} функции нескольких измеренных переменных вычисляется по формуле:

$$\bar{y} = f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k). \quad (3.17)$$

2. Оценка среднеквадратического отклонения вычисляется способом, основанном на дифференциальном исчислении. Если среднеквадратичные ошибки $\sigma_{x_1}, \sigma_{x_2}, \dots, \sigma_{x_k}$ величин x_1, x_2, \dots, x_k , определенные по данным измерений малы по сравнению с измеряемыми величинами, то

$$\sigma_y = \sqrt{\left(\frac{\partial f}{\partial x_1} \sigma_{x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2} \sigma_{x_2}\right)^2 + \dots + \left(\frac{\partial f}{\partial x_k} \sigma_{x_k}\right)^2}, \quad (3.18)$$

где $\partial f / \partial x_i$ - частная производная функции f , вычисленная при значении переменных, соответствующих средним значениям $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$.

3. Для определения доверительного интервала используется та же процедура, что и при определении доверительного интервала непосредственно измеряемой величины.

Сравнивая слагаемые в подкоренном выражении, легко определить, ошибки какой из измеряемой величин вносят основной вклад в суммарную ошибку. Предварительный анализ выражения для σ_y полезен для того, чтобы при планировании постановки эксперимента попытаться добиться повышенной точности измерения величины, вносящей основной вклад в ошибку, а также понять, какие измерения следует производить тщательно, а на какие не следует тратить больших усилий.

Лекция 4. Основные законы распределения, применяемые при обработке данных научного эксперимента

Нормальный закон распределения и его основные свойства

На прошлой лекции говорилось, что основной целью статистической обработки экспериментальных данных является формулировка закономерностей по данным выборки и распространение их на генеральную совокупность. Основной получения точных статистических результатов является правильное определение закона распределения. Какими могут быть законы распределения?

При обработке данных измерений в науке и технике обычно предполагают **нормальный закон распределения**. При анализе в молекулярно-кинетической теории данный закон называется **распределением Максвелла**, а при анализе случайных ошибок название определяется типом данных. Для непрерывных данных закон называется **распределением Гаусса**, а для дискретных данных –

Учебно-исследовательская работа студента

распределением Пуассона. Нормальный закон распределения хорошо описывает разброс непрерывной случайной величины при большом числе независимых случайных погрешностей и всегда проявляется тогда, когда суммарная погрешность есть результат неучтенного совместного воздействия множества причин, каждая из которых дает малый вклад в погрешность. Причем совершенно неважно, по какому закону распределен каждый из вкладов в отдельности.

Функция плотности вероятности для нормального закона имеет вид:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - M_x)^2}{2\sigma^2}\right). \quad (4.1)$$

График функции плотности вероятности для нормального закона распределения для трех значений среднеквадратичного отклонения σ приведен на рисунке 4.1. Анализ данного рисунка показывает, что чем меньше величина σ , тем плотнее вблизи среднего значения (математического ожидания $M_x = X_0$) лежат результаты измерения.

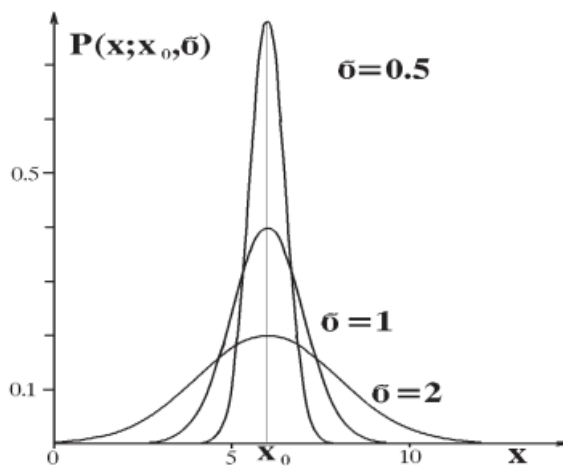


Рисунок 4.1 – График плотности вероятности для нормального закона распределения

Пусть $\Delta x = x - M_x$ - произвольное отклонение от средней величины. Введем ε - величину отношения полуширины интервала Δx к среднему квадратичному отклонению σ :

$$\varepsilon = \frac{\Delta x}{\sigma}. \quad (4.2)$$

В Таблице 4.1 указана вероятность α , позволяющая оценить попадание случайной величины в интервал $M_x - \varepsilon\sigma \leq x \leq M_x + \varepsilon\sigma$:

$$\alpha = P(M_x - \varepsilon\sigma \leq x \leq M_x + \varepsilon\sigma). \quad (4.3)$$

Вероятность α можно также рассчитать по приближенному выражению:

Учебно-исследовательская работа студента

Т а б л и ц а 4.1 – Доверительные интервалы $(M_x - \Delta x; M_x + \Delta x)$ для доверительной вероятности α (в долях ε)

α	0,68	0,90	0,95	0,990	0,997	0,999
ε	1,0	1,65	2,0	2,6	3,0	3,3

$$\alpha \approx \sqrt{1 - \exp\left(-\frac{2\varepsilon^2}{\pi}\right)}. \quad (4.4)$$

Анализ данных Таблицы 4.1 показывает, что с вероятностью 68% результат измерения попадет в интервал $(M_x - \sigma \leq x \leq M_x + \sigma)$, 95% результатов измерений попадут в интервал $(M_x - 2\sigma \leq x \leq M_x + 2\sigma)$ и 99% результатов попадут в интервал $(M_x - 3\sigma \leq x \leq M_x + 3\sigma)$, т.е. в указанный интервал не попадет только один результат из трехсот. Изображение распределения данных интервалов на графике функции плотности вероятности приведено на рисунке 4.2.

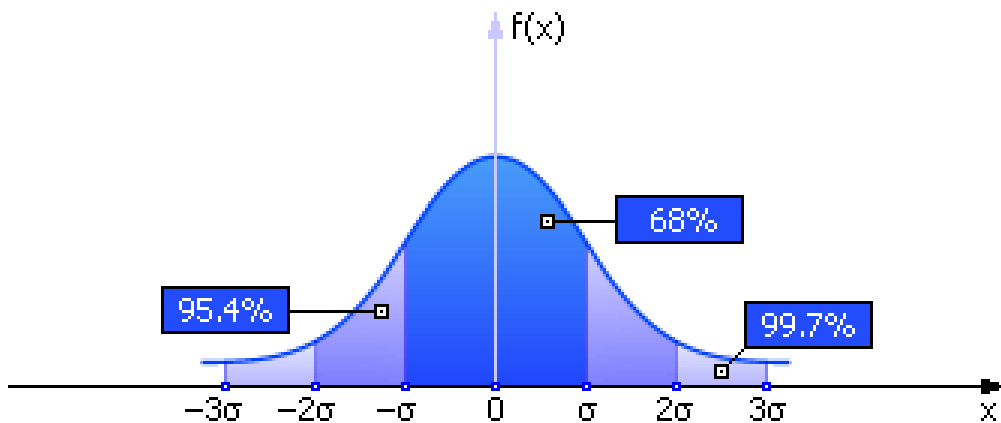


Рисунок 4.2 – Графический вид нормального закона распределения

При обработке данных результатов эксперимента часто используют «правило 3σ » или «правило трех стандартгов», которое основано на указанном свойстве нормального распределения. С учетом проведенного выше анализа можно установить наличие промаха в результате отдельного измерения, а значит, отбросить его, если результат измерения отличается более чем на 3σ от измеренного среднего значения случайной величины.

В то же время стоит более тщательно повторить измерения в этой области параметров. Возможно, данный результат является не промахом, а свидетельствует о наличии необычного поведения изучаемой системы, например, о резонансе.

Поскольку данный закон распределения наиболее часто используется в науке и технике, рассмотрим его **основные свойства**:

1. Изменение величины математического ожидания приводит к сдвигу кривой по оси Ox , как показано на рисунке 4.3,а.
2. Изменение среднеквадратичного отклонения σ приводит к масштабированию формы по оси Ox , как показано на рисунке 4.1. Чем более не

Учебно-исследовательская работа студента

случаен процесс, тем меньше его среднеквадратичное отклонение, тем уже и выше колокол на графике.

3. Для нормального закона распределения интегральная функция распределения $F(x)$ определяется *функцией Лапласа*:

$$F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt. \quad (4.5)$$

В общем виде данный интеграл не берется, поэтому функция Лапласа задана в виде таблицы. Причем в таблице учтено, что для нормального распределения $F(-x) = F(x)$.

Если задается интервал интегрирования функции Лапласа $[a, b]$, то

$$\begin{aligned} P(a < x < b) &= \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\left(-\frac{(x - M_x)^2}{2\sigma^2}\right) dx = \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\frac{a - M_x}{\sigma}}^{\frac{b - M_x}{\sigma}} \exp\left(-\frac{t^2}{2}\right) dt = F(b) - F(a). \end{aligned} \quad (4.6)$$

Например, для правила «трех сигм»

$$P(|x - M_x| < 3\sigma) = 2F(3) - 1 = 2 \cdot 0.9987 - 1 = 0.9973.$$

4. Нормальное распределение является пределом для различного вида распределений, вытекающее из центральной предельной теоремы: «Для большого числа N случайных величин X с любым законом распределения их сумма есть случайное число с нормальным законом распределения».

Частным случаем нормального распределения является **логарифмически-нормальное распределение (логнормальное распределение)**. Функция плотности вероятности для него имеет вид:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln x - M_x)^2}{2\sigma^2}\right). \quad (4.7)$$

Вид функции плотности логнормального распределения приведен на рисунке 4.3. Данному распределению подчиняется, например, размер частиц при дроблении какого-либо материала.

Математическое ожидание и дисперсия для данного закона определяются выражениями:

$$M_x = \exp(M_{xn}), \quad D_x = \exp(D_{xn}), \quad (4.8)$$

где M_{xn} , D_{xn} - соответственно математическое ожидание и дисперсия нормального закона распределения $N(M_{xn}, D_{xn})$, связанного с логнормальным соотношением:

$$X \sim \ln(N(M_{xn}, D_{xn})).$$

Учебно-исследовательская работа студента

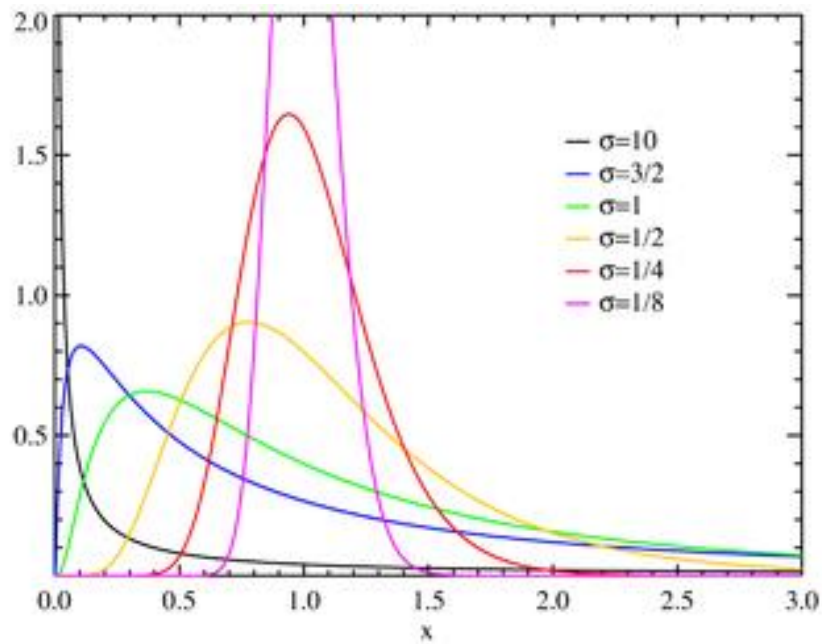


Рисунок 4.3 – Плотность вероятности логнормального распределения

Логнормальный закон распределения и распределение Вейбулла

Еще одними из видов распределений, встречающихся при анализе экспериментальных данных в радиотехнике, являются двухпараметрические распределения: **двойное экспоненциальное распределение (распределение Лапласа)** и **распределение Вейбулла (распределение Релея)**.

Функция плотности распределения для распределения Лапласа имеет вид:

$$f(x) = \frac{1}{2\lambda} \exp\left(-\frac{|x - M_x|}{\lambda}\right). \quad (4.9)$$

На рисунке 4.4 приведены графики функции плотности распределения и интегральной функции распределения для распределения Лапласа. С помощью двойного распределения Лапласа описываются, в частности, процессы теплообмена океана с атмосферой, а также кинетика квазихрупкого разрушения.

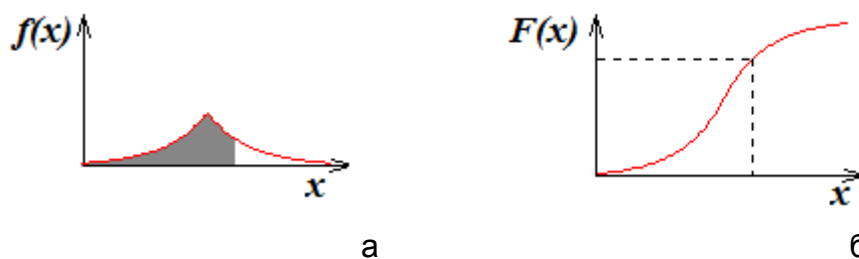


Рисунок 4.4 – Распределение Лапласа: а – функция плотности вероятности; б – интегральная функция распределения

Учебно-исследовательская работа студента

Распределение Вейбулла используется для исследования интенсивности отказов для периодов приработки и старения в электрической сети. Как известно, надежность наиболее распространенных элементов электрических сетей в значительной степени определяется надежностью работы изоляции, «прочность» которой изменяется в течение эксплуатации, которая, в свою очередь, определяется механической прочностью, эластичностью, исключающей возможности образования неоднородностей.

Неоднородности в изоляции возникают в основном в результате нагревания токами нагрузок и температурных воздействий внешней среды. Механические нагрузки (вибрации, деформации, удары, а так же нагрузки, вызванные электрическими процессами и др.) также приводят к разрушению изоляции. При воздействии переменных неблагоприятных условий неоднородности материала увеличиваются, например микротрещина распространяется вглубь изоляции и при случайном повышении напряжения может вызвать пробой изоляции. Причиной отказа может быть даже небольшая неоднородность материала.

Число неблагоприятных воздействий (тепловых или электромеханических), вызывающих пробой изоляции, есть функция, убывающая в зависимости от размеров неоднородности. Чем больше по размерам неоднородность, тем меньше факторов в совокупности нужно, чтобы пробить изоляцию.

В случае отсутствия так называемого «порога чувствительности», то есть периода времени, за которое сеть гарантировано не откажет, безотказную работу такой сети можно описать при помощи распределения Вейбулла. Для него функция плотности распределения определяется формулой:

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right), & x \geq 0, \\ 0, & x < 0 \end{cases}, \quad (4.10)$$

где $\lambda > 0$ - коэффициент масштаба, $k > 0$ - параметры распределения.

В случае $k=1$ закон Вейбулла сводится к экспоненциальному закону распределения. При $k \gg 2$ функция распределения времени безотказной работы достаточно хорошо аппроксимируется нормальным законом распределения в окрестности среднего времени безотказной работы.

При соответствующем подборе параметра k можно с помощью закона Вейбулла описывать надежность как стареющих элементов (период старения и износа), так и надежность элементов, имеющих скрытые дефекты (период приработки).

Математическое ожидание и дисперсия данного распределения определяются формулами:

$$M_x = \lambda \Gamma\left(1 + \frac{1}{k}\right), \quad D_x = \lambda^2 \left(\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right). \quad (4.11)$$

В данных соотношениях $\Gamma(\cdot)$ - гамма-функция.

Оценка согласия эмпирических и теоретических распределений

Для оценки степени совпадения эмпирического закона распределения с теоретическим применяют стандартные критерии согласия. Всего их известно до 200, но применяются практически, в основном, два: критерий Колмогорова и критерий Пирсона (критерий χ^2).

Рассмотрим **применение критерия Колмогорова**, известный также как критерий согласия Колмогорова-Смирнова.

Оценка степени совпадения эмпирического (т.е. полученного с помощью эксперимента) закона распределения с теоретическим осуществляется следующим образом:

1. Определяется, на сколько интервалов K надо разбить при дискретизации распределение. Это необходимо сделать, поскольку оценка касается непрерывного распределения, а измерения дают дискретные данные.

При определении K используется правило Стерджеса:

$$K = 1 + \log_2 N = 1 + 3.322 \lg N. \quad (4.12)$$

Здесь N - количество случайных значений (проведенных опытов).

2. Полученная в ходе экспериментальных данных выборка сортируется в порядке возрастания значений случайной величины, и определяются границы ее изменения (нижняя граница интервала $a = \min(X)$, верхняя граница интервала $b = \max(X)$).

3. Интервал $[a, b]$ изменения случайной величины разбивается на K подынтервалов размером Δ , для каждого из которых вычисляется нижняя граница x_j ($1 \leq j \leq K$).

4. Вычисляется эмпирическая функция распределения F_j^{emp} , построенная по выборке и определяющая частоту попадания случайной величины в j -й интервал:

$$F_j^{emp} = \frac{1}{N} \sum_{i=1}^N I_{x_i \leq x_j}, \quad (4.13)$$

где

$$I_{x_i \leq x_j} = \begin{cases} 1, & x_i \leq x_j; \\ 0, & x_i > x_j. \end{cases}$$

(4.14)

Для полученной интегральной функции $F_j^{emp} = P(x_i \leq x_j)$ строится ступенчатый график, как показано на рисунке 4.5.

5. Для полученной сгруппированной выборки вычисляются моменты распределения m_1, m_2, m_3, \dots . Число моментов равно числу неизвестных в теоретическом законе распределения. Для гипотезы о нормальном законе распределения необходимо вычислить два момента: первый (математическое

Учебно-исследовательская работа студента

ожидание или среднеарифметическое значение) M_x и второй (дисперсию или разброс) $D_x = \sigma_r^2$.

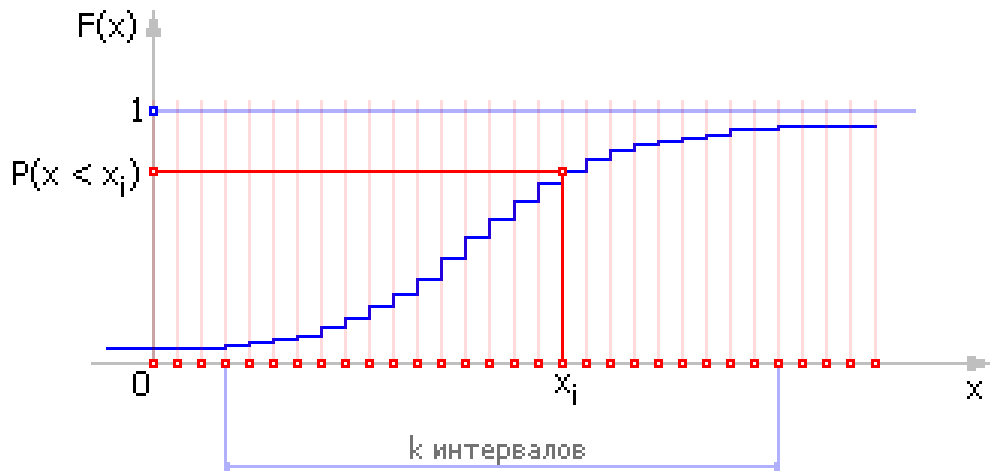


Рисунок 4.5 – Интегральный закон эмпирического распределения, дискретный вариант

6. Строится ступенчатый график для теоретического (непрерывного) распределения в предположении, что последний является нормальным. Интегральная функция при этом будет определяться соотношением:

$$F_j^{teor} = \frac{1}{2} \operatorname{erf}\left(\frac{x_j - Mx}{\sigma_r}\right) + \frac{1}{2}. \quad (4.15)$$

Данная функция в виде ступенчатого графика наносится на поле зависимости эмпирической функции $F_j^{emp} = P(x_i \leq x_j)$ (см. рисунок 4.6).

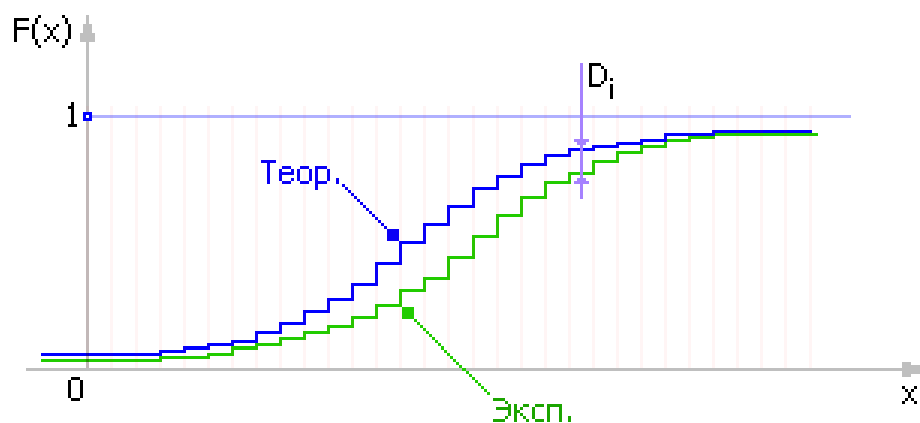


Рисунок 4.6 – Сравнение теоретического и эмпирического интегральных распределений случайной величины (дискретный вариант)

7. Вычисляется статистика критерия для эмпирической функции распределения и определяется точная верхняя ее граница:

Учебно-исследовательская работа студента

$$D = \max_j |F_j^{emp} - F_j^{teor}|. \quad (4.16)$$

8. Вычисление квантиля распределения Колмогорова K_α , имеющего правостороннюю критическую область. Если N достаточно велико ($N > 15$), то для его нахождения может быть использована формула:

$$K_\alpha \approx \sqrt{-\frac{1}{2} \ln \frac{\alpha}{2}}, \quad (4.17)$$

где $\alpha = 1 - Q$ - заданный уровень значимости; Q - вероятность получения достоверных результатов (надежность результатов). Значения квантиля распределения Колмогорова для некоторых значений уровня значимости приведены в Таблице 4.2.

Т а б л и ц а 4.2 – Таблица критерия Колмогорова

Q	0,85	0,90	0,95	0,99
λ	1,14	1,22	1,36	1,63

9. Гипотеза о нормальном законе распределения случайной величины X считается достоверной, если выполняется неравенство:

$$\sqrt{KD} < K_\alpha. \quad (4.18)$$

При обработке данных измерений в науке и технике обычно предполагают нормальный закон распределения случайных погрешностей измерений. Однако могут встречаться и другие законы распределений, например, при оценке электрической прочности диэлектриков.

Достоинством критерия Колмогорова является простота его применения, однако по степени доверия к результатам идентификации он уступает **критерию Пирсона (критерию хи-квадрат)**. Преимуществом данного критерия является возможность сопоставлять распределения признаков, представленных в любой шкале, начиная со шкалы наименований. Чем больше расхождение между двумя сопоставляемыми распределениями, тем больше эмпирическое значение хи-квадрат.

В соответствии с ним после определения числа интервалов разбиения по формуле (4.12) строится статистический ряд. Длина получаемой выборки меньше исходной и представляет так называемую *сгруппированную выборку*:

Варианты: $x_1, x_2, \dots, x_s,$

Частоты: $n_1, n_2, \dots, n_s.$

Значения вариант, попавших в каждый интервал, приближенно равны числу, задающему середину интервала, т.е. x_i - значения середин вариант; n_i - число наблюдений (вариант), попавших в i -й интервал (эмпирические частоты).

По полученным данным вычисляются выборочное среднее \bar{x}_B и выборочное среднее квадратическое отклонение σ_B .

Выдвигается предположение о том, что генеральная совокупность распределена по нормальному закону.

Учебно-исследовательская работа студента

На основании выдвинутой гипотезы о виде изучаемого распределения (называемого теоретическим) определяют его параметры $M_x = \bar{x}_B$ и $D_x = \sigma_B^2$, а также количество чисел из выборки объема n , которое должно оказаться в каждом интервале (теоретические частоты). Для этого по таблице значений функции Лапласа (интеграла вероятности) $\Phi(x)$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt. \quad (4.18)$$

находится вероятность попадания в i -й интервал:

$$p_i = \Phi\left(\frac{b_i - \bar{x}_B}{\sigma_B}\right) - \Phi\left(\frac{a_i - \bar{x}_B}{\sigma_B}\right), \quad (4.19)$$

где a_i, b_i - границы i -го интервала. Умножив полученные вероятности на объем выборки n , найдем теоретические частоты $n \cdot p_i$. С учетом сказанного в качестве критерия Пирсона используется случайная величина $K = \chi^2$. Наблюдаемое значение данного параметра определяет меру расхождения эмпирических и теоретических частот и вычисляется по формуле:

$$\chi_H^2 = \sum_{i=1}^s \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}. \quad (4.20)$$

Для выбранной величины значимости α по таблице χ^2 -распределения находят критическое значение $\chi_{кр}^2 = \chi^2(\alpha, k)$ при числе степеней свободы $k = s - r - 1$, где s - число интервалов эмпирического распределения (вариационного ряда); r - число параметров теоретического распределения, определяемых по опытным данным (для нормального закона распределения $r = 2$). Если фактически наблюдаемое значение χ_H^2 больше критического, то гипотеза отвергается, в противном случае гипотеза принимается. Следует заметить, что критерий Пирсона дает удовлетворительные результаты, если в каждом группировочном интервале достаточное число наблюдений n_i . Если в каком-нибудь интервале число наблюдений меньше 5, имеет смысл объединить соседние интервалы.

Лекция 5. Регрессионный, корреляционный и дисперсионный виды анализа

Линейные регрессионные модели

Рассмотрим применение методов статистической обработки при проведении регрессионного анализа.

Учебно-исследовательская работа студента

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 обозначают на схемах в виде многополюсника, показанного на рисунке 5.1. Значения на входах и выходах черного ящика можно наблюдать и измерять. Содержимое ящика неизвестно. 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.



Рисунок 5.1 – Обозначение черного ящика на схемах

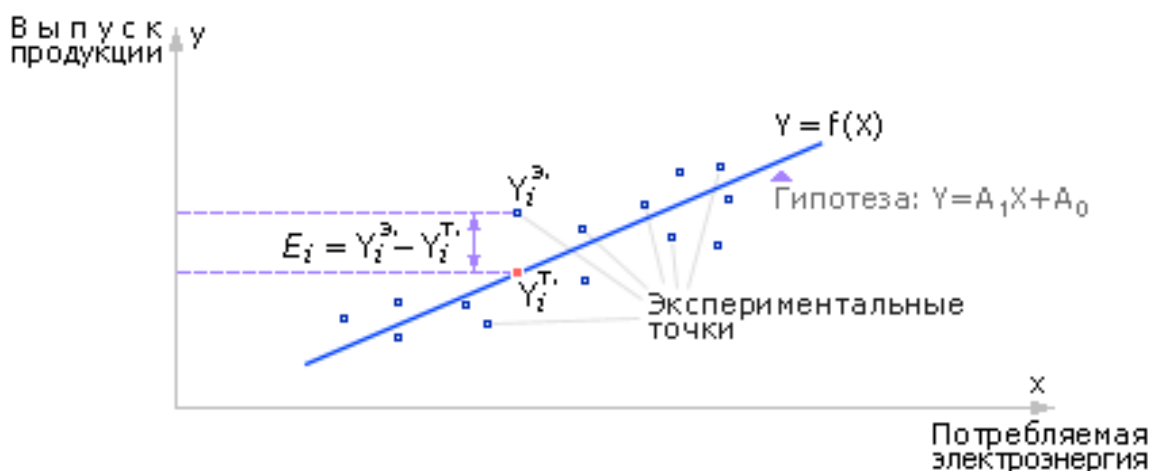


Рисунок 5.2 – Графический вид представления результатов наблюдений над черным ящиком

Решим задачу на примере черного ящика с одним входом и одним выходом. Допустим для простоты, что зависимость между входом и выходом линейная или

Учебно-исследовательская работа студента

почти линейная. Тогда данная модель будет называться **линейной одномерной регрессионной моделью**. Результаты n наблюдений показаны на графике 5.2.

Этапы построения моделей включают в себя:

1. **Выдвижение гипотезы о структуре ящика.** Рассматривая экспериментально полученные данные, предположим, что они подчиняются линейной гипотезе, т.е. выход Y зависит от входа X линейно, т.е. гипотеза имеет вид:

$$Y = A_1 X + A_0. \quad (5.1)$$

2. **Определение неизвестных коэффициентов A_0 и A_1 модели.** Для каждой из n снятых значений экспериментально вычислим ошибку E_i между экспериментальными значениями $Y_i^{эксн}$ и теоретическим значением $Y_i^{теор}$, лежащим на гипотетической прямой $A_1 X + A_0$:

$$E_i = (Y_i^{эксн} - Y_i^{теор}) = Y_i - A_0 - A_1 \cdot X_i, \quad i = 1, \dots, n. \quad (5.2)$$

Для всех n точек ошибки складываются, причем для избегания влияния знака ошибок складываются их суммарные значения:

$$F(A_0, A_1) = \sum_{i=1}^n E_i^2. \quad (5.3)$$

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

$$F(A_0, A_1) = \sum_{i=1}^n E_i^2 = \sum_{i=1}^n (Y_i - A_0 - A_1 \cdot X_i)^2 \Rightarrow \min_{A_0, A_1}. \quad (5.4)$$

Примерный график данной функции показан на рисунке 5.3.

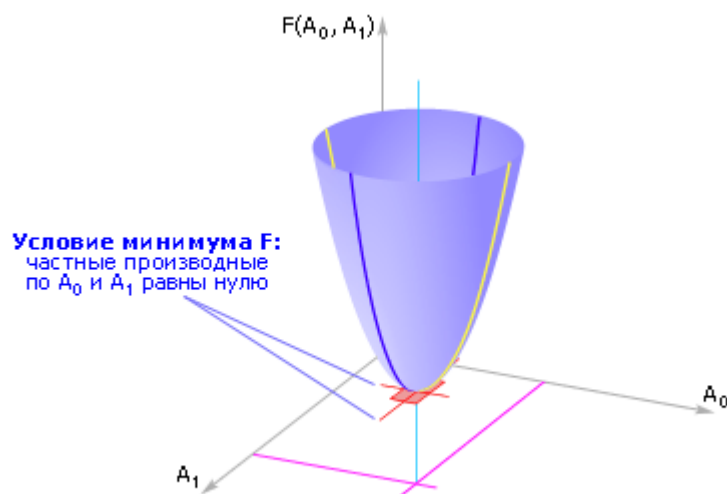


Рисунок 5.3 – Примерный график функции ошибки

Учебно-исследовательская работа студента

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

$$\frac{\partial F}{\partial A_0} = -2 \sum_{i=1}^n (Y_i - A_0 - A_1 X_i) = 0, \quad \frac{\partial F}{\partial A_1} = -2 \sum_{i=1}^n (Y_i - A_0 - A_1 X_i) X_i = 0. \quad (5.5)$$

Раскрывая скобки, получаем систему из двух линейных уравнений:

$$\sum_{i=1}^n A_0 + A_1 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i, \quad (5.6a)$$

$$A_0 \sum_{i=1}^n X_i + A_1 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i. \quad (5.6b)$$

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

$$\begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix} \begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix}. \quad (5.7)$$

Решение системы (5.7) имеет вид:

$$A_0 = \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i^2 - \sum_{i=1}^n X_i Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}, \quad (5.8a)$$

$$A_1 = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}. \quad (5.8b)$$

По формулам (5.8) вычисляются значения коэффициентов A_0 и A_1 .

3. *Проверка гипотезы.* Для проверки гипотезы необходимо, во-первых, рассчитать ошибку между точками заданной экспериментальной и полученной теоретической зависимости и суммарную ошибку, определяемые по формулам (5.2) и (5.3). Во-вторых, необходимо найти оценку среднеквадратического отклонения по формуле

$$\sigma = \sqrt{\frac{F}{n}}. \quad (5.9)$$

Если в полосу, ограниченную линиями $Y^{теор} - S$ и $Y^{теор} + S$, как показано на рисунке 5.4, попадает 68,28% и более экспериментальных точек $Y_i^{эксн}$, то выдвинутая гипотеза принимается. В противном случае выбирают более сложную гипотезу или проверяют исходные данные. Если потребуется большая

Учебно-исследовательская работа студента

уверенность в результате, то используют дополнительное условие: в полосу, ограниченную линиями $Y^{теор} - 2S$ и $Y^{теор} + 2S$, должны попасть 95,44% и более экспериментальных точек $Y_i^{эксп}$.

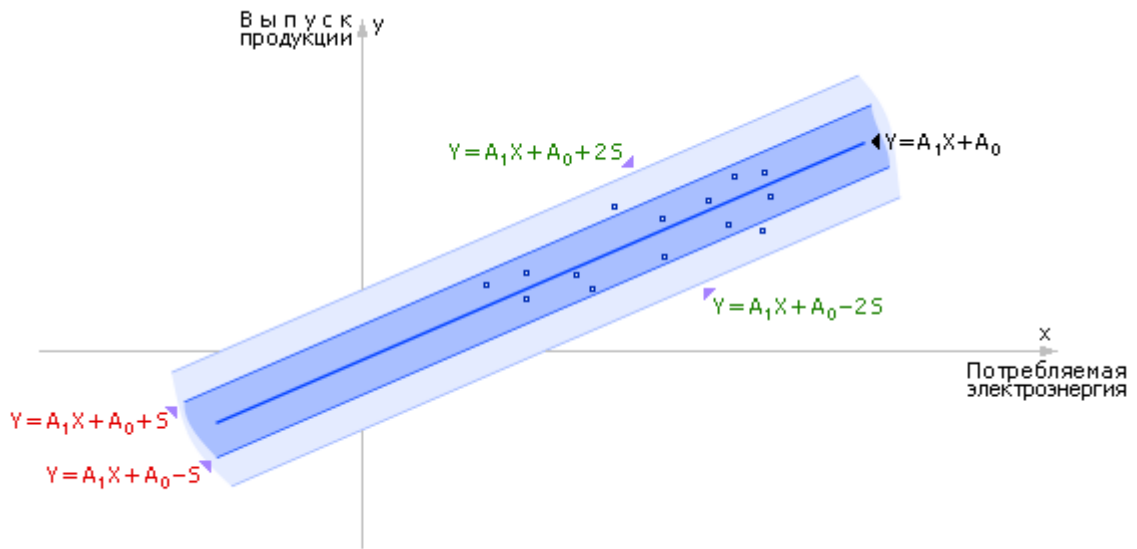


Рисунок 5.4 – Исследования допустимости принятия гипотезы

На графиках расстояние S связано со среднеквадратическим отклонением σ соотношением:

$$S = \frac{\sigma}{\sin \beta} = \frac{\sigma}{\sin(90^\circ - \arctg A_1)} = \frac{\sigma}{\cos(\arctg A_1)}. \quad (5.10)$$

Геометрия, на основе которой получено данное соотношение, приведена на рисунке 5.5. Условие принятия гипотезы выведено из нормального закона распределения случайных ошибок, что проиллюстрировано рисунком 5.6.

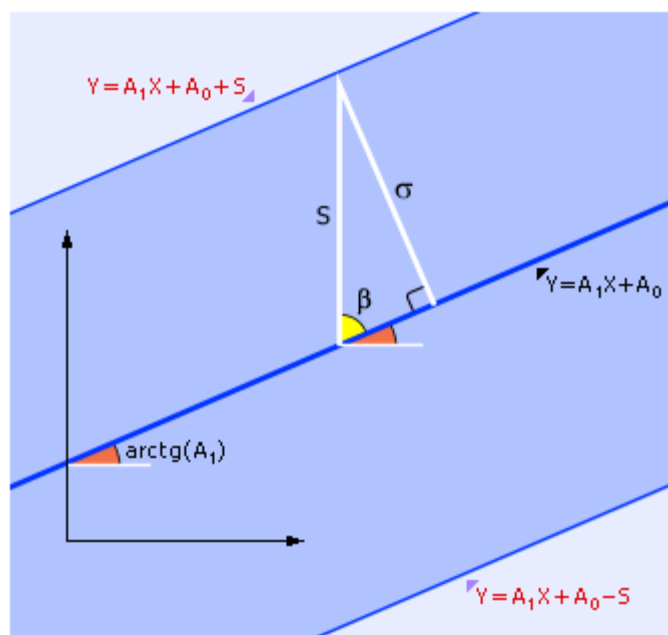


Рисунок 5.5 – Связь значений S и σ

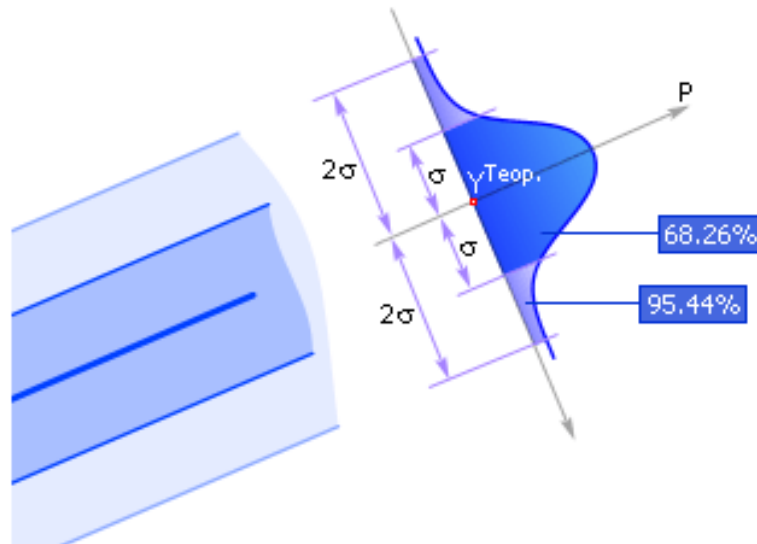


Рисунок 5.6 – Иллюстрация закона нормального распределения ошибок

Корреляционный анализ

Корреляционный анализ – метод, позволяющий обнаружить зависимость между несколькими случайными величинами.

Допустим, проводится независимое измерение различных параметров у одного типа объектов. Из этих данных можно получить качественно новую информацию – о взаимосвязи этих параметров. Например, между ростом и весом человека. Корреляция является частным случаем стохастической (вероятностной) связи.

Взаимосвязь между переменными определяется коэффициентом корреляции, который может быть положительным (при увеличении одного параметра второй тоже увеличивается), отрицательным (при увеличении одного параметра второй уменьшается) и нулевым (взаимосвязь параметров отсутствует).

Коэффициент корреляции вычисляется следующим образом. Пусть имеется массив из n точек $x_{1,i}, x_{2,i}$ ($i=1,2,\dots,n$). Для каждого из параметров рассчитывается среднее значение:

$$\bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1,i}, \quad \bar{x}_2 = \frac{1}{n} \sum_{i=1}^n x_{2,i}. \quad (5.11)$$

Затем вычисляется коэффициент корреляции по формуле:

$$r = \frac{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2} \sqrt{\sum_{i=1}^n (x_{2,i} - \bar{x}_2)^2}}. \quad (5.12)$$

Коэффициент корреляции изменяется в пределах от -1 до 1. В зависимости от величины коэффициента корреляции различают:

Учебно-исследовательская работа студента

- сильную связь $\pm 0.7 \leq r \leq \pm 1.0$;
- умеренную связь $\pm 0.5 \leq r \leq \pm 0.7$;
- слабую связь $\pm 0.3 \leq r \leq \pm 0.5$;
- связь практически отсутствует $r \leq \pm 0.3$.

Знак минус или плюс у коэффициента корреляции указывает на направление связи. Знак плюс означает, что связь между признаками x_1 и x_2 прямая (положительная), а знак минус – связь обратная (отрицательная).

Величина коэффициента корреляции служит оценкой соответствия уравнения регрессии выявленным связям. Корреляция и регрессия тесно связаны между собой. Корреляция оценивает силу связи, а регрессия исследует ее форму.

Коэффициент корреляции r является случайной величиной, поскольку вычисляется из случайных величин. Для него можно выдвигать и **проверять следующие гипотезы**.

1. Коэффициент корреляции значимо отличается от нуля (т.е. есть взаимосвязь между величинами).

Тестовая статистика вычисляется по формуле:

$$\xi = \left(0.5 \ln \left(\frac{1+r}{1-r} \right) - \frac{|r|}{2(n-1)} \right) \sqrt{n-3}. \quad (5.13)$$

Вычисленная величина сравнивается с табличным значением коэффициента Стьюдента $t(P = 0.95, n = \infty) = 1.96$.

Если тестовая статистика больше табличного значения, то коэффициент значимо отличается от нуля. Анализ формулы (5.13) показывает, что чем больше измерений n , тем вероятнее, что коэффициент значимо отличается от нуля.

2. Отличие между двумя коэффициентами корреляции значимо.

Тестовая статистика вычисляется по формуле:

$$\xi = 0.5 \ln \left(\frac{(1+r_1)(1-r_2)}{(1-r_1)(1+r_2)} \right) \frac{1}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}. \quad (5.14)$$

Результат вычислений также сравнивается со значением коэффициента Стьюдента.

Методами корреляционного анализа решаются следующие **задачи**:

Определение взаимосвязи между параметрами.

Осуществляется прогнозирование. Если известно поведение одного параметра, то можно предсказать поведение другого параметра, коррелирующего с первым.

Классификация и идентификация объектов. Корреляционный анализ помогает подобрать набор независимых признаков для классификации.

Корреляционный анализ является начальным этапом анализа, после которого возможны следующие переходы:

- в паттерн-анализ для выявления связанного («синхронно изменяющегося») набора факторов. Этот набор (паттерн) может быть затем подвергнут регрессионному анализу для установления количественных связей;

Учебно-исследовательская работа студента

- в кластерный анализ для разделения выборки на части с разным типом связи между факторами. Например, смешанная выборка с приближительной нулевой корреляцией может распасться на две подвыборки: с высокой положительной и высокой отрицательной корреляцией;

- в регрессионный анализ для построения регрессии выбранного фактора на все другие факторы, с которыми он достаточно сильно коррелирован, или только на какой-либо один выбранный фактор.

Дисперсионный анализ

Дисперсионный анализ включает в себя проверку гипотез, связанных с оценкой выборочной дисперсии.

Можно выделить три основных вида гипотез:

Значимо ли различие между двумя дисперсиями:

$$\sigma_1^2 = \sigma_2^2. \tag{5.15}$$

Превышает ли одна дисперсия другую:

$$\sigma_1^2 > \sigma_2^2. \tag{5.16}$$

Значимо ли различие между несколькими дисперсиями

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2. \tag{5.17}$$

Дисперсии вычисляются из случайных величин, поэтому сама также является случайной величиной. Напомним, что дисперсии, в отличие от средних, подчиняются распределению хи-квадрат.

Рассмотрим применение дисперсионного анализа на примере однофакторного анализа.

Пусть генеральные совокупности X_1, X_2, \dots, X_P распределены нормально и имеют одинаковую дисперсию, значение которой неизвестно. Найдем выборочные средние по выборками из этих генеральных совокупностей и проверим на заданном уровне значимости нулевую гипотезу H_0 : $M(X_1) = M(X_2) = \dots = M(X_P)$ о равенстве всех математических ожиданий.

При решении задачи будем считать, что на случайную величину X воздействует некоторый качественный фактор F , имеющий P уровней: F_1, F_2, \dots, F_P . Требуется сравнить «факторную дисперсию», т.е. рассеяние, порождаемое изменением уровня фактора, и «остаточную дисперсию», обусловленную случайными причинами. Если их различие значимо, то фактор существенно влияет на X , и при изменении его уровня групповые средние различаются значимо.

Будем считать, что количество наблюдений на каждом уровне фактора одинаково и равно q . Оформим результаты наблюдений в виде Таблицы 5.1.

Определим различные виды дисперсий.

Учебно-исследовательская работа студента

Т а б л и ц а 5.1 – Результаты экспериментальных данных

Номер испытания	Уровни фактора F_j			
	F_1	F_2	...	F_P
1	x_{11}	x_{12}	...	x_{1P}
2	x_{21}	x_{22}	...	x_{2P}
...
q	x_{q1}	x_{q2}	...	x_{qP}
Групповое среднее	\bar{x}_{zp1}	\bar{x}_{zp2}	...	\bar{x}_{zpP}

Общая сумма квадратов отклонений наблюдаемых значений от общего среднего \bar{x} определяется выражением:

$$S_{общ} = \sum_{j=1}^P \sum_{i=1}^q (x_{ij} - \bar{x})^2. \quad (5.18)$$

Факторная сумма отклонений групповых средних от общей средней, характеризующая рассеяние между группами, находится по формуле:

$$S_{факт} = q \sum_{j=1}^P (\bar{x}_{zpj} - \bar{x})^2, \quad (5.19)$$

Остаточная сумма квадратов отклонений наблюдаемых значений группы от своего среднего группового среднего, характеризующая рассеяние внутри групп, описывается выражением:

$$S_{ост} = \sum_{i=1}^q (x_{i1} - \bar{x}_{zp1})^2 + \sum_{i=1}^q (x_{i2} - \bar{x}_{zp2})^2 + \dots + \sum_{i=1}^q (x_{iP} - \bar{x}_{zpP})^2. \quad (5.20)$$

Для сокращения объема вычислений остаточную сумму можно вычислить не по формуле (5.20), а как

$$S_{ост} = S_{общ} - S_{факт}. \quad (5.21)$$

Вводя обозначения $R_j = \sum_{i=1}^q x_{ij}$, $Q_j = \sum_{i=1}^q x_{ij}^2$, получим формулы, более удобные для расчетов:

$$S_{общ} = \sum_{j=1}^P Q_j - \frac{1}{pq} \left(\sum_{j=1}^P R_j \right)^2, \quad S_{факт} = \frac{1}{q} \sum_{j=1}^P R_j^2 - \frac{1}{pq} \left(\sum_{j=1}^P R_j \right)^2. \quad (5.22)$$

Разделив суммы квадратов на соответствующее число степеней свободы, получим общую, факторную и остаточную дисперсии:

$$\sigma_{общ}^2 = \frac{S_{общ}}{pq - 1}, \quad \sigma_{факт}^2 = \frac{S_{факт}}{p - 1}, \quad \sigma_{ост}^2 = \frac{S_{ост}}{p(q - 1)}. \quad (5.23)$$

Если справедлива гипотеза H_0 , то все эти дисперсии являются несмещенными оценками генеральной дисперсии. Покажем, что проверка нулевой

Учебно-исследовательская работа студента

гипотезы сводится к сравнению факторной и остаточной дисперсии по критерию Фишера-Снедекора.

1. Пусть гипотеза H_0 правильна, Тогда факторная и остаточная дисперсии являются несмещенными оценками неизвестной генеральной дисперсии и, следовательно, различаются незначительно. В связи с этим результат оценки по критерию Фишера-Снедекора $F = \frac{\sigma_1^2}{\sigma_2^2}$ (σ_1^2 - большая дисперсия, σ_2^2 - меньшая

дисперсия) покажет, что нулевая гипотеза принимается. Таким образом, если верна гипотеза о равенстве математических ожиданий генеральных совокупностей, то верная гипотеза и о равенстве факторной и остаточной дисперсиях.

2. Если нулевая гипотеза неверна, то с возрастанием расхождения между математическими ожиданиями увеличивается и факторная дисперсия, а вместе с

ней и отношение $F_{набл} = \frac{\sigma_{факт}^2}{\sigma_{ост}^2}$. В результате $F_{набл} > F_{кр}$, и гипотеза о равенстве

дисперсий будет отвергнута. Следовательно, если гипотеза о равенстве математических ожиданий генеральных совокупностей ложна, то ложна и гипотеза о равенстве факторной и остаточной дисперсий.

Таким образом, метод дисперсионного анализа состоит в *проверке по критерию F нулевой гипотезы о равенстве факторной и остаточной дисперсий*. Если факторная дисперсия окажется меньше остаточной, то гипотеза о равенстве математических ожиданий генеральных совокупностей верна. При этом нет необходимости использовать критерий F .

Если число испытаний на разных уровнях различно (q_1 испытаний на уровне F_1 , q_2 - на уровне F_2 , ..., q_p - на уровне F_p), то

$$S_{общ} = (Q_1 + Q_2 + \dots + Q_p) - (R_1 + R_2 + \dots + R_p), \tag{5.24}$$

$$S_{факт} = \left(\frac{R_1^2}{q_1} + \frac{R_2^2}{q_2} + \dots + \frac{R_p^2}{q_p} \right) - \frac{(R_1 + R_2 + \dots + R_p)^2}{n}. \tag{5.25}$$

Остальные вычисления проводятся так же, как и в случае одинакового числа испытаний.

Лекция 6. Модель динамической системы в виде Фурье представления

Общие сведения о динамических системах

До настоящего времени мы рассматривали статические модели, т.е. случай, когда один эксперимент не зависит от другого (система не обладала памятью). Если же каждый раз значение на выходе при одном и том же значении на входе

Учебно-исследовательская работа студента

разное, т.е. 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, т.е. обладают памятью. В математической записи прошлое отражается присутствием производной. Чем большей памятью обладает система, тем больше состояний из прошлого влияют на настоящее, тем большая степень старшей производной используется в записи модели.

Пусть на входе и выходе черного ящика имеются зависимости параметров X и Y от времени t . Задача состоит в том, чтобы адекватно определить черный ящик.

Графики зависимостей $X(t)$ и $Y(t)$ могут быть самыми разными, например, такими, как показано на рисунке 6.1. Для моделирования системы переведем аналоговый сигнал в дискретный вид, как проиллюстрировано на рисунке 6.2, и строят таблицу отсчетов, имеющую вид, показанный в Таблице 6.1.

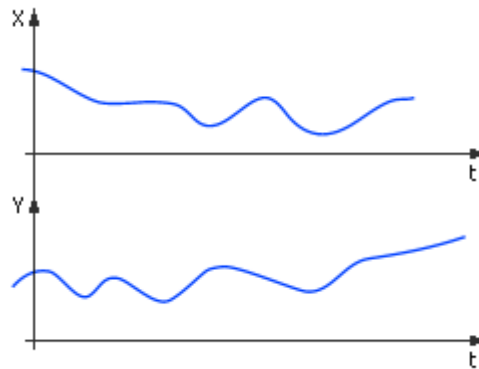


Рисунок 6.1 – Временные зависимости входной и выходного сигналов

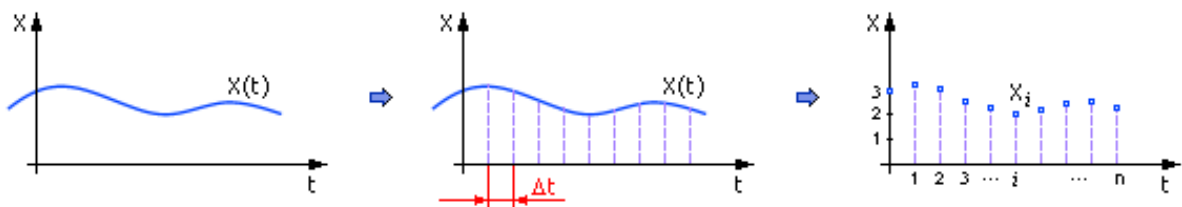


Рисунок 6.2 – Дискретизированный временной сигнал

Т а б л и ц а 6.1 – 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 6.1, упорядоченных во времени, часто называют **динамическим рядом**. Естественно, часть информации при такой операции теряется. Чем меньше расстояние между отсчетами, тем больше частота дискретизации, тем меньше потери информации. Частоту

Учебно-исследовательская работа студента

дискретизации принимают такой, чтобы не потерять высокочастотные составляющие в сигнале, отдельные пики.

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.

Для определения параметров динамической системы необходимо сначала оценить ее порядок: он совпадает со степенью свободы 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 рисунке 6.3.

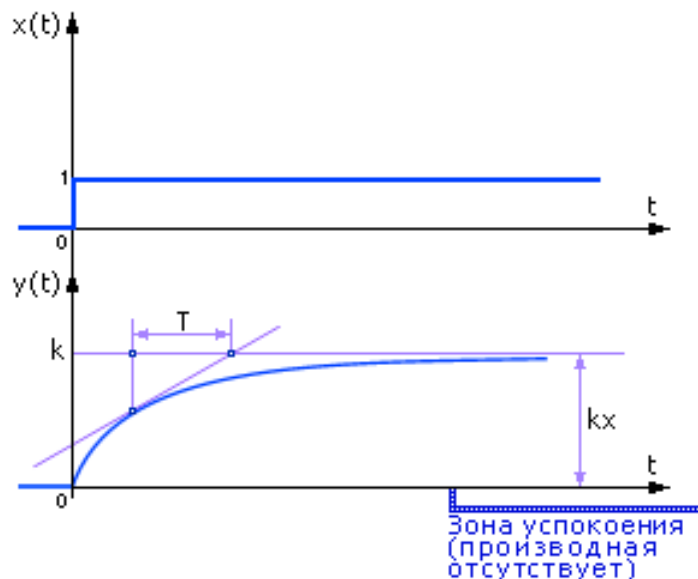


Рисунок 6.3 – Входной и выходной сигналы, типичные для системы первого порядка

Поясним смысл графика. 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 прийти до значения kx , то есть увеличить сигнал x в k раз (k — коэффициент усиления входного сигнала). Однако происходит это 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

Учебно-исследовательская работа студента

описывается производной, и выход оказывается меньше входа на некоторую величину f :

$$y = kx - f\left(\frac{dy}{dt}\right). \quad (6.1)$$

Когда система достигнет на выходе значения равного kx , то изменений не будет, значение производной станет равной нулю. $y = kx$. Данная функция определяет 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 (в решении модели будет присутствовать один интеграл):

$$T \frac{dy}{dt} + y = kx. \quad (6.2)$$

Данная система имеет два параметра: T и k .

Следует отметить, что один интеграл у линейных динамических систем всегда «порождает» одну экспоненту, двойной интеграл – сумму двух экспонент, и так далее. Чтобы определить, является ли кривая экспонентой, в каждой ее точке проводится касательная до пересечения с линией установившегося уровня (на рисунке 6.3 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 $3T$.

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

Помимо рассмотренного звена известны также и другие типовые звенья (усилительное звено, апериодическое, колебательное и т.д.).

Модель объекта

Одной из наиболее часто встречающихся динамических моделей в радиотехнике является ее представление в виде Фурье преобразования. Это обусловлено тем, что при приеме и передаче сигналов используются гармонические сигналы, которые могут быть легко представлены в виде

Учебно-исследовательская работа студента

разложения в ряд Фурье. При таком моделировании используется типовой прием – переход из описания сигнала во временной области к описанию в частотной области. Это позволяет перейти от дифференциальных уравнений к частотной характеристике объекта. Фактически это означает замену сигнала на частотную модель сигнала, а объекта – на частотную модель объекта. Получаемая при таком переходе схема моделирования показана на рисунке 6.4.

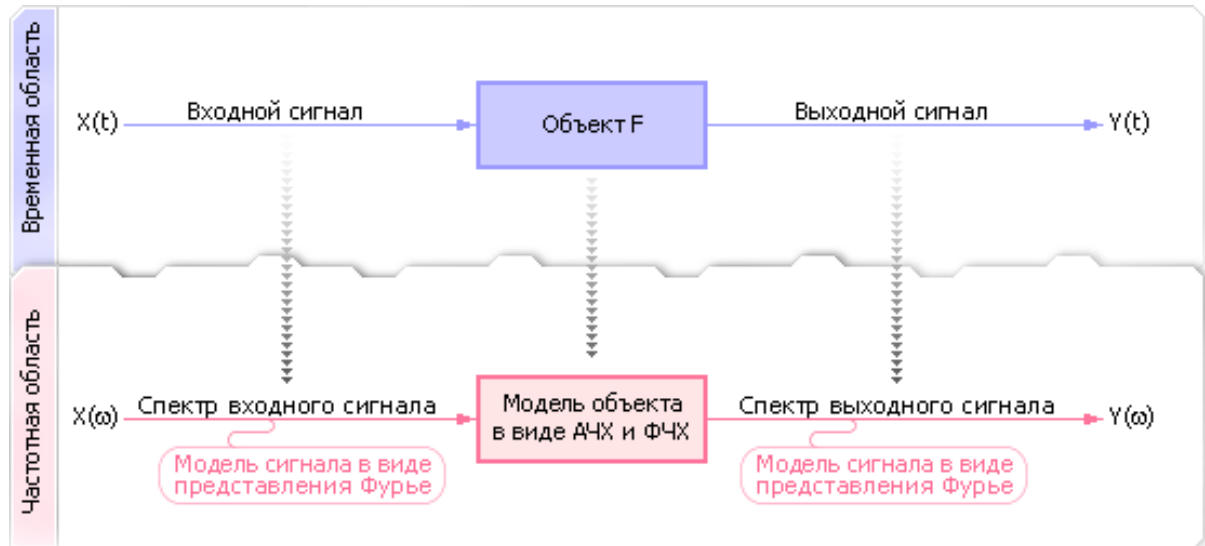


Рисунок 6.4 – Схема моделирования динамического объекта при переходе из временной области представления в частотную

Полученный при моделировании результат будет частотной моделью выходного сигнала, которую для нахождения окончательного результата придется преобразовать во временную область $Y(t)$. Процесс такого преобразования из частотной области во временную и обратно называется преобразованием Фурье.

Модель объекта в частотном виде называется передаточной функцией или АЧХ (амплитудно-частотной характеристикой). Объекты, для которых известны АЧХ, обычно называют типовыми звеньями. Амплитудно-частотная характеристика 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:

$$A_i^* = A_i(\omega_i)k_i(\omega_i). \tag{6.3}$$

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.

Таким образом, моделирование прохождения сигнала через динамический объект свелось к операции умножения двух переменных, точнее, к операции

Учебно-исследовательская работа студента

поэлементного умножения сектора одних переменных на вектор других переменных. Кроме того, достоинством метода является замена дифференциальных уравнений модели на алгебраические. Схема преобразования показана на рисунке 6.5.

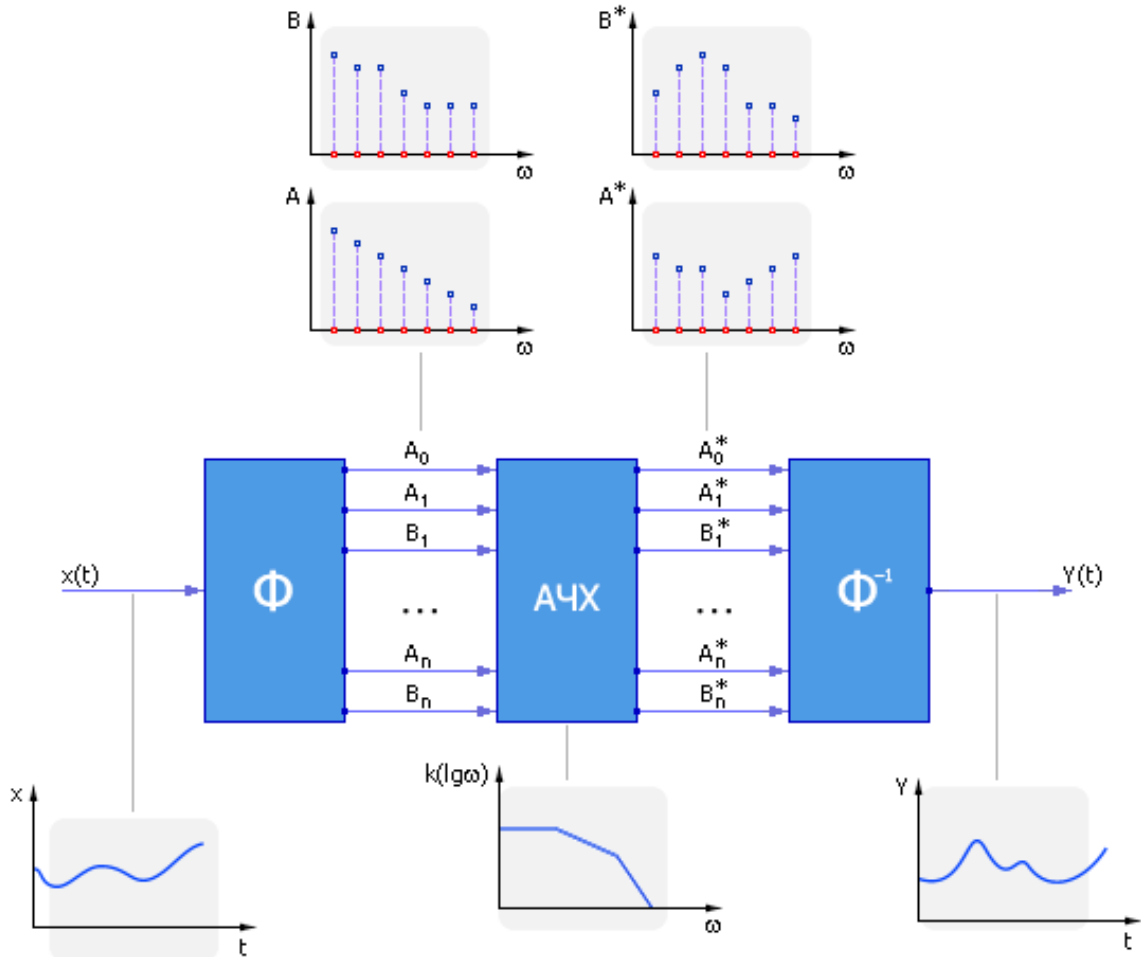


Рисунок 6.5 – Схема процедуры преобразования сигнала при использовании метода Фурье

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 радиоэлектронное устройство, состоящее из 5 блоков, показанных на рисунке 6.6. 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. Примером нелинейного блока может служить устройство ограничения сигнала (срез) по амплитуде.

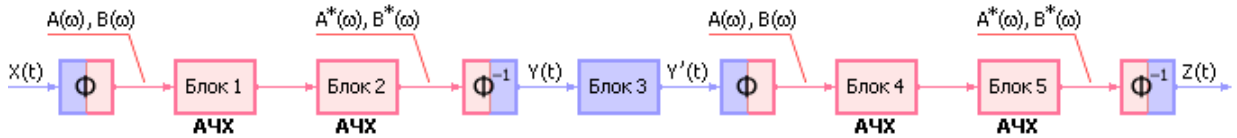


Рисунок 6.6 – Пример моделирования тракта, содержащего нелинейные блоки, с использованием метода Фурье

Анализ рисунка 6.6 показывает, что 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.

Рассмотренный метод является одним из самых быстродействующих. Это связано, как уже отмечалось, 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.

На эффективность метода сильно влияет частота дискретизации сигнала при разложении в ряд Фурье. При малой частоте дискретизации отсчеты в сигнале следуют редко, и часть сигнала остается потерянной. Это обусловлено тем, что между отсчетами может оказаться резко возросший и опавший пик, информация о котором пропадет. Другими словами, 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:

$$2W_{\max} \leq W_{\text{дискр}} \quad (6.4)$$

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

Модель сигнала

Рассмотрим применение метода Фурье для моделирования радиосигналов. Возможность его применения 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**.

Учебно-исследовательская работа студента

Пусть, например, в некотором сигнале присутствует сумма трех гармоник:

$$3\cos(t) + 2\cos(3t) + 0.5\cos(5t) \tag{6.5}$$

Это означает, что в сигнале присутствует первая гармоника с амплитудой 3, третья гармоника с амплитудой 2, пятая гармоника с амплитудой 0.5. Суммарный сигнал и его разложение в спектр показаны на рисунках 6.7, 6.8 соответственно.

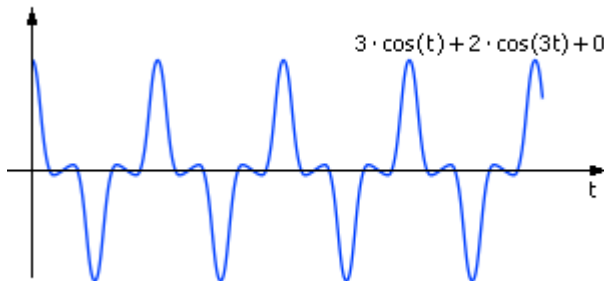


Рисунок 6.7 – Пример гармонического сигнала

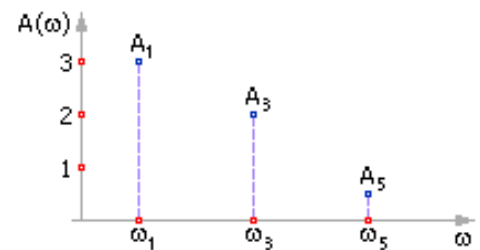


Рисунок 6.8 – Пример спектра гармонического сигнала

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 некоторый сигнал $X(t)$, имеющий вид, приведенный на рисунке 6.9. 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.

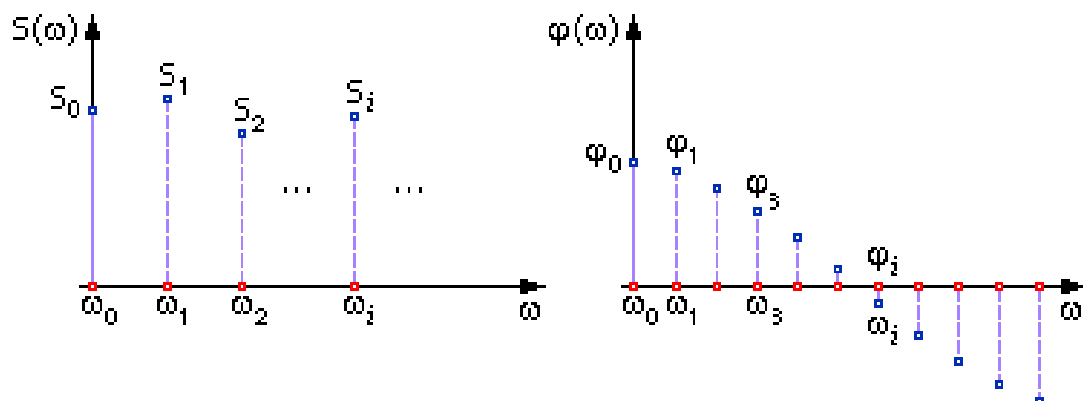


Рисунок 6.9 – Сигнал, представленный в частотной области, АЧХ и ФЧХ

В этом случае веса гармоник разложения Фурье будут описываться соотношениями, называемыми *прямым преобразованием Фурье*:

$$A_0 = \frac{2}{p} \int_0^p X(t) dt, \quad B_0 = 0, \quad (6.6)$$

$$A_1 = \frac{2}{p} \int_0^p X(t) \cos\left(\frac{2\pi t}{p}\right) dt, \quad B_1 = \frac{2}{p} \int_0^p X(t) \sin\left(\frac{2\pi t}{p}\right) dt, \quad (6.7)$$

$$A_2 = \frac{2}{p} \int_0^p X(t) \cos\left(\frac{2\pi 2t}{p}\right) dt, \quad B_2 = \frac{2}{p} \int_0^p X(t) \sin\left(\frac{2\pi 2t}{p}\right) dt, \quad (6.8)$$

$$A_i = \frac{2}{p} \int_0^p X(t) \cos\left(\frac{2\pi i t}{p}\right) dt, \quad B_i = \frac{2}{p} \int_0^p X(t) \sin\left(\frac{2\pi i t}{p}\right) dt, \quad (6.9)$$

где A_i, B_i - веса i -й гармоники в сигнале (его спектральные составляющие); отношение $i/p = \omega_i$ описывает частоту i -й гармоники.

Исходный сигнал восстанавливается с помощью формулы **обратного преобразования Фурье**:

$$x(t) = \frac{A_0}{2} + \sum_{i=1}^M \left(A_i \cos\left(\frac{2\pi i t}{p}\right) + B_i \sin\left(\frac{2\pi i t}{p}\right) \right), \quad (6.10)$$

где M - число учитываемых в разложении гармоник Фурье.

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.

Система чисел A_i и B_i является **полной характеристикой сигнала**. Такой же полной характеристикой сигнала является система чисел из амплитудно-частотной характеристики S и фазо-частотной характеристики (ФЧХ) φ , показанные на рисунке 6.9 и определяемые соотношениями:

$$S_i = \sqrt{A_i^2 + B_i^2}, \quad \varphi_i = \arctg\left(\frac{B_i}{A_i}\right). \quad (6.11)$$

Системы « A и B » и « S и φ » являются полностью равнозначными. При сложении гармоник в системе АЧХ, ФЧХ необходимо учитывать сдвиг фаз. Кроме того, обратное преобразование Фурье в данной системе будет определяться формулой:

$$x(t) = \sum_{i=1}^M S_i \sin(\omega_i t + \varphi_i). \quad (6.12)$$

Помимо рассмотренных на данной лекции примеров динамических систем известны и другие динамические модели. К ним, в частности, можно отнести модель в виде фильтра Каллмана.

Лекция 7. Статистическое моделирование

Общие сведения о статистическом моделировании

При исследовании сложных систем, подверженных случайным возмущениям, используются статистические методы. **Основным отличием** данных методов от рассмотренных в предыдущих лекциях является построение генеральной совокупности: последовательность вариантов исходных данных, поступающих на вход системы, определяется не самим исследователем в зависимости от плана эксперимента, а генерируются с помощью датчика случайных чисел на компьютере. Далее реакция проверяется не на реальном объекте исследований, а на модели. Таким образом, основное место при использовании статистических методов занимает компьютер.

В качестве моделей, на которых проверяется возможная реакция системы, применяются вероятностно аналитические и вероятностно имитационные модели. В *вероятностных аналитических моделях* влияние случайных факторов учитывается с помощью задания вероятностных характеристик случайных процессов (законы распределения вероятностей, спектральные плотности или корреляционные функции). Это приводит к усложнению вычислительной задачи и ограничивает применение данных моделей сравнительно простыми системами.

В *имитационных моделях* введение случайных возмущений не вносит принципиальных усложнений, что делает их наиболее часто применяемыми. В связи с этим исследование сложных процессов и систем, подверженных случайным возмущениям, с помощью имитационного моделирования принято называть **статистическим моделированием**.

Статистическая модель случайного процесса - это алгоритм, с помощью которого имитируют работу сложной системы, подверженной случайным возмущениям, причем полагается, что взаимодействие элементов системы носит вероятностный характер.

Оценка параметров модели осуществляется с помощью статистических методов: метода максимального правдоподобия, метода наименьших квадратов, метода моментов.

Таким образом, **методика статистического моделирования** состоит из следующих этапов:

1. Моделирование на компьютере псевдослучайных последовательностей с заданной корреляцией и законом распределения вероятностей (метод Монте-Карло), имитирующих случайные значения параметров при каждом испытании.

2. Преобразование полученных числовых последовательностей на имитационных математических моделях в генеральную совокупность.

3. Статистическая обработка результатов моделирования.

Обобщенный алгоритм метода статистических испытаний представлен на рисунке 7.1.

Различают две области применения метода статистического моделирования:

- для изучения стохастических систем;
- для решения детерминированных задач.

Учебно-исследовательская работа студента

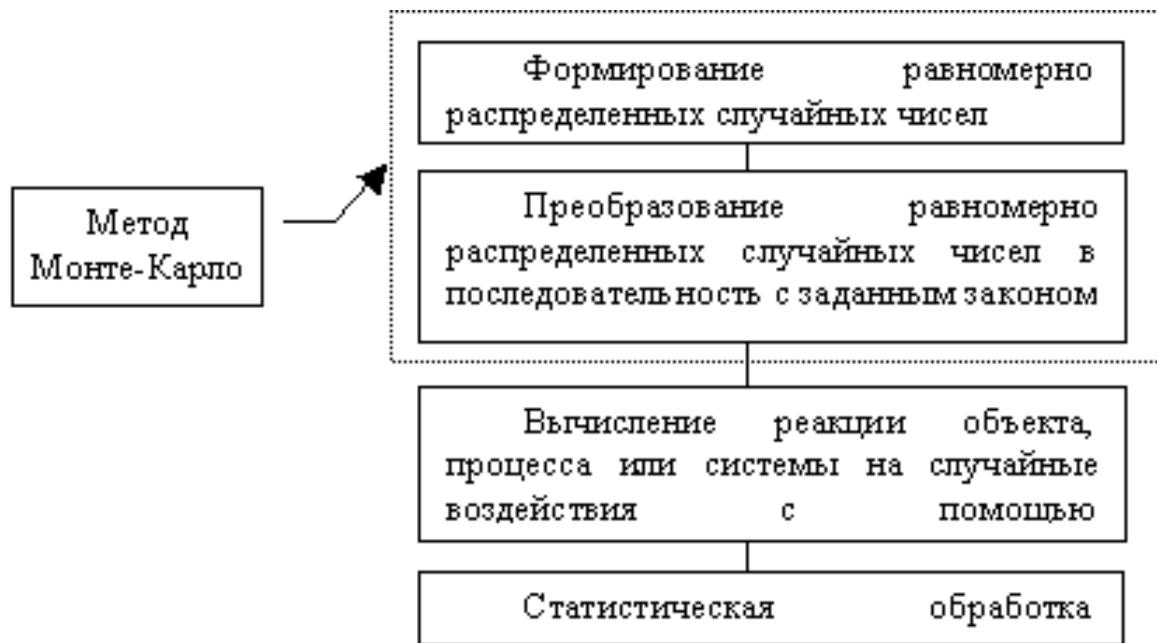


Рисунок 7.1 – Обобщенный алгоритм метода статистических испытаний

Напомним, что в детерминированных системах предсказываемые значения могут быть вычислены точно, а в стохастических – лишь с некоторой долей вероятности.

Основной идеей, которая используется для решения детерминированных задач методом статистического моделирования, является замена детерминированной задачи эквивалентной схемой некоторой стохастической системы, выходные характеристики последней совпадают с результатом решения детерминированной задачи. При такой замене погрешность уменьшается с увеличением числа испытаний N при реализации моделирующего алгоритма.

Достоинство: статистическая обработка получаемой серии частных значений искомых величин или функций позволяет получить сведения о поведении реального объекта или процесса в произвольные моменты времени. При этом если количество реализаций достаточно велико, то результаты моделирования приобретают статистическую устойчивость и с достаточной точностью могут быть приняты в качестве оценок искомых характеристик процесса функционирования системы.

Основной сложностью при использовании статистического моделирования является учет стохастических воздействий. Здесь можно выделить следующие моменты:

- точность получаемых оценок зависит от размера совокупности случайных чисел, генерируемых системой, что приводит к *росту вычислительных затрат*, обусловленных созданием данной совокупности;

- качество получаемых на основе статистических моделей результатов, их точность и достоверность определяются исходными (базовыми) последовательностями случайных чисел. Это приводит к *необходимости разработки простых и экономичных способов формирования последовательностей случайных чисел* требуемого качества.

Методы генерирования случайной величины

Для получения случайных числовых последовательностей с заданными вероятностными характеристиками могут применяться несколько методов, различающихся видом распределения случайной величины на заданном интервале (a, b) . В общем случае случайная величина может быть распределена на интервале в соответствии с различными законами распределения:

- равномерным, при котором на интервале (a, b) функция плотности вероятности и распределение соответственно определяются соотношениями:

$$f(x) = \begin{cases} 1/(b-a), & a \leq x \leq b \\ 0, & x < a, x > b \end{cases} \quad F(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x \leq b \\ 1 & x > b \end{cases} \quad (7.1)$$

- нормальным;

- распределением Бернулли, для которого случайная величина принимает значение 1 с вероятностью p и 0 с вероятностью $1-p$;

- биномиальным, при котором случайное число подчиняется распределению

$$P_n(m) = C_n^m p^m (1-p)^{n-m}, \quad (7.2)$$

где n - общее число испытаний; m - число успешных опытов;

- Пуассона, при котором вероятность реализации случайной величины со значением m и параметром распределения λ равна

$$P(m) = \frac{\lambda^m}{m!} \exp(-\lambda). \quad (7.3)$$

Численный метод, моделирующий случайные величины, равномерно распределенные на интервале $(0, 1)$, получил название "**метод статистических испытаний**" или "**метод Монте-Карло**".

Задачу **моделирования случайных чисел** с нормальным законом распределения решают в несколько **этапов**:

1. Вначале имитируют равномерное распределение и получают последовательность псевдослучайных чисел, равномерно распределенных на интервале $(0, 1)$.

2. Затем, используя равномерно распределенную псевдослучайную величину, получают последовательность псевдослучайных чисел с нормальным законом распределения (чаще всего в нормированном виде, т.е. $M_x = 0, \sigma = 1$).

Различают три **основных способа формирования последовательности** нормально распределенных случайных величин:

1. **Прямое преобразование** псевдослучайного числа u являющегося реализацией случайной величины Y , равномерно распределенной на интервале $[0, 1]$, с помощью некоторой функции W в число x , которое может рассматриваться как реализация случайной величины X , имеющей нормальный закон распределения.

Учебно-исследовательская работа студента

2. *Отсевание* псевдослучайных чисел из первоначальной последовательности Y равномерно распределенной на интервале $[0,1]$, таким образом, чтобы оставшиеся числа были распределены по нормальному закону.

3. *Моделирование условий*, соответствующих центральной предельной теореме теории вероятности.

Для моделирования нормально распределенной случайной величины применяют различные *методы*.

Метод полярных координат относится к первому способу получения последовательности псевдослучайных чисел с нормальным законом распределения. Метод вычисляет две независимые нормально распределенные случайные величины X_1 и X_2 с $M_x = 0$, $\sigma = 1$ по двум заданным независимым равномерно распределенным случайным числам Y_1 и Y_2 .

Метод, основанный на центральной предельной теореме, относится к третьему способу получения последовательности чисел с нормальным законом распределения. Метод основан на приближенном воспроизводстве условий, при которых справедлива центральная предельная теорема теории вероятности. Согласно данной теореме при сложении достаточно большого независимых случайных величин с произвольным законом распределения получается случайная величина, распределенная по нормальному закону. Опыт показывает, что при сложении всего шести ($k=6$) случайных величин равномерно распределенных на интервале $[0,1]$, получается случайная величина, которая с точностью, достаточной для большинства прикладных задач, может считаться нормальной.

Марковские цепи

К статистическим методам относится также применяемый в статистической радиофизике метод, основанный на применении марковских процессов.

Под **марковским процессом** понимается случайный процесс, эволюция которого после любого заданного значения временного параметра t не зависит от эволюции, предшествовавшей t , при условии, что значение процесса в этот момент фиксировано. Это означает, что «будущее» процесса не зависит от «прошлого» при известном «настоящем».

Определяющее марковский процесс свойство принято называть марковским. Впервые оно было сформулировано А. А. Марковым, который в работах 1907 г. положил начало изучению последовательностей зависимых испытаний и связанных с ними сумм случайных величин. Это направление исследований известно под названием **теории цепей Маркова или «динамики вероятностей»**. Основы общей теории марковских процессов с непрерывным временем были заложены Колмогоровым.

Марковские цепи по существу аналогичны методу динамического программирования, отличием является лишь то, что на каждом шаге учитывается вероятность попадания системы в то или иное состояние. В связи с этим этот метод называют **стохастическим динамическим программированием**.

Учебно-исследовательская работа студента

Благодаря сравнительной простоте и наглядности математического аппарата, высокой достоверности и точности получаемых решений особое внимание марковские процессы приобрели у специалистов, занимающихся также исследованием операций и теорией принятия оптимальных решений.

Марковские случайные процессы относятся к частным случаям случайных процессов и основаны на понятии случайной функции.

Случайной функцией называется функция, значение которой при любом значении аргумента является случайной величиной. Таким образом, случайной можно назвать функцию, которая при каждом испытании принимает какой-либо заранее неизвестный вид.

Примерами случайных функция являются колебания напряжения в электрической цепи, скорость движения автомобиля на участке дороги с ограничением скорости, шероховатость поверхности детали на определенном участке и т. д.

Считают, что если аргументом случайной функции является время или какой-то другой аргумент, то такой **процесс** называют **случайным**. При этом случайные процессы могут быть с дискретным или непрерывным состоянием или временем. Например, любой выборочный контроль продукции будет относиться к случайным процессам с дискретными состояниями, а случай отказа любой машины можно отнести к случайным процессам с дискретными состояниями, но непрерывным временем. Любая осциллограмма будет записью случайного процесса с непрерывными состояниями и временем.

Важным свойством случайных процессов является вероятностная связь между состояниями случайного процесса. Так, например, если в случайном процессе вероятность перехода системы в каждое последующее состояние зависит только от предыдущего состояния, то такой процесс называется **процессом без последствия**. Такие процессы также рассматривались А.А. Марковым, который предложил называть их в отличие от простой цепи без взаимосвязи сложной цепью. В настоящее время теория таких цепей разработана слабо и обычно применяют так называемый процесс укрупнения состояний, путем математических преобразований, объединяя предшествующие состояния в одно. Это обстоятельство должно обязательно учитываться при составлении математических моделей принятия решений.

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. Если вероятность перехода не зависит от времени, то **марковскую цепь** называют **однородной**.

Представим модель марковского процесса в виде графа, в котором состояния (вершины) связаны между собой связями (переходами из i -го

Учебно-исследовательская работа студента

состояния в j -е), как показано на рисунке 7.2,а. 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. Переходы происходят случайно, однако если измерить частоту переходов за достаточно большое время, то окажется, что эта частота будет совпадать с заданной вероятностью перехода. У каждого состояния сумма вероятностей всех переходов (исходящих стрелок) из него в другие состояния должна быть всегда равна 1.

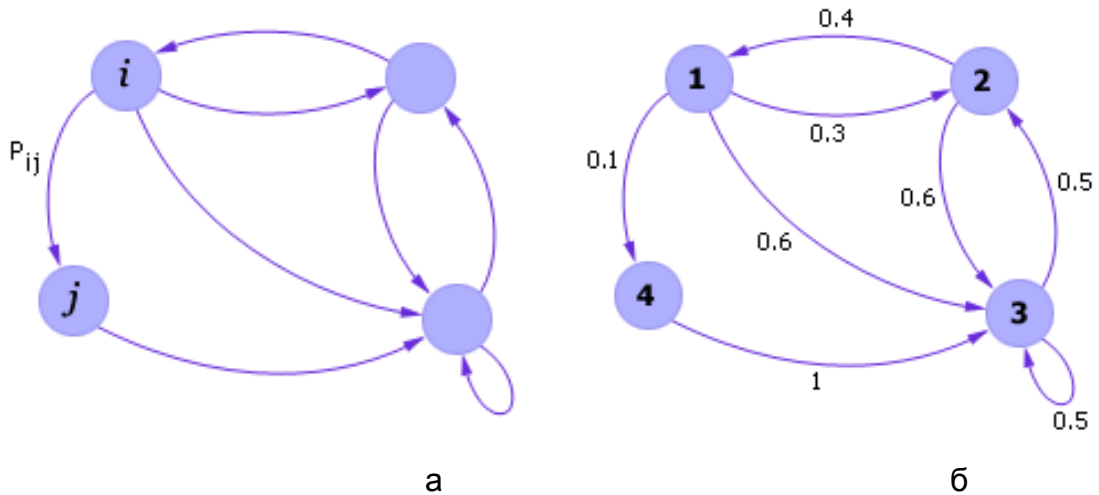


Рисунок 7.2 – Пример графа переходов

Реализация марковского процесса (процесс его моделирования) представляет собой вычисление последовательности (цепи) переходов из состояния в состояние, как показано на рисунке 7.3. Причем цепь является случайной последовательностью и может иметь также и другие варианты реализации.

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. рисунок 7.4. Далее с помощью генератора случайных чисел (ГСЧ) надо получить очередное равномерно распределенное в интервале $[0; 1]$ случайное число r_{pp} и определить, в какой из интервалов оно попадает. После этого осуществляется переход в состояние, определенное ГСЧ, и повтор описанной процедуры для нового состояния. Результатом работы модели является марковская цепь, показанная на рисунке 7.3.



Рисунок 7.3 – Пример марковской цепи

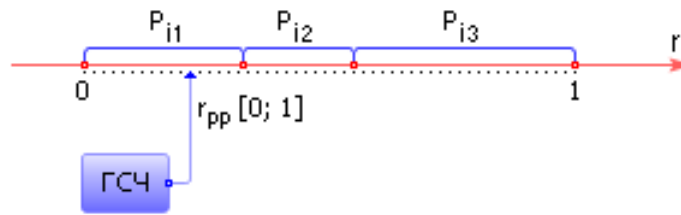


Рисунок 7.4 – Процесс моделирования перехода с использованием генератора случайных чисел

Рассмотрим численный пример, в котором имитируется стрельба из пушки по цели. Определим следующие три состояния: S_0 — цель не повреждена; S_1 — цель повреждена; S_2 — цель разрушена. Вектор начальных вероятностей задан в Таблице 7.1. В ней значение P_0 для каждого из состояний показывает, какова вероятность каждого из состояний объекта до начала стрельбы.

Таблица 7.1 – Вектор начальных вероятностей

	S_0	S_1	S_2
P_0	0.8	0.2	0

Таблица 7.2 – 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 перехода состояний, определяющая вероятность перехода из каждого состояния в каждое, приведена в Таблице 7.2. Наглядно модель марковского процесса можно представить себе в виде следующего графа, представленного на рисунке 7.5.

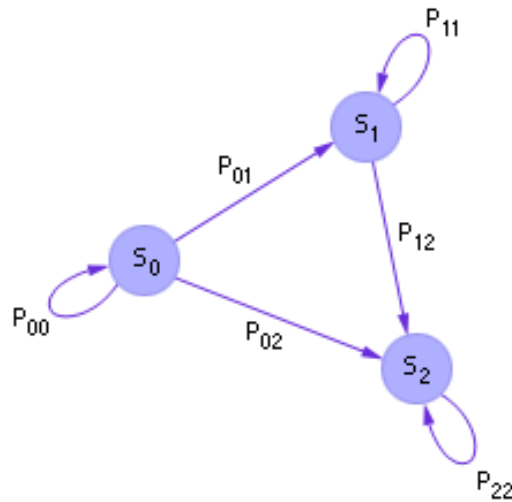


Рисунок 7.5 – Графа марковского процесса, моделирующего стрельбу из пушки по цели

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,

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 рисунке 7.6 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. Процедура уничтожения цели совершена за 5 тактов, то есть марковская цепь этой реализации выглядит следующим образом: $S_0—S_0—S_1—S_1—S_1—S_2$. Конечно, ответом задачи данное число быть не может, так как в разных реализациях получатся разные ответы. Ответ же у задачи может быть только один.

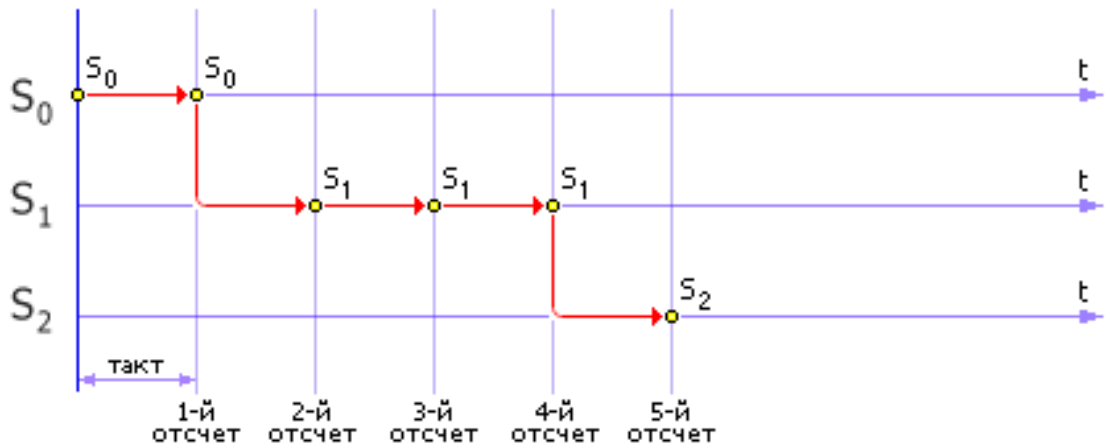


Рисунок 7.6 – Временная диаграмма переходов в марковском процессе (пример имитации)

Повторяя данную имитацию, можно получить, например, еще такие реализации (это зависит от того, какие конкретно случайные числа выпадут): 4 ($S_0—S_0—S_1—S_1—S_2$); 11 ($S_0—S_0—S_0—S_0—S_0—S_1—S_1—S_1—S_1—S_1—S_1—S_2$); 5 ($S_1—S_1—S_1—S_1—S_1—S_2$); 6 ($S_0—S_0—S_1—S_1—S_1—S_1—S_2$); 4 ($S_1—S_1—S_1—S_1—S_2$); 6 ($S_0—S_0—S_1—S_1—S_1—S_1—S_2$); 5 ($S_0—S_0—S_1—S_1—S_1—S_2$). 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.

Определим точность полученного результата. Для этого проследим, как сходится последовательность случайных (приближенных) ответов к правильному (точному) результату. 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, как показано на рисунке 7.7, то обнаружится, что данный ряд сходится к некоторой величине, которая и является ответом. Визуально мы можем наблюдать, что график «успокаивается», разброс между вычисляемой текущей величиной и ее теоретическим значением со временем уменьшается, стремясь к статистически точному результату. Таким образом, в некоторый момент график входит в некоторую «трубку», размер которой и определяет точность ответа.

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.

Учебно-исследовательская работа студента

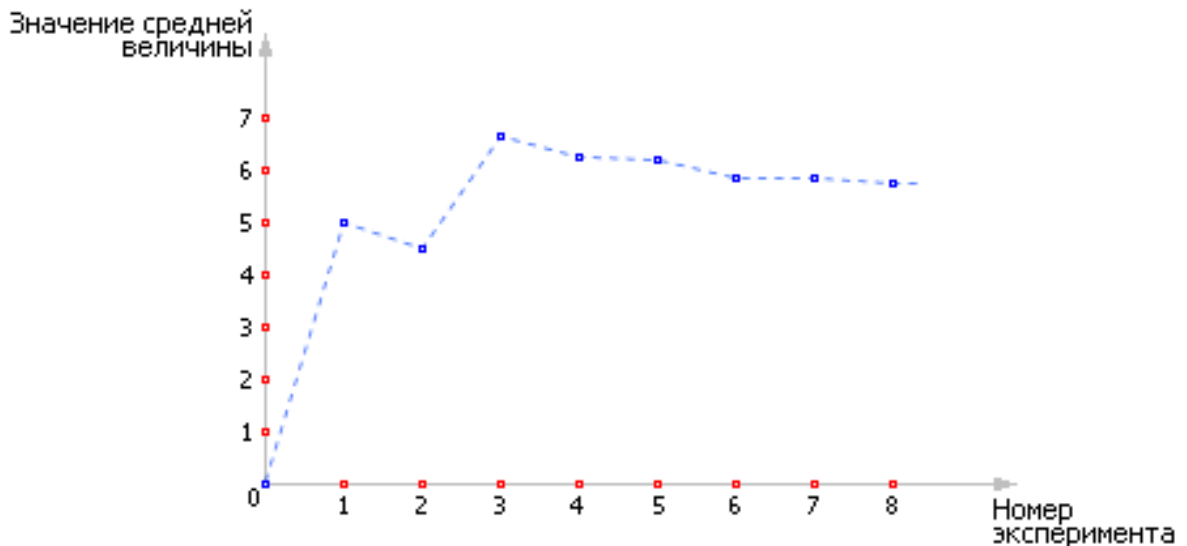


Рисунок 7.7 – Изменение средней величины в зависимости от номера эксперимента

Лекция 8. Процесс экспертизы

Методы получения оценок экспертизы. Голосование

При сравнении нескольких научных проектов (моделей) для оценки их качества может проводиться их экспертиза, т.е. целесообразный отбор одного решения из множества возможных.

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.

Существует много методов проведения экспертизы: голосование, ранжирование, попарное сравнение, разбиение на классы. Рассмотрим некоторые из них.

Голосование можно применять только в том случае, если мнений, которые предложены к рассмотрению, меньше, чем собравшихся.

Способы голосования:

1. Выбор большинством голосов. Достоинство – простота, недостаток - решение принимается неквалифицированным большинством.

2. Голосованием большинством 2/3. Применяется в особо ответственных случаях, повышая порог принятия решения.

3. 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).

Учебно-исследовательская работа студента

4. 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.

5. Туровое голосование. В первом туре голосования отбирается как результат 2-3 проекта, которые набирают большинство голосов и проходят установленный порог. Во втором туре все эксперты голосуют снова, но только уже за эти 2-3 отобранных проекта.

6. Метод Дельфы. Экспертам дается возможность знакомиться с результатами голосования после каждого тура и не убирать проекты из матрицы от тура к туру. Это позволяет экспертам менять собственные предпочтения, исключая случайность и приходя к равновесному мнению.

Методы ранжирования

Вторым классом методов экспертизы является ранжирование. **Ранжирование** есть упорядочение, разбиение множества на элементы с введением между ними некоторого порядка, в данном случае порядка «лучше-хуже».

Различают несколько **методов ранжирования**.

1. **Метод «Медиана Кемени»**. Вычисляются значения r_{ij} - расстояние между мнениями i -го и j -го экспертов. Вычислять расстояние (норму) между двумя векторами, описывающими мнение эксперта по кортежу проектов, можно различными способами. В математике известны пифагорова норма, манхэттенская, Минковского и т. д. Полученные нормы заносятся в таблицу (матрицу расстояний), имеющую вид, показанный в Таблице 8.1. Затем вычисляются значения R_i - расстояние от i -го эксперта до всех остальных - как сумма расстояний между мнением этого эксперта и остальных экспертов. Среди вычисленных суммарных расстояний ищется минимум из R_i . Мнение соответствующего i -го эксперта (с минимальным R) является самым средним и объявляется результатом экспертизы.

Т а б л и ц а 8.1 – 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	$r_{41} + r_{42} + r_{43} + 0 = R_4$
					$\min(R_1, R_2, R_3, R_4)$

Учебно-исследовательская работа студента

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. Самая удаленная от центра вершина характеризует самое ненадежное мнение и может этим характеризовать самого эксперта.

В качестве примера рассмотрим экспертизу четырех проектов (A, B, C и D), выполненную четырьмя экспертами. Распределение мест проектов приведено в Таблице 8.2.

Т а б л и ц а 8.2 – Расчет расстояний между мнениями экспертов

	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

Для рассматриваемого примера матрица расстояний будет иметь вид, показанный в Таблице 8.3. Анализ данной таблицы показывает, что наименьшее расстояние от остальных мнений — у эксперта 1, поэтому его мнение объявляется результатом. Ответ: проект A — 1 место, проект B — 2, проект C — 3, проект D — 4. Наилучшим проектом является проект A.

Т а б л и ц а 8.3 – Расчет 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

2. 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. Таблицу 8.2) проект A займет в итоге первое место, так как трое экспертов отдали ему первое место. Проекты B и C 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.

3. Альтернатива Кондорсе. В соответствии с данным методом сначала 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. Результат соотношений заносят в Таблицу 8.4.

Учебно-исследовательская работа студента

Т а б л и ц а 8.4 – Расчет отношений «лучше-хуже» между проектами

S_{ij}	Отношение	S_{ji}	Пояснение
$S_{AB} = 3$	>	$S_{BA} = 1$	A лучше B
$S_{AC} = 4$	>	$S_{CA} = 0$	A лучше C
$S_{AD} = 4$	>	$S_{DA} = 0$	A лучше D
$S_{BC} = 3$	>	$S_{CB} = 1$	B лучше C
$S_{BD} = 3$	>	$S_{DB} = 1$	B лучше D
$S_{CD} = 3$	>	$S_{DC} = 1$	C лучше D

Далее строится граф, на котором стрелками показываются отношения между параметрами проектов, как показано на рисунке 8.1,а. 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 A направлена к B, так как проект A лучше проекта B (трое экспертов предпочли проекту B проект A, и лишь только один эксперт - наоборот).

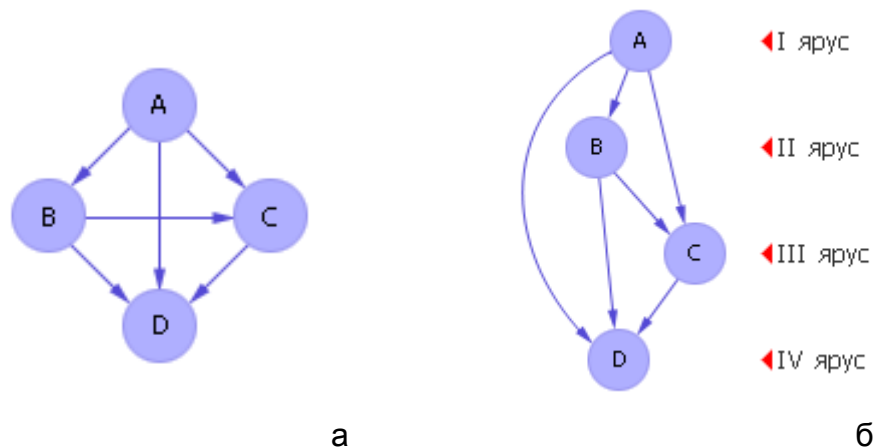


Рисунок 8.1 – Граф предпочтений проектов

После упорядочения графа к ярусно-параллельной форме он примет вид, показанный на рисунке 8.2,б. При этом в верхних ярусах находятся вершины, в которые нет входящих стрелок из вершин, лежащих в нижних ярусах. Анализ графа показывает, что вершиной, в которую нет стрелок, является вершина A, следовательно, абсолютным лидером среди проектов является проект A. Проект A лучше B, одновременно он лучше C, одновременно он лучше D. Далее, среди оставшихся проектов (BCD) проект B лучший, так как он одновременно лучше C и D. И, наконец, C лучше D.

4. **Принцип Борда.** В соответствии с ним за первое место каждому проекту дается 1 балл, за второе место — 2 балла, за третье — 3 балла и так далее. Далее подсчитывается количество баллов, которое получит каждый проект в сумме. Первое место в итоге получит проект, который наберет наименьшее число баллов, остальные места определяются сортировкой набранных баллов.

Учебно-исследовательская работа студента

В нашем примере получим распределение мест, показанное в Таблице 8.5. Анализ таблицы показывает, что и как в случае применения рассмотренных ранее методов ранжирования первое место занимает проект А.

Т а б л и ц а 8.5 – Распределение мест проектов методом Борда

	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

5. **Метод «Электра».** Рассмотрим данный метод на примере решения о покупке автомобиля. Сначала составим таблицу критериев, по которым будут оценивать проекты (Таблицу 8.6).

Т а б л и ц а 8.6 – 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 Таблице 8.7.

Рассматриваем все пары проектов i и j . Если по какому-либо критерию i -й проект лучше, чем j -й, 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 -й проект оказывается лучше, чем i -й, 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

Учебно-исследовательская работа студента

$D \leq 1$ отбрасываются. Заметим, что $D_{ji} = 1/D_{ij}$ (и наоборот), таким образом, вычисления можно несколько упростить.

Т а б л и ц а 8.7 – 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

Рассмотрим 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, как показано в Таблице 8.8.

Учебно-исследовательская работа студента

Т а б л и ц а 8.8 – Полная матрица предпочтений проектов по методу «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							

Далее начинаем разреживать матрицу, задавая порог принятия решения, например $C = 1.33$, и оставляем в матрице те числа, которые больше или равны значению порога C . По данным полученной матрицы строим граф предпочтений, показанный на рисунке 8.2,а. 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. В то же время, 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 разрежается. Соответствующий ей граф предпочтений показан на рисунке 8.2,б. При этом несложно перевести граф к ярусно-параллельной форме, что свидетельствует о том, что решение получено. 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.

Таким образом, рассмотренные на данной лекции методы позволяют проводить анализ и сравнение проектов различной сложности.

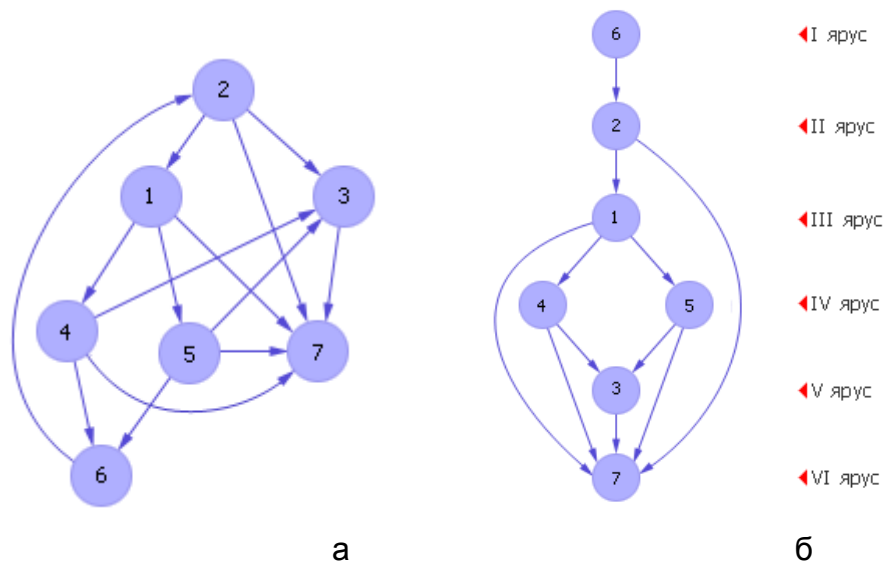


Рисунок 8.2 – Вид графа предпочтений

Лекция 8. Построение отчета о выполненной учебно-исследовательской работе

План отчета об учебно-исследовательской работе

Написание отчета о проделанной учебно-исследовательской работе является важным этапом в ходе научных исследований, поскольку по нему другие исследователи смогут ознакомиться с Вашими результатами. Учебно-исследовательская работа имеет следующую *структуру*:

- *Содержание*;

- *Введение*. Здесь отражается программа исследования, в процессе описания которой доказываемся актуальность выбранной проблемы и дается обоснование темы. Оформляется научный аппарат работы, выдвигается гипотеза, ставятся задачи, определяются методы исследования, описывается организация исследования, формулируются проблема и цель, объект и предмет исследования, его новизна и практическая значимость, коротко говорится о содержании каждой главы работы;

- *Раздел 1. Теоретическая часть работы*. Проводится обзор и анализ научных источников по выбранной проблеме. Делается акцент на неисследованных аспектах области исследования, даются определения используемых в работе понятий или обосновывается правильность и необходимость ссылки на авторские концепции или определения, делаются выводы. На страницах первой главы должны быть ссылки на исследователей, занимавшихся этой проблемой. Заниматься плагиатом, т.е. присваивать себе авторство, небезопасно. В случае уличения исследователь может считать свою научную карьеру завершенной.

Название раздела 1 связано с теоретическим обоснованием изучаемой проблемы. Названия параграфов соотносятся с отдельными задачами исследования;

- *Раздел 2. Практическая часть научной работы (констатирующий эксперимент)*. Содержит описание программы опытно-экспериментальной работы. Здесь отражаются цель, задачи, методы и методики, описание базы исследования, средства обработки данных. В качестве подразделов обычно применяются следующие:

Методика измерений. Излагается основная идея эксперимента и дано детальное описание методики измерений. Необходимо остановиться на средствах измерений и кратко на возможных ошибках. Особое внимание обратить на надежность представляемых результатов.

Описание установки. Подробно описывается экспериментальная установка. Особое внимание уделяется тем ее элементам, которые могут влиять на результаты измерений. Обязательно приводится рисунок или блок-схема установки.

Результаты измерений. Приводятся первичные необработанные данные измерений в виде таблиц и графиков. Необходимо детально описать условия, в которых они получены.

Учебно-исследовательская работа студента

- *Раздел 3. Анализ результатов измерений.* Содержит два подраздела:

Обработка результатов. В данном подразделе описываются используемые алгоритмы обработки результатов измерений. В текст обязательно должны быть включены формулы, поясняющие теоретические выкладки, для понимания тех модельных функциональных зависимостей, в рамках которых происходит интерпретация полученных экспериментальных данных.

При использовании программных пакетов необходимо указывать их название, версию и значения численных параметров, используемых при обработке данных.

Оценка погрешностей. Подробно описываются погрешности полученных экспериментальных данных и возможные ошибки измерений. Необходимо детально описать методику, с помощью которой определялись погрешности.

- *Раздел 4. Обсуждение полученных результатов.* Выделяется главный, основной результат, необходимо показать, насколько устойчивы полученные данные к изменениям условий измерений, четко определить область параметров проведения экспериментов, где данные верны. Дается оценка параметров функциональных зависимостей, их погрешностей, выполняется статистическая проверка гипотез об адекватности используемых моделей.

Далее осуществляется сравнение полученных экспериментальных данных с другими известными из литературы измерениями, а также с теорией. Для наглядности сравнения полученные и известные данные размещаются на одном графике или в одной таблице. Выполняется анализ сравнения полученных совпадений и расхождений, дается его объяснение.

Делаются выводы. В выводах делается заключение о подтверждении или опровержении рассматриваемых гипотез. При их написании следует использовать глаголы: исследованы, проверены, измерены и т.п.

Название раздела 2 соотносено с опытно-экспериментальной частью работы;

- *Выводы и рекомендации. "Ответ введению".* Краткие ответы на поставленные в начале исследования задачи, гипотезу и предположение. Отмечаются основные выводы исследования и новизна. Можно дать собственную оценку работе и указать пути улучшения методики проведения эксперимента, увеличения точности измерений.

Фразой типа "проведённое нами исследование не претендует на исчерпывающее исследование, требуют внимания ... вопросы, связанные с темой исследования ..." работа заканчивается;

- *Библиографический список.* Нумерованный перечень использованных при написании работы литературных источников по проблеме. Составляется в алфавитном порядке с обязательным указанием автора, названия работы, городом издания, названием издательства, годом издания и количеством страниц.

Содержание работы

Рассмотрим более подробно содержание работы.

Любое исследование начинается с обоснования актуальности темы исследования. Этот раздел отражает Ваши длительные поиски - как Вы к этому

Учебно-исследовательская работа студента

пришли, почему Ваша тема действительно интересна, каков замысел Вашего исследования, какие методологические подходы Вы намерены реализовать. Надо учесть, что Вы должны отразить актуальность именно Вашей конкретной темы, а не какой-либо иной из этого направления. При обосновании актуальности темы вопрос рассматривается с нескольких точек зрения. Во-первых, с практической точки зрения. Во-вторых, со стороны развития теории. Показывается масштаб и глобальность теории вопроса. На этой основе формируется противоречие - это взаимодействие между взаимоисключающими, но при этом взаимообуславливающими и взаимопроникающими друг в друга противоположностями внутри единого объекта и его состояний. Фраза типа *"Несмотря на имеющиеся теоретические работы в области ... наблюдается нехватка практических разработок, направленных на..."*.

На основе выявленного противоречия формулируется проблема - это объективно возникающий в ходе развития познания вопрос или целостный комплекс вопросов, решение которых представляет существенный практический или теоретический интерес. Таким образом, проблема исследования логически вытекает из установленного противоречия, из него вычленено то, что имеет отношение только к науке и переведено в плоскость познания.

Решение сформулированной проблемы и составляет цель исследования. На основе проблемы исследования устанавливается цель исследования. Цель - это то, что Вы намерены достигнуть в итоге работы. Например: *решение данной проблемы составляет цель исследования; или цель исследования - разработка (создание, апробация, формирование) у кого-либо чего-либо.*

Вслед за проблемой определяются объект и предмет исследования. Объект - то, что противостоит познающему субъекту - Вам, в познавательной деятельности, та часть практики, с которой Вы имеете дело. Предмет - это та сторона, тот аспект, та точка зрения, с которой исследователь познает целостный объект, выделяя при этом главные, наиболее существенные признаки объекта. Предмет либо совпадает с формулировкой темы, либо близок с ней по звучанию.

Предмет - то, что именно исследуют; объект - то, в чём исследуют.

Гипотеза - научное предположение, допущение, истинное значение которого неопределенно. Формулируя гипотезу, Вы строите предположение о том, каким образом Вы намерены достичь поставленной цели. Например: *Гипотезой исследования стало предположение о том, что ... повысится (улучшится, разовьется) за счет применения (использования, внедрения) ...*

В соответствии с целью, объектом, предметом и гипотезой исследования подразумевается выполнение ряда задач. Задача - это данная в определенных конкретных условиях цель деятельности.

1. *Теоретическая задача* (вариант формулировки по выбору):

- описать (выявить) теоретические основы...;
- провести научный анализ состояния теории и практики;
- проанализировать техническую литературу по направлению ...;

2. *Опытно-экспериментальная задача:*

- выявить влияние ... на ...;
- выявить закономерности, обусловленные

Практическая задача:

Учебно-исследовательская работа студента

- показать пути и способы практического применения ...;
- разработать и апробировать комплекс мероприятий ...;
- экспериментально проверить эффективность предложенной ...;
- наметить возможные пути

Методы исследований могут быть следующими: методы векторного анализа, методы факторного анализа, методы физической оптики и т.д.

Новизна исследования может заключаться в разработке, раскрытии, дополнении, обосновании, создании, предложении и т.п. чего-либо.

Оформление графической части отчета

Правильно оформленная графическая часть отчета о учебно-исследовательской работе позволяет эффективно донести суть работы до читателя. Это связано с тем, что графики, дающие визуальное представление о связи между величинами, что крайне важно при интерпретации полученных данных, легко воспринимаются, обладают значительной емкостью и вызывают больше доверия. На основе графика легче делать вывод о соответствии теоретических представлений данным эксперимента.

Графическое представление экспериментальных материалов используется для различных целей:

- наглядное изображение экспериментальных данных для качественного анализа поведения исследуемого объекта;
- определение (качественное, количественное) характерных точек и параметров процессов по наличию особенностей - максимумов, минимумов, точек перегиба, излома, скачка и т.п.;
- аппроксимация экспериментальных точек кривыми и определение по ним законов и закономерностей поведения объекта;
- проверка предполагаемых и обнаружение неизвестных зависимостей, выражаемых аналитическими функциями и т.д.

В зависимости от целей вид и способ графического представления может быть различным. Однако при оформлении графиков необходимо учитывать *основные требования их построения*:

- иметь оси координат с нанесенными на них наименованиями и единицами измерений. Оси графиков, за некоторым исключением, строят в декартовой системе координат, где по оси абсцисс (горизонтальной оси) откладывают аргумент (независимую величину), а по оси ординат (вертикальной оси) - функцию, зависящую от аргумента. Масштаб по осям должен позволять проводить анализ получаемых зависимостей. Оси необязательно должны содержать начало координат. При необходимости выбирают логарифмический или двойной логарифмический масштаб. Оси графика должны быть подписаны и включать масштабный коэффициент, если был выбран логарифмический или двойной логарифмический масштаб;
- содержать экспериментальные точки (если график строится по точкам, полученным в результате измерений). Если на одном графике показаны несколько зависимостей, то они обозначаются различными символами. При этом

Учебно-исследовательская работа студента

желательно указать масштаб ошибок, характерных для различных участков графика. Масштаб ошибок изображается отрезками прямых, проведенных через экспериментальные точки параллельно осям координат, с длиной, равной ошибке в масштабе соответствующей оси координат. В выбранном масштабе длина каждого отрезка должна равняться удвоенной погрешности величины, откладываемой параллельно оси. Центр отрезка должен приходиться на экспериментальную точку;

- аппроксимирующие кривые, проведенные по экспериментальным точкам, не должны скрывать самих точек и их реального разброса. При построении кривых следует следить за соотношением масштаба на графике и наносимых данных. Правильно построенные кривые должны заполнять все поле графика;

- если целью графического представления является проверка предполагаемой аналитической зависимости экспериментальных данных, то целесообразно использовать линеаризацию графика путем выбора нелинейного масштаба по соответствующей оси координат, например, логарифмического. Линеаризация значительно упрощает проверку типа зависимости;

- название графика должно быть приведено под ним и пояснять, к чему относится изображенная зависимость.

Для оформления схемы установки целесообразно использовать графические пакеты: Visio, Русплан и др. Выбор пакета определяется темой исследований и составом установки, которую необходимо изобразить.

Структура выступления на защите научной работы

На защите перед комиссией Вы должны придерживаться определенного порядка в изложении мыслей.

В начале доклада Вы представляетесь и обращаетесь к членам комиссии: *Уважаемый председатель Государственной аттестационной комиссии, уважаемые члены Государственной аттестационной комиссии, гости! Вашему вниманию предлагается ...*. Далее Вы называете тему работы и обосновываете:

- актуальность (2-3 предложения) выбранной Вами темы;
- противоречие;
- проблему и тему исследования;
- объект исследования;
- предмет исследования;
- гипотезу исследования (причем гипотеза может подтвердиться, частично подтвердиться или не подтвердиться);
- задачи исследования.

Изложение задач исследования - это 3-4 предложения, в которых Вы говорите о том, какие подходы к проблеме были Вами изучены, какие близки Вам лично, называете фамилии авторов. Имена и отчества ученых можно не называть, предварительно сказав фразу: *"Уважаемые члены Государственной аттестационной комиссии, позвольте в целях экономии времени не называть имена и отчества авторов ..."*. Однако будьте готовы к тому, что Вас спросят,

Учебно-исследовательская работа студента

как звали того или иного ученого, какой проблемой он занимался и на какие его работы Вы ссылаетесь.

Далее Ваш монолог строится следующим образом: *Решая первую задачу, был проведен анализ литературы по проблеме исследования... . Решая вторую задачу, мы провели констатирующий эксперимент, который показал, что... В ходе эксперимента были отмечены такие-то особенности... . Решая третью задачу, мы использовали такие-то методики* (то есть говорите о формирующем эксперименте), *мы добились таких-то результатов*. Другими словами, через задачи Вы излагаете основное содержание работы.

Время выступления составляет 7-8 минут, т.е. примерно 1.5-2 страницы текста 14 шрифтом через полтора интервала с учетом возможных заминок, обусловленных волнением. Темп изложения средний, как при обычном разговоре, без так называемого «тарактения», «как на пожар», что случается, когда объем доклада не был проверен на время. Желательно, несмотря на значение излагаемого материала исследований, доклад выучить наизусть. Это позволит избежать ненужных остановок. Читать доклад не принято.

В ходе своего выступления Вы демонстрируете членам комиссии сравнительные таблицы (диаграммы) с результатами экспериментов, материалы из приложения и т.п.

В конце Вашего выступления могут быть заданы вопросы. Вы должны до конца выслушать вопрос, а свой ответ желательно начать со слов: *"Благодарю за вопрос, ..."*. Во время ответа недопустимы словосочетания «как я уже говорил в докладе». Лучше повторить часть выступления без данного вступления. Отвечать на вопросы следует корректно и дипломатично. Помните, что вопросы, задаваемые членами комиссии, будут касаться, в основном, изложенного в докладе материала, и непосредственно темы Вашего исследования.