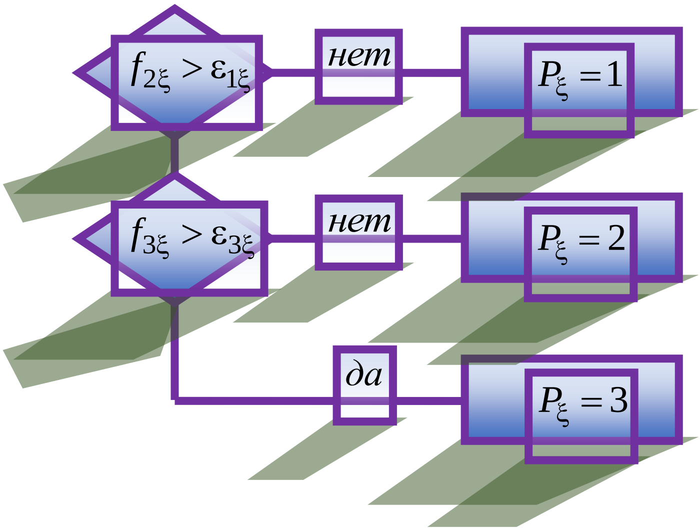
В. В. КНЯЗЕВ, А. В. ЗАЙЦЕВ

ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ



Тамбов
♦ Издательский центр ФГБОУ ВО «ТГТУ» ♦
2025

Министерство науки и высшего образования Российской Федерации

**Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Тамбовский государственный технический университет»**

В. В. КНЯЗЕВ, А. В. ЗАЙЦЕВ

ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ

Утверждено Ученым советом
ФГБОУ ВО «Тамбовский государственный технический университет»
в качестве учебного пособия для студентов 1–2 курсов,
обучающихся по направлению подготовки
15.04.06 «Мехатроника и робототехника»

Учебное электронное издание

Тамбов
♦Издательский центр ФГБОУ ВО «ТГТУ»♦
2025

УДК 001.89(075.8)
ББК 72.5я73
К54

Рецензенты:

Доктор технических наук, профессор кафедры баллистики
и систем управления летательных аппаратов Военной академии
имени Петра Великого, начальник отдела Научно-производственного
центра автоматики и приборостроения имени академика Н. А. Пилюгина
В. С. Гаврилов

Доктор технических наук, профессор, профессор кафедры
«Информационные системы и защита информации» ФГБОУ ВО «ТГТУ»
В. Е. Дидрих

Князев, В. В.

К54 Основы научных исследований [Электронный ресурс] : учебное пособие / В. В. Князев, А. В. Зайцев. – Тамбов : Издательский центр ФГБОУ ВО «ТГТУ», 2025. – 1 электрон. опт. диск (CD-ROM). – Системные требования : ПК не ниже класса Pentium IV ; RAM 512 Mb ; необходимое место на HDD 9,5 Mb ; Windows 7/8/10/11 ; дисковод CD-ROM ; мышь. – Загл. с экрана.

ISBN 978-5-8265-2970-6

Приведены основы классической и современной математической теории, которая является фундаментом научных исследований. Изложены положения теории вероятностей и математической статистики, рассмотрены случайные события и случайные величины. Особенно выделены определения теории эффективности и принятия решений как в условиях определенности и риска, так и в условиях неопределенности и активного противоборства. Описан алгоритм иерархий В. В. Подиновского, дан анализ современных информационных теорий, изложена структура моделирования в научных исследованиях от классификации моделей до имитационного моделирования. Приведено описание популярных пакетов моделирования, к которым относятся Mathcad, MATLAB, программный комплекс ELECTRONICS WORKBENCH, даны примеры решения задач проектирования узлов, устройств и систем автоматического управления.

Предназначено для студентов 1–2 курсов, обучающихся по направлению подготовки 15.04.06 «Мехатроника и робототехника», изучающих теорию управления, также может быть полезно научным сотрудникам, специалистам, занимающимся оценкой, анализом и конструированием сложных организационно-технических систем и комплексов.

УДК 001.89(075.8)
ББК 72.5я73

ISBN 978-5-8265-2970-6

ВВЕДЕНИЕ

В последние десятилетия наблюдается высокий рост технологичности оборудования. Если говорить о таких областях как авиация, ракетостроение, медицинская техника, автомобилестроение, то необходимо отметить высокую роль профилактического обслуживания оборудования. В приведенных областях цена отказа или сбоя оборудования может быть очень высокой, а иногда и приводит к неоценимым потерям, связанным с гибелью людей.

Обслуживание подобных систем требует высоких материальных затрат, но в то же время не исключает возможного возникновения сбоев. Очень важную роль занимает процесс принятия решения о целесообразности ремонта данного оборудования или о необходимости отказа от дальнейшего использования и списания.

В связи с высокой стоимостью приведенных видов техники в настоящее время на первое место выходит обеспечение длительного срока службы объектов при минимизации затрат на содержание и техническое обслуживание.

Возникает вопрос, какими качествами должен обладать специалист, который уверенно чувствует себя в названной профессиональной сфере. Ответ на данный вопрос дает краткий анализ индикаторов достижимости в рамках компетентностного подхода.

Данный специалист способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования в профессиональной деятельности. Он умеет применять и владеет методами статистической обработки, интерполяции и аппроксимации при оценке и анализе полученных экспериментальных данных.

За счет введения в структуру системы модуля прогнозирования технического состояния возможно предотвратить возникновение сбоев системы в процессе выполнения основной задачи, поставленной перед объектом.

Для более глубокого обоснования и рассмотрения предлагаемого подхода требуется рассмотрение основ научных исследований, которые изложены в данном учебном пособии.

В основе учебного пособия лежит многолетний опыт авторов, который накоплен в ходе проведения занятий с будущими специалистами, а также соискателями ученых степеней кандидата и доктора наук.

1. АЛГЕБРА

Алгебра – это раздел математики, посвященный изучению операций над элементами множества произвольной природы, обобщающий многие математические дисциплины и являющийся основой научных исследований как прикладных, так и фундаментальных. Овладение основами алгебры является необходимым условием успешного проведения любых научных исследований. Основой алгебры является *теория множеств,* которая является предметом текущего занятия.

1.1. ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ МНОЖЕСТВ

1.1.1. ПОНЯТИЕ МНОЖЕСТВА

Множество и его элементы

Множество – одно из первичных фундаментальных понятий математики. Строгого единого определения термина «множество» в рамках самой теории множеств не существует.

Определение 1.1.1. Множество – совокупность элементов (объектов) произвольной природы.

Определение 1.1.2. Множество – совокупность, набор элементов, объединенных общими свойствами.

В Определении 1.1.2 к элементам предъявляется требование – наличие общих свойств. Обратим внимание на это обстоятельство.

Пример: множество обучающихся в учебном заведении.

Множество бесструктурно!!!

Примечание.

Простра́нством называется математическое множество, имеющее структуру, определяемую аксиоматикой свойств его элементов (например, точек в геометрии, векторов в линейной алгебре, событий в теории вероятностей и так далее), т.е. пространство уже не множество.

Математи́ческая структу́ра – название, объединяющее понятия, общей чертой которых является их применимость к множествам, природа которых не определена. Для определения самой структуры задают отношения, в которых находятся элементы этих множеств. Затем постулируют, что данные отношения удовлетворяют неким условиям, которые являются аксиомами рассматриваемой структуры [Математическая энциклопедия (в 5 т.). – М.: Советская Энциклопедия, 1985. – Т. 5].

Множества, как правило, обозначаются заглавными латинскими буквами A, B, X, Y, но можно и сочетанием букв и цифр.

Определение 1.1.3. Элементы множества – объекты, составляющие множество.

Элементы множества обозначаются строчными латинскими буквами a, b, x, y или a_1, a_2, a_3.

Если объект a – элемент множества A, то говорят, что «элемент a принадлежит множеству A» и записывают $a \in A$.

Если объект a не является элементом множества A, то говорят, что «элемент a НЕ принадлежит множеству A» и записывают $a \notin A$.

Элементы во множествах не могут повторяться[1], т.е. если есть некий элемент x и множество A, то можно лишь сказать, что $x \in A$ либо $x \notin A$, но нельзя сказать, сколько раз он содержится в A. Если записать множество $A = \{x, x\}$, то запись не будет ошибочной – просто она будет значить то же самое, что и $A = \{x\}$.

Определение 1.1.1.* Множество, элементами которого являются множества, называется **семейством множеств**.

Количество элементов во множестве называется *мощностью множества*.

---

[1] Если в решаемой задаче важно количество одинаковых элементов, то следует применять математический аппарат теории комплектов.

Определение 1.1.4. Мощность множества – это количество его элементов.

Записывают $|A|$ или card(A) – от «кардинальное число множества». Реже #A.

Пример: $|A| = 5$, card(A) = 5 или #A = 5.

Различают конечные и бесконечные множества.

Определение 1.1.5. Конечное множество – множество, состоящее из конечного количества элементов.

Причем не важно: известно или неизвестно это количество, главное – оно конечное.

Пример: множество обучающихся в учебном заведении конечно.

Определение 1.1.6. Бесконечное множество – множество, состоящее из бесконечного количества элементов: card(A) = ∞.

Примечание. Бесконечные множества бывают *счетные* (счетной мощности), и *несчетные* (например, мощности континуума). Для этих мощностей существуют особые обозначения – мощность счетного множества обозначается через $\aleph_0$ (читается «алеф нуль»), а мощность континуума – C.

Вопрос: счетно ли множество натуральных чисел? четных чисел, целых чисел, рациональных чисел, действительных чисел?

Определение 1.1.7. Пустое множество – множество, не содержащее ни одного элемента: $A = \varnothing$.

Пример: $B = \{\varnothing\}$ – не пустое множество. Это семейство множеств, содержащее один элемнт – пустое множество, card(B) = 1.

Способы задания множеств

1. Перечислением его элементов:

$$A = \{a, b, c\},\ A = \{a_1, a_2, a_3\}.$$

Этим способом можно задать множество обучающихся вашей группы, но уже множество обучающихся в учебном заведении задать сложно, а бесконечное множество и вовсе невозможно.

2. Описанием свойств (свойства) его элементов.

Определение 1.1.8. Свойство элемента, определяющее принадлежит ли элемент множеству или не принадлежит, называется *характеристическое свойство множества* (не элемента, а множества!).

$$A = \{x\colon P(x)\} \quad \text{или} \quad A = \{x \mid P(x)\}.$$

Примеры конечных и бесконечных множеств, заданных данным способом:

$$A = \{x\colon x \in N \text{ и } x < 10\}, \quad A = \{x\colon x > 0\},$$

$$A = \{a_i,\ i = \overline{1, I}\}, \quad A = \{a_i,\ i = \overline{1, |A|}\},$$

Примечание.

Пусть даны два множества $A = \{a_i\}, i = \overline{1, I}, \quad B = \{b_i, i = \overline{1, I}\}$, тогда $card(A) = 1, card(B) = I$, так как в соответствии с ГОСТ Р 54521 – 2011[2] вся информация об элементах множества содержится внутри скобок. Таким образом, множество A содержит один элемент, некое a_i, а множество B содержит совокупность из I элементов. Фигурные скобки играют роль, подобную операторным скобкам в программировании.

Обратимся к определениям термина «множество» (Определение 1.1.1 и Определение 1.1.2).

Определение 1.1.2 требует обязательного наличия общих свойств у элементов. Определение 1.1.8 называет это свойство характеристическим свойством. А первый способ задания множеств (перечислением его элементов) ничего не говорит о наличии каких-либо общих свойств элементов. По нему элементы множества можно просто перечислить без упоминания каких-либо их свойств, общих или различающихся, т.е. элементы множества могут иметь общие свойства, а могут и не иметь таковых. Поэтому правильным будет Определение 1.1.1, хотя в практических задачах, как правило, в множества включают элементы, обладающие каким-то общим для них свойством.

Подытожим: запись $A = \{\ldots\}$ означает следующее. Слева от знака равенства *название множества* «*A*», справа, обязательно в фигурных скобках, информация о его элементах – либо их перечисление, либо указание характеристического свойства.

---

[2] ГОСТ Р 54521–2011. Статистические методы. Математические символы и знаки для применения в стандартах. – М. : Стандартинформ 2012. – 32 с.

Записи $A\{\ldots\}$, A: $\{\ldots\}$ и т.п. бессмысленны.

Запись $\{A\}$ означает множество без указания его названия, состоящее из одного элемента – «A».

Определение 1.1.9. Если два множества состоят из одних и тех же элементов, то эти множества называют *равными множествами*.

$$A = B, \quad A = \{a_1, a_2, a_3\}=\{a_2, a_1, a_3\}.$$

Подмножество

Определение 1.1.10. Если множество A содержит только часть элементов множества B (и не содержит иных, не принадлежащих множеству B элементов), то множество A является *подмножеством множества B*. Различают собственное и несобственное подмножества и обозначают, соответственно, $A \subset B$, $A \subseteq B$.

Собственное подмножество некоторого множества не может содержать все элементы этого множества $A \subset B \Rightarrow |A| < |B|$.

Несобственное подмножество некоторого множества может содержать все элементы этого множества $A \subseteq B \Rightarrow |A| \le |B|$.

Примеры подмножеств.

Пусть $B = \{x, y, z\}$, $A = \{x\}$.

Тогда *верными* будут записи

$$A \subset B,\ \{x\} \subset B,\ x \in B.$$

Неверные записи

$$A \in B,\ x \subset B.\ \text{(Почему?)}$$

Запись $A \in B$ некорректна, так как A не является элементом B.

Запись $x \subset B$ некорректна, так как x не является множеством.

множество $\subset$ множество
элемент $\in$ множество

Пустое множество является подмножеством любого множества.

Определение 1.1.11. Булеан множества A это множество всех его подмножеств. Обозначается как «два» с верхним индексом, где в этом индексе – название множества

$$2^A = \{X : X \subseteq A\}.$$

Неправильно читать «два в степени A».

Определение 1.1.12. Универсальное множество – такое множество, что все остальные рассматриваемые множества являются его подмножествами. Обозначается буквой E.

Пример: обучающиеся в учебном заведении.

Кванторы

В математике известны и применяются кванторы, из которых наиболее часто применяются:

1. Квантор существования $\exists$ – «существует хотя бы один», $\nexists$ – соответственно, «не существует».
2. Квантор всеобщности $\forall$ – «для всех», «для любого».

Задание. Записать собственное подмножество через квантор.

1.1.2. АЛГЕБРА МНОЖЕСТВ

Алгебра

В теории множеств рассматривают *операции*, позволяющие из одних элементов множества получать другие элементы этого множества.

Определение 1.1.13. Алгебраическая операция – действие над элементами множества (которые называются операнды), в результате которого опять получается элемент этого множества. По количеству операндов различают унарные, бинарные и n-арные операции. Если в операции порядок следования операндов не имеет значения, то операция коммутативная, иначе – некоммутативная (сложение чисел – коммутативна, вычитание – некоммутативна).

Определение 1.1.14. Алгебраическая система (алгебра) – упорядоченная пара множеств $A(M, O)$. Первое множество (M) – элементы произвольной природы – *носитель, основание алгебры*. Второе множество (O) – операции над элементами первого множества, результатом которых является элемент из первого множества – *сигнатура алгебры*.

Пример. Алгебра натуральных чисел с основанием – множеством натуральных чисел N и сигнатурой из операций сложения, умножения, возведения в степень $A(N, \{+, *, \wedge\})$. Обратим внимание, что операции вычитания и деления не включены в сигнатуру.

Еще примеры алгебр: целых, рациональных, действительных, комплексных чисел, множеств, матриц, векторов, булева.

Задание. Написать сигнатуры для этих алгебр.

Основная ошибка обучающихся: смешение алгебр – основание от одной алгебры, операции от другой, например, основание из алгебры множеств, сигнатура из булевой алгебры.

Алгебра множеств

Пусть заданы множества A, B, C, E.

Определение 1.1.15. Объединением множеств A и B называется множество элементов, принадлежащих по крайней мере одному из данных множеств (т.е. либо A, либо B, либо и A и B). Обозначают $A \cup B$ и читают «объединение A и B».

$$C = A \cup B = \{x\text{: } x \in A \text{ или } x \in B\}.$$

Пусть $A = \{x\text{: } P(x)\}$, $B = \{x\text{: } Q(x)\}$.

Тогда $A \cup B = \{x\text{: } P(x) \vee Q(x)\}$.

Это бинарная коммутативная операция.

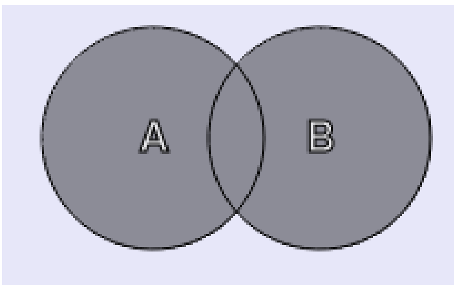


Рис. 1.1. Иллюстрация к алгебраической операции «объединение множеств»

Задание. Привести примеры.

Определение 1.1.16. Пересечением множеств A и B называется множество элементов, общих для A и B, т.е. принадлежащих одновременно и A и B.

$$C = A \cap B = \{x\colon x \in A \text{ и } x \in B\}, \quad A \cap B = \{x\colon P(x) \wedge Q(x)\}.$$

Это бинарная коммутативная операция.

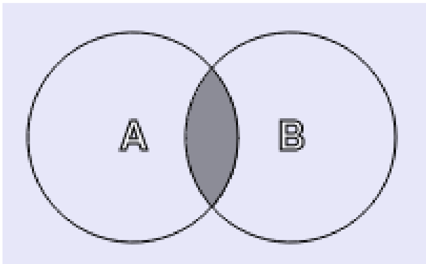


Рис. 1.2. Иллюстрация к алгебраической операции «пересечение множеств»

Задание. Привести примеры.

Вопросы. Пусть $C = A \cap B$.

Верно ли, что $C \subset A$?

$C \subset B$?

card (C) > card (A)?

$C \subset A \cup B$?

Напишите еще верные и неверные выражения.

Все эти бинарные операции можно распространить на несколько множеств:

$$D = A \cup B \cup C.$$

Пусть дано n множеств $A_1, A_2, \ldots, A_i, \ldots, A_n$. Тогда множество A есть объединение всех этих n множеств:

$$A = \bigcup_{i=1}^{n} A_i.$$

Аналогично пересечение

$$A = \bigcap_{i=1}^{n} A_i .$$

Определение 1.1.17. Дополнение множества A (до универсального E).

Это унарная (одноместная) операция.

$$\overline{A} = \{x : x \notin A\}.$$

Еще обозначение $\neg A$.

Задание. Привести примеры.

Теперь можно уточнить определение универсального множества.

Универсальным называется множество E такое, что для любого множества A справедливы соотношения

$$A \cap E = A, A \cup E = E.$$

Рис. 1.3. Иллюстрация к алгебраической операции «дополнение множества до универсального множества»

Объединение, пересечение и дополнение образуют *функционально полную систему* операций *алгебры* множеств, т.е. комбинацией (суперпозицией) этих операций можно представить любую операцию.

Еще операции:

Разность множеств – бинарная НЕкоммутативная операция:

$A \backslash B = \{x: x \in A$ и $x \notin B\}$.

$A \backslash B = \{x: P(x) \wedge \neg Q(x)\}$.

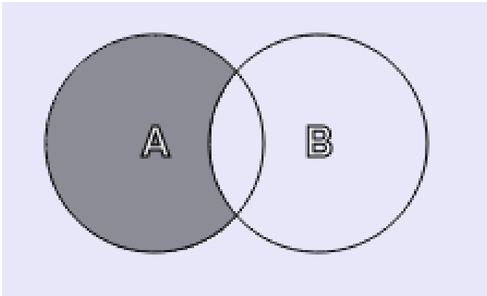


Рис. 1.4. Иллюстрация к алгебраической операции «разность множеств»

Выразим ее через известные нам операции

$$A \setminus B = A \cap \overline{B}.$$

Задание. Привести примеры.

Правило приоритетов выполнения операций: наивысший – дополнение, потом – пересечение, наименьший – объединение.

Изменить приоритет можно скобками. У выражения в скобках приоритет наивысший.

Симметрическая разность множеств – бинарная коммутативная операция

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

Задание. Привести примеры.

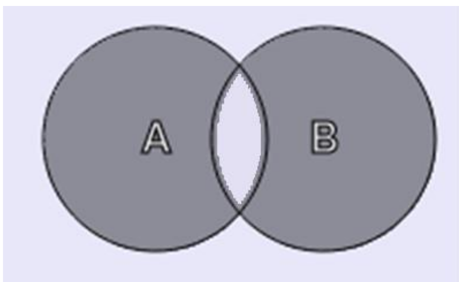


Рис. 1.5. Иллюстрация к алгебраической операции «симметрическая разность множеств»

1.1.3. ОСНОВНЫЕ ТОЖДЕСТВА АЛГЕБРЫ МНОЖЕСТВ

1	$A \cup B = B \cup A$	$A \cap B = B \cap A$
2	$A \cup B \cup C = A \cup (B \cup C)$	$A \cap B \cap C = A \cap (B \cap C)$
3	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
4	$A \cup \varnothing = A$	$A \cap E = A$
5	$A \cup \overline{A} = E$	$A \cap \overline{A} = \varnothing$

Тождества слева и справа ***двойственны***, т.е. получены заменой $\cap$ на $\cup$, E на $\varnothing$ и наоборот.

Тождества используются для доказательства тех или иных утверждений относительно множеств.

1.1.4. ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ ПОДМНОЖЕСТВА

Характеристическая функция (Х.Ф.) *подмножества* X, $X \subseteq U$ – функция с областью значений {0, 1}, вычисляемая для любого элемента u, $u \in U$ (т.е. определена на множестве U), результатом которой является число, обозначающее принадлежность элемента подмножеству.

Или иначе:

Характеристическая функция χ *подмножества* $X \subseteq U$ для элемента $u \in U$ показывает, принадлежит ли u множеству X в соответствии с правилом:

$$\chi_X(u) = \begin{cases} 0, & u \notin X, \\ 1, & u \in X, \end{cases} \quad X \subseteq U,\ u \in U.$$

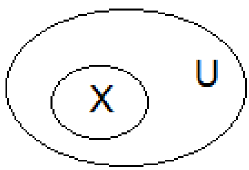


Рис. 1.6. Характеристическая функция

Операции над подмножествами множества U соответствуют операциям с их характеристическими функциями. В частности, пересечению множеств соответствует произведение характеристических функций: $\chi_{A\cap B}(u) = \chi_A(u)\chi_B(u)$. Дополнению (до U) соответствует функция $1 - \chi$, если χ – характеристическая функция исходного множества.

Развитие понятия «характеристическая функция» дал Ло(ю)тфи Заде (Лютфали Рагим оглы Алескерзаде) в теории нечетких множеств.

1.2. СООТВЕТСТВИЯ. СВОЙСТВА СООТВЕТСТВИЙ

1.2.1. ДЕКАРТОВО ПРОИЗВЕДЕНИЕ МНОЖЕСТВ

Пусть A и B – произвольные множества.

Неупорядоченная пара на множествах *A* и *B* – это любое множество $\{a, b\}$, где $a \in A$, $b \in B$.

Упорядоченная пара на множествах A и B, обозначаемая записью (a, b), или $<a, b>$ определяется не только самими элементами $a \in A$ и $b \in B$, *но и порядком, в котором они записаны*. И в этом состоит ее существенное отличие от неупорядоченной пары.

Обобщением понятия упорядоченной пары является упорядоченный n-набор, или *кортеж*. Если кортеж состоит только из чисел, его называют «вектор» (вектор-строка) и охватывают круглыми скобками, иначе кортеж охватывают угловыми скобками.

Определение 1.2.1. Множество всех(!) кортежей длины n на множествах $A_1, A_2, \ldots, A_n$ называют **декартовым (прямым) произведением этих множеств** и обозначают

$$A_1 \times A_2 \times \ldots A_n.$$

Таким образом

$$A_1 \times A_2 \times \ldots A_n = \{(a_1, a_2, \ldots, a_n) : a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n\}.$$

Пример.

Пусть $A = \{a_1, a_2, a_3\}$, $B = \{b_1, b_2\}$

Тогда $C = A \times B$, где $C = \{c_1, c_2, c_3, c_4, c_5, c_6\} = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2), (a_3, b_1), (a_3, b_2)\}$.

Верно ли:

$a_1 \in C, \ A \subset C$?

Ответ – нет.

Важно: $|A \times B| = |A| \cdot |B|$.

Пример – пространство характеристик объекта.

$$M = \{m, m \in [50, 300] \text{ кг}\}, \ Ц = \{\text{"красный"}, \text{"синий"}\},$$

$$V = \{1 \text{ м}^3, 2 \text{ м}^3, 3 \text{ м}^3\}, \quad O = М \times Ц \times V.$$

Вопрос: $|O| = ?$

Тогда

$o_1 = \left\langle 70 \text{ кг}, \text{"синий"}, 3 \text{ м}^3 \right\rangle, \ o_1 \in O,$

$o_2 = \left\langle 3 \text{ м}^3, \text{"синий"}, 70 \text{ кг} \right\rangle, \ o_2 \notin O.$

Обычно реальное пространство характеристик ПХ содержит не все элементы декартового произведения, т.е. возможны не все комбинации параметров, тогда ПХ $\subset$ М×Ц×V.

Определение 1.2.2. Степенью *n* множества *A* называется его прямое произведение самого на себя *n* раз.

Примеры: R, R^2, R^3 – числовые прямая, плоскость, трехмерное пространство соответственно, $x_1 \in R, \ x_2 \in R^2, \ x_3 \in R^3$, x_1, x_2, x_3 – действительное число, точка на плоскости и точка в трехмерном пространстве, соответственно.

$x_1 = (2, \ 3) \neq x_2 = (3, \ 2)$ – различные точки плоскости.

$e = \{0, 1\}, \ E = e^8, \ (0, 1, 1, 0, 1, 0, 0, 1) \in E$, E – множество, содержащее все байты.

Пусть $A = [a, \ b]$ и $B = [c, \ d]$ – два отрезка вещественной прямой. Тогда их декартовым произведением будет множество точек плоскости, которое имеет вид (рис. 1.7).

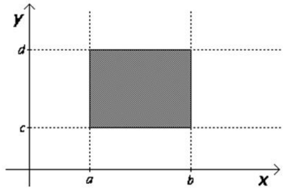


Рис. 1.7. Декартово произведение

Еще *пример*:

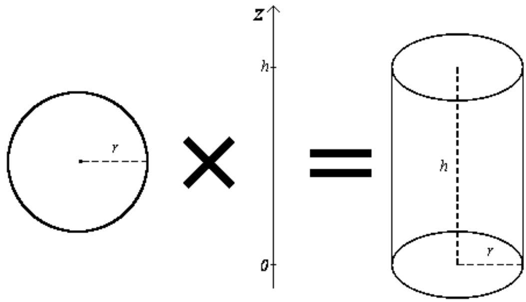


Рис. 1.8. Декартово произведение

Важно: $A \times \varnothing = \varnothing \times A = \varnothing$.

1.2.2. СООТВЕТСТВИЕ

В математике, как и в реальной жизни, различные объекты могут чему-то соответствовать или не соответствовать, находиться меж собой в определенных отношениях или наоборот – не находиться. И основой формализации, если угодно – математизации, здесь также служат множества. Человек может соответствовать профессии, зарплата соответствовать должности, оценка – знаниям.

Определение 1.2.3. Соответствием между множествами X и Y называется всякое подмножество декартова произведения этих множеств. Соответствия принято обозначать буквами P, S, T и др. Если xSy – соответствие между элементами множеств X и Y, то, согласно определению, $S \subseteq X \times Y$.

Поскольку соответствие – это подмножество, то его можно задать как любое множество, либо перечислив все пары элементов, находящихся в данном соответствии, либо указав характеристическое свойство элементов этого подмножества.

Например, соответствие между множествами $X = \{1, 2, 4, 6\}$ и $Y = \{3, 5\}$ можно задать:

1) при помощи предложения с двумя переменными: $a < b$ при условии, что $a \in X$, $b \in Y$;

2) перечислив пары чисел, принадлежащих подмножеству декартова произведения $X \times Y$: $\{(1, 3), (1, 5), (2, 3), (2, 5), (4, 5)\}$. К этому способу задания относят также задание соответствия при помощи ориентированного графа на плоскости.

Пример некоторого соответствия (рис. 1.9).

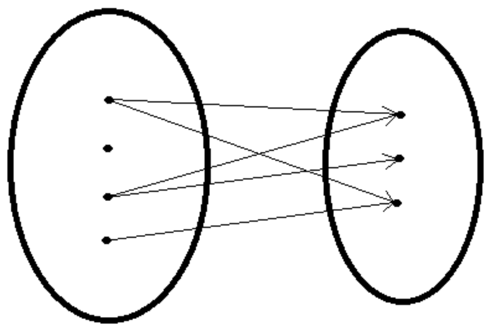


Рис. 1.9. Пример некоторого соответствия

Определение 1.2.4. Область определения соответствия $D(S)$ ($\mathrm{Dom}(S)$) – множество всех первых компонент упорядоченных пар соответствия S.

$$D(S) = \{x \in X \mid \exists y \in Y : (x, y) \in S\}.$$

Определение 1.2.5. Область значений соответствия $R(S)$ ($\mathrm{Im}(S)$, $E(S)$, $\mathrm{cod}(S)$, $\mathrm{ran}(S)$) – множество всех вторых компонент упорядоченных пар соответствия S.

$$R(S) = \{y \in Y \mid \exists x \in X : (x, y) \in S\}.$$

Пример:

Множество X = {Иванов, Петров, Сидоров, Федоров} – группа обучающихся, сдающая экзамен;

Множество Y = {отл., хор., удовл., неудовл.} – множество оценок.

Итоги экзамена множества пар (соответствие) S = {<Иванов, отл.>, <Петров, хор.>, <Сидоров, отл.>}.

Федоров – не явился.

Тогда область определения $D(S)$ = {Иванов, Петров, Сидоров}, область значений $R(S)$ = { отл., хор }.

Важно: на одной и той же паре множеств можно задать различные соответствия, в том числе – с различными $D(S)$ и $R(S)$.

1.2.3. СВОЙСТВА СООТВЕТСТВИЙ

Определение 1.2.6. Соответствие S называется *функциональным,* если каждому элементу множества X соответствует *не более одного элемента во множестве Y*.

То есть $\forall x \in X, y_1 \in Y, y_2 \in Y$ из $(x, y_1) \in S$ и $(x, y_2) \in S \Rightarrow y_1 = y_2$.

Определение 1.2.7. Соответствие S называется *инъективным,* если каждому элементу множества Y соответствует *не более одного элемента во множестве X*.

То есть $\forall y \in Y, x_1 \in X, x_2 \in X$ из $(x_1, y) \in S$ и $(x_2, y) \in S \Rightarrow x_1 = x_2$.

Определение 1.2.8. Соответствие S называется *всюду определенным*, если каждому элементу множества X соответствует *хотя бы один элемент во множестве Y*.

$$D(S) = X.$$

Определение 1.2.9. Соответствие S называется *сюръективным,* если каждому элементу множества Y соответствует *хотя бы один элемент во множестве X*.

$$R(S) = Y.$$

1.3. ОТОБРАЖЕНИЯ. ОТНОШЕНИЯ. СВОЙСТВА ОТНОШЕНИЙ

1.3.1. ОТОБРАЖЕНИЯ

Пусть X и Y – произвольные множества. *Отображением* множества X на множество Y называют правило (соответствие), которое *каждому элементу множества X ставит в соответствие единственный для этого элемента элемент множества Y*.

Обозначается:

$$f\colon X \to Y \text{ или } X \xrightarrow{f} Y.$$

При этом, к элементам множества Y никакие требования не предъявляются. В частности, в множестве Y могут существовать элементы, не соответствующие ни одному элементу из множества X, или один элемент множества Y соответствует нескольким элементам множества X.

Если элементу x, $x \in X$, ставится в соответствие элемент y, $y \in Y$, то это обозначают $y = f(x)$ и называют значением отображения «в точке x» или *образом* элемента x, *значением функции в точке* x. При этом элемент x называют *прообразом* элемента y, *значением аргумента функции*. Само «отображение в точке» называется функцией или функциональной зависимостью.

Важно. К элеменам множеств X и Y никаких требований не предъявляется. Они могут быть объектами произвольной природы, а не обязательно числами. Но принципиально важно, что при функциональной зависимости каждому значению аргумента соответствует единственное значение функции.

Множество X называют областью определения функции (отображения) и обозначают $D(f)$, а множество значений обозначают $\mathrm{Im}(f)$ и называют образом отображения f. $\mathrm{Im}(f)$ является подмножеством множества Y: $\mathrm{Im}(f) \subseteq Y$.

Определение 1.3.1. Всюду определенное функциональное соответствие из множества X в множество Y называется *отображением множества X в множество Y*.

Если в отображении выполняются требования Определения 1.2.7 или Определения 1.2.9, то отображение называется соответственно *инъективным* или *сюръективным* (отображение множества в множество и отображение множества на множество, соответственно).

В общем случае в отображении требования Определения 1.2.7 или Определения 1.2.9 могут не выполняются.

Определение 1.3.2. Биективное отображение множества (биекция) – взаимооднозначное отображение (инъективное и сюръективное отображение).

Важно. Функция отображает точки одного множества в точки другого множества независимо друг от друга. Но если на обоих множествах заданы *структуры* и функция отображает *не только элементы множества, но и его структуру*, то такая функция называется *оператор*.

Пример: преобразование системы координат в Евклидовом пространстве.

Задание отображений

Для того, чтобы определить (задать) отображение множества X в множество Y нужно задать сами множества X и Y, а затем задать правило, с помощью которого можно для каждого x, $x \in X$ находить соответствующий ему элемент y, $y \in Y$. Это правило можно задать:

1. Таблицей, если множество X конечное и имеет небольшое количество элементов.
2. С помощью формулы (математического выражения).
3. С помощью некоторого алгоритма (процедуры). Все зависит от конкретной ситуации.

Отображение f множества X на множество Y считается заданным, если каждому элементу $x \in X$ сопоставлен единственный элемент $y \in Y$.

Важно. Некоторые отображения имеют свои собственные названия:

Название	$D(S)$	$R(S)$
Функция (отображение)	Требований не предъявляется	Требований не предъявляется
Функционал	Требований не предъявляется	$R(S) \subseteq R$
Числовая функция	$D(S) \subseteq R^n$	$R(S) \subseteq R$
Предикат	Требований не предъявляется	$R(S) = \{0, 1\}$

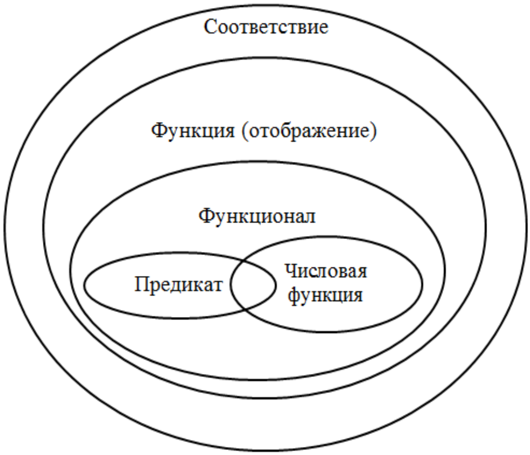


Рис. 1.10. Соотношение терминов типов соответствий

1.3.2. БИНАРНЫЕ ОТНОШЕНИЯ НА МНОЖЕСТВЕ

Определение 1.3.3. *Бинарным отношением* на множестве X (между элементами одного и того же множества) называется соответствие из множества X в X, т.е. всякое подмножество декартова квадрата X^2.

$$S: X \to X.$$

n-арное отношение – это всякое подмножество декартова произведения X^n.

Рассмотрим бинарные отношения.

Примеры.

Отношения на множестве целых чисел *Z* «делится», «делит», «равно», «больше», «меньше», «взаимно просты».

1. Отношение $\leq$ выполняется для (7, 9), (7, 7), а для (9, 7) не выполняется.

2. Отношение «иметь общий делитель, отличный от единицы», выполняется для пар (6, 9), (4, 2), (2, 4), (4, 4), но не выполняется (7, 9), (9, 7).

3. Отношение «a делитель b» выполняется для (2, 4), (4, 4), не выполняется для (4, 2), (7, 9).

Отношение на множестве точек действительной плоскости:

1. Отношение «находится на одинаковом расстоянии от начала координат».

2. Симметричны относительно какой-либо оси координат.

На множестве прямых пространства отношения «параллельны», «взаимно перпендикулярны», «скрещиваются», «пересекаются», «совпадают».

На множестве окружностей плоскости «пересекаются», «касаются», «концентричны».

На множестве людей в организации: «начальник–подчиненный» и т.д.

Свойства бинарных отношений

Определение 1.3.4. Если для любого $x \in X$ имеет место xSx, то отношение S называется *рефлексивным*.

Определение 1.3.5. Если для любых x, $y \in X$ из xSy следует ySx то отношение S называется *симметричным*.

Определение 1.3.6. Если для любых x, y, $z \in X$ из xSy и ySz следует xSz, то отношение S называется *транзитивным*.

Определение 1.3.7. Симметричное, рефлексивное транзитивное отношение – отношение *эквивалентности*.

Определение 1.3.8. *Разбиение множества* $A = \{A_i,\ i = 1..\ N,\ \forall i\ A_i \subseteq A,\ \forall\ i, j,\ i \neq j\ A_i \cap A_j = \varnothing,\ \cup_i A_i = A\}$ – семейство множеств, являющихся подмножествами данного множества, попарное пересечение которых равно пустому множеству, а объединение всех подмножеств равно исходному множеству.

Определение 1.3.9. *Фактор-множество* – разбиение множества на подмножества, содержащие элементы одного класса эквивалентности.

Выводы.

1. Знание основ алгебры и теории множеств является фундаментально необходимым для успешного формулирования постановки задачи и проведения научных исследований.
2. Каждый символ, знак, слово в алгебре имеет вполне определенное значение и не терпит произвольного толкования и применения.
3. К сожалению, в различных научных школах иногда имеют место различия в терминологии, поэтому, приступая к чтению научного труда, целесообразно уяснить терминологию его автора.

2. ТЕОРЕТИКО-ВЕРОЯТНОСТНЫЕ ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ. СЛУЧАЙНЫЕ СОБЫТИЯ

Окружающий нас мир полон таких явлений, которые имеют случайный характер[3]. Например:

– результат бросания монеты;

– бросание игральной кости;

– попадание в мишень;

– при проверке деталь оказалась годной/бракованной;

– при вытаскивании карандаша из коробки попался красный карандаш;

– после оплаты проезда в автобусе пассажиру достался «счастливый» билет;

– из колоды вытащили карту бубновой масти;

– выпадение снега.

При многократном выполнении эксперимента даже при одинаковых начальных условиях каждый раз его результат оказывается случайным.

Здесь случайным будет *наступление того или иного события*.

Еще:

– сколько лет проживет только что родившийся человек?

– сколько времени проработает новое техническое устройство?

– сколько попаданий в мишень будет при стрельбе?

Здесь *случайным будет уже число*.

Но если эксперимент проводить очень большое количество раз, то в некоторых случаях можно обнаружить закономерности, свойствен-

[3] Письменный Д. Т. Конспект лекций по теории вероятностей, математической статистике и случайным процессам. – М. : АЙРЕСС ПРЕСС 2010. – 288 с.

ные именно массовым случайным явлениям. А в некоторых случаях даже при массовых экспериментах закономерности не проявляются.

Стохастическими называются случайные события, в которых при массовых экспериментах проявляются закономерности. Предметом исследования теории вероятностей (ТВ) являются именно эти закономерности.

Теория вероятностей – математическая наука, изучающая закономерности, присущие массовым случайным явлениям.

Методы ТВ не дают возможности предсказывать, к чему приведет отдельное случайное явление, но дают возможность *предсказать средний суммарный результат массы однородных случайных явлений.*

Например: при отдельном бросании монеты нельзя указать, какой будет результат, при нескольких бросаниях закономерности тоже не будет просматриваться, а при большом количестве экспериментов орел и решка будут появляться примерно одинаковое количество раз. И это будет уже проявляться закономерность.

Цель ТВ – осуществление прогноза в области случайных явлений, влияние на ход этих явлений, их контроль.

2.1. СЛУЧАЙНЫЕ СОБЫТИЯ

2.1.1. ОПРЕДЕЛЕНИЕ СЛУЧАЙНОГО СОБЫТИЯ

Случайный эксперимент (СЭ) (опыт, испытание, наблюдение) – это реализация той или иной модели случайного явления.

Результат случайного эксперимента – *случайный исход*.

В результате СЭ может появиться только один из исходов. Поэтому говорят, что *исходы взаимоисключающие.*

Различают составные и элементарные исходы.

Элементарный исход – неразложимый исход.

Множество взаимоисключающих исходов называется *пространством исходов*. В частности, основным является пространство *элементарных исходов* Ω.

Это пространство может быть как с *конечным количеством исходов*, так и с *бесконечным*.

Может случиться так, что некоторые исходы для исследователя равноценны, равнозначны. Например: попадание в любую точку мишени приводит к ее поражению.

Тогда можно говорить о *случайном событии* (СС) «поражение цели».

Здесь исходов бесконечное количество, а событий – два.

Тогда все множество исходов делится на подмножества, приводящие к тем или иным СС.

То есть СС *отождествляется с подмножеством исходов, к нему приводящих*[4].

Еще пример: бросание двух игральных костей. СС здесь могут быть – выпадение конкретного суммарного значения очков (пример – 5 очков из исходов 4-1, 3-2, 2-3, 1-4), или превышение некоторого суммарного значения очков, или выпадения двух одинаковых значений и т.п.

Еще пример. В корзине несколько черных и несколько белых шаров. Выбор любого из белых шаров есть событие – белый шар.

В терминологии теории множеств:

– *пространство* Ω *элементарных исходов* – это универсальное множество;

– *элементарный исход* – это элемент этого множества Ω;

– *случайное событие* – подмножество множества Ω.

Если СС есть пустое множество, то это событие – *невозможное*.

Если СС – это множество тождественно множеству исходов Ω, то событие – *достоверное* (дальше через Ω будем обозначать достоверное событие).

Если событие не достоверное и не невозможное, оно *возможное, вероятное.*

---

[4] В некоторых источниках не различают исходы из одного подмножества и говорят, что СЭ непосредственно приводит к СС. т.е. результатом СЭ является не исход, а СС.

2.1.2. ОСНОВНЫЕ ОПЕРАЦИИ НАД СОБЫТИЯМИ

Поскольку события – множества, то к ним применимы операции алгебры множеств. Но поскольку в данном случае множества имеют интерпретацию событий, операции тоже должны иметь соответствующую интерпретацию, т.е. существует алгебра событий: $A(2^\Omega, S)$, где 2^Ω – носитель, S – сигнатура.

События и действия над ними можно наглядно иллюстрировать с помощью диаграмм Эйлера–Венна.

Достоверное событие Ω обозначается прямоугольником, элементарные исходы – точками прямоугольника, СС – областью внутри прямоугольника.

1. **Сумма событий** (говорят, что произошло хотя бы одно из событий или оба)

$$C = A + B.$$

Соответствует *объединению* соответствующих подмножеств исходов

$$\Omega_C = \Omega_A \cup \Omega_B.$$

2. **Произведение событий** (говорят, что произошли оба события).

$$C = A \cdot B$$

(иногда пишут $A \times B$, но не советую, можно спутать с декартовым произведением множеств).

Произведение событий соответствует *пересечению* соответствующих подмножеств исходов.

$$\Omega_C = \Omega_A \cap \Omega_B.$$

Если $\Omega_C = \varnothing$, то события *несовместны*.

Совокупность всех несовместных событий (т.е. их сумма – достоверное событие, а множество исходов равно Ω) называется *полной группой несовместных событий* (или реже – *полный набор несовместных событий*) – это разбиение множества – пространства элементарных исходов.

3. **Разность событий.**

$$C = A - B \ (\Omega_C = \Omega_A \setminus \Omega_B)$$

происходит тогда и только тогда, когда происходит событие A и не происходит событие B.

4. **Противоположное событие.** Пусть для события A множество исходов Ω_A. Тогда событие B, множество исходов которого $\Omega_B = \overline{\Omega_A}$, называется *противоположным событием для* A.

$\neg A$ означает, что событие не наступило.

5. Событие A *влечет* событие B $(\Omega_A \subseteq \Omega_B)$ (A является частным случаем события B), если из того, что происходит событие A следует, что происходит и событие B.

Верно ли обратное утверждение?

6. Если $\Omega_A \subseteq \Omega_B$ и $\Omega_B \subseteq \Omega_A$, то события называются равными.

Система тождеств алгебры событий

1	$A + B = B + A$	1’	$A \cdot B = B \cdot A$
2	$A + B + C = (A + B) + C$	2’	$A \cdot B \cdot C = (A \cdot B) \cdot C$
3	$A + (B \cdot C) = (A + B) \cdot (A + C)$	3’	$A \cdot (B + C) = (A \cdot B) + (A \cdot C)$
4	$A + \varnothing = A$	4’	$A \cdot \Omega = A$
5	$A + \neg A = \Omega$	5’	$A \cdot \neg A = \varnothing$
6	$A + \Omega = \Omega$	6’	$A \cdot \varnothing = \varnothing$
7	$A + A = A$	7’	$A \cdot A = A$
8	$A + (A \cdot B) = A$	8’	$A \cdot (A + B) = A$
9	$(A \cdot B) + (A \cdot \neg B) = A$	9’	$(A + B) \cdot (A + \neg B) = A$
10	$\neg(A + B) = \neg A \cdot \neg B$	10’	$\neg(A \cdot B) = \neg A + \neg B$
11	$\neg\varnothing = \Omega$	11’	$\neg\, \Omega = \varnothing$
12	$\neg\neg A = A$	12’	

Задача.

Дано. Система защиты информации (СЗИ) состоит из 6 элементов. Пусть случайные события A_i, $i = 1…6$ состоят в том, что злоумышленник преодолел соответствующий элемент СЗИ. Выразить через A событие, состоящее в том, что злоумышленник преодолел СЗИ в целом.

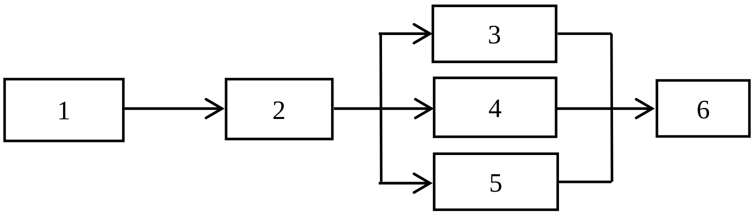


Рис. 2.1. Иллюстрация к задаче: структура системы защиты информации

Решение.

$$A = A_1 \cdot A_2 \cdot (A_3 + A_4 + A_5) \cdot A_6.$$

2.2. ПОНЯТИЕ И СВОЙСТВА ВЕРОЯТНОСТЕЙ

2.2.1. СВОЙСТВО СТАТИСТИЧЕСКОЙ УСТОЙЧИВОСТИ ОТНОСИТЕЛЬНОЙ ЧАСТОТЫ СОБЫТИЯ

Пусть в n повторяющихся опытах некоторое событие A наступило n_A раз.

Число n_A называется *частотой события A*, а отношение

$$\frac{n_A}{n} = P^*(A)$$

называется *относительной частотой* (или *частостью*) события A в рассматриваемой серии опытов.

Относительная частота события обладает следующими свойствами:

1. Частость любого события заключена между нулем и единицей

$$0 \le P^*(A) \le 1.$$

2. Частость невозможного события равна нулю

$$P^*(\varnothing)=0.$$

3. Частость достоверного события равна единице

$$P^*(\Omega)=1.$$

4. Частость суммы двух *несовместных* событий равна сумме частостей этих событий, т.е. если $A \cdot B = \varnothing$, то

$$P^*(A+B)=P^*(A)+P^*(B).$$

А если они совместны?

Частость обладает еще одним важнейшим фундаментальным свойством, называемым *свойством статистической устойчивости*: *с увеличением количества опытов* (n) *она принимает значения, близкие к некоторому постоянному числу.*

Говорят, что частотность стабилизируется.

Теория вероятности изучает только те массовые случайные явления с неопределенным исходом, для которых предполагается наличие устойчивости относительной частоты – т.е. СТОХАСТИЧЕСКИЕ случайные явления.

2.2.2. ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ

Каждое случайное событие обладает той или иной степенью возможности (невозможности).

Количественная мера возможности наступления случайного события называется **вероятность** (стохастическая и нестохастическая).

Определения опираются на принцип равновозможности элементарных исходов.

Статистическое определение вероятности

Вероятность наступления события A – предел частости события A при количестве опытов n, стремящемся к бесконечности.

$$P(A) = \lim_{n\to\infty} P^*(A) = \lim_{n\to\infty} \frac{n_A}{n}.$$

Классическое определение вероятности

Пусть:

1. card (Ω) = n и *все элементарные исходы равновозможные*.

2. card (Ω_A) = m.

Тогда

$$P(A) = \frac{m}{n}.$$

Геометрическое определение вероятности

Все исходы из Ω *равновозможны*.

Пусть:

1. Ω – замкнутая область на плоскости, card (Ω) = ∞ = continuum, и площадь этой области $S(\Omega) = S_\Omega$.

2. Ω_A – замкнутая область на плоскости, $\Omega_A \subseteq \Omega$, $S(\Omega_A) = S_A$.

Тогда

$$P(A) = \frac{S_A}{S_\Omega}.$$

2.2.3. СВОЙСТВА ВЕРОЯТНОСТИ

Вероятность события обладает следующими свойствами:

1. Вероятность любого события заключена между нулем и единицей

$$0 \le P(A) \le 1.$$

2. Вероятность невозможного события равна нулю

$$P(\varnothing) = 0.$$

3. Вероятность достоверного события равна единице

$$P(\Omega) = 1.$$

4. Вероятность суммы двух событий равна сумме вероятностей этих событий минус вероятность их совместного наступления

$$P(A + B) = P(A) + P(B) - P(A \cdot B).$$

4’. Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий, т.е. если $A \cdot B = \varnothing$, то

$$P(A + B) = P(A) + P(B).$$

4’’. Вероятность суммы трех событий равна

$$P(A + B + C) = P(A) + P(B) + P(C) -$$
$$- P(A \cdot B) - P(A \cdot C) - P(B \cdot C) + P(A \cdot B \cdot C).$$

5. Вероятность события, противоположного событию A, равна

$$P(\overline{A}) = 1 - P(A).$$

6. Сумма вероятностей полной группы несовместных событий $\Omega = \{A_1, A_2, \ldots, A_i, \ldots, A_n, \forall\ i \neq j\ A_i \cdot A_j = \varnothing, A_1 + A_2 + \ldots + A_i + \ldots + A_n = \Omega\}$ равна единице

$$\sum_{i=1}^{n} P(A_i) = 1.$$

Задачи.

1. Найти вероятность того, что в наудачу написанном двузначном числе цифры разные[5].

[5] Всего двузначных чисел 90 (с 10 по 99). Из них с одинаковыми цифрами 9. Тогда 1 – 9/90.

2. Набирая номер телефона, абонент забыл 2 последние цифры и набрал их наугад. Какова вероятность того, что он набрал правильный номер[6].

2.2.4. УСЛОВНАЯ ВЕРОЯТНОСТЬ

Понятие и выражение для вычисления условной вероятности

Рассмотрим вопрос о том, как вычислить вероятность $P(B/A)$ события B при том условии, что в этом же испытании уже наступило событие A?

Пример: брошен игральный кубик и результат нам не известен, но известно, что выпало четное число (событие A). Определить, зная этот факт, вероятность того, что выпало число больше «3»?

$$A = \{2, 4, 6\},$$

$$B = \{4, 5, 6\},$$

$$B/A = \{4, 6\},$$

$$P(B/A) = 2/3.$$

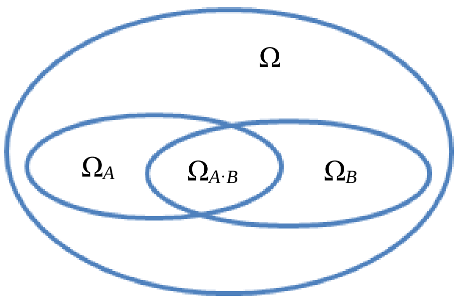


Рис. 2.2. Иллюстрация к определению вероятности произведения событий

[6] 1/100.

Выведем формулу для вычисления условной вероятности.

$$P(A)=\frac{S(\Omega_A)}{S(\Omega)}\Rightarrow S(\Omega_A)=P(A)\cdot S(\Omega).$$

Но $\Omega_{B/A}$ это по сути $\Omega_{A\cdot B}$.

Если событие A наступило (а именно этот случай и рассматривается), то для события B/A пространство Ω_A будет *пространством элементарных исходов* и в соответствии с геометрическим определением вероятности

$$P(B/A)=\frac{S(\Omega_{A\cdot B})}{S(\Omega_A)}.$$

Проведем в этом выражении замены

$$P(B/A)=\frac{S(\Omega_{A\cdot B})}{S(\Omega_A)}=\frac{S(\Omega_{A\cdot B})}{P(A)\cdot S(\Omega)}=\frac{P(A\cdot B)}{P(A)}.$$

Итак

$$P(B/A)=\frac{P(A\cdot B)}{P(A)}. \tag{2.1}$$

Аналогично через классическое определение вероятности:

$$P(B/A)=\frac{\mathrm{card}(\Omega_{A\cdot B})}{\mathrm{card}(\Omega_A)}.$$

Поделим числитель и знаменатель на $\mathrm{card}(\Omega)$

$$P(B/A)=\frac{\dfrac{\mathrm{card}(\Omega_{A\cdot B})}{\mathrm{card}(\Omega)}}{\dfrac{\mathrm{card}(\Omega_A)}{\mathrm{card}(\Omega)}}=\frac{P(A\cdot B)}{P(A)}.$$

Вопрос: Равны ли между собой $P(B/A)$ и $P(A/B)$?

Чему равна условная вероятность $P(B/A)$, если события несовместны[7]?

Задание. Показать на диаграмме Венна.

Свойства условной вероятности

1. $P(\Omega/A) = 1$ «Вероятность того, что произойдет достоверное событие, при условии, что наступило событие A».

2. $P(\varnothing/A) = 0$ «Вероятность того, что произойдет невозможное событие, при условии, что наступило событие A».

3. $0 \le P(B/A) \le 1$.

4. $P(\overline{B}/A) = 1 - P(B/A)$.

5. $P((A+B)/C) = P(A/C) + P(B/C) - P((A \cdot B)/C)$.

Вероятность произведения событий

Выражение для вычисления вероятности произведения событий непосредственно вытекает из (2.1)

$$P(A \cdot B) = P(A) \cdot P(B/A). \tag{2.2}$$

Для трех событий

$$P(A \cdot B \cdot C) = P(A) \cdot P(B/A) \cdot P(C/(A \cdot B)). \tag{2.3}$$

Вопросы:

1. Чему равны условная вероятность $P(B/A)$ и $P(A \cdot B)$, если $\Omega_A \subseteq \Omega_B$[8]?

2. Чему равны условная вероятность $P(B/A)$ и $P(A \cdot B)$, если $\Omega_A \supseteq \Omega_B$[9]?

---

[7] $P(B/A) = 0$

[8] $P(B/A) = 1$, $P(A \cdot B) = 1 \cdot P(A) = P(A)$.

Пример.

В корзине лежат три белых, три черных и три красных шара.

Наугад вынимают один за другим три шара.

Какова вероятность того, что все вынутые шары одного цвета?

Решение.

Первый шар может быть любого цвета (событие A и это событие достоверное, т.е. $P(A) = 1$).

Второй шар должен быть того же цвета $P(B/A) = 2/8$.

Третий тоже того же цвета $P(C/(B \cdot A)) = 1/7$.

А все вместе события $D = A \cdot (B/A) \cdot (C/(A \cdot B))$ произойдут с вероятностью, равной их произведению

$P(D) = P(A) \cdot P(B/A) \cdot P(C/(A \cdot B)) = 1 \cdot 2/8 \cdot 1/7 = 1/28$.

Вопросы:

Какова вероятность, что это будут красные шары?

Какова вероятность что это будут не красные шары?

Какова вероятность вытянуть все шары разного цвета?

$$\frac{9}{9} \cdot \frac{8-2}{8} \cdot \frac{7-2-2}{7}\,.$$

Независимые события

События A и B называются независимыми, если вероятность каждого из них не зависит от того, произошло ли другое или нет.

Для них

$$P(B/A) = P(B) \text{ и } P(A/B = P(A).$$

[9] $P(B/A) = P(B) / P(A)$, $P(A \cdot B) = (P(B) / P(A)) \cdot P(A) = P(B)$.

2.3. ОСНОВНЫЕ ФОРМУЛЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Одним из следствий совместного применения теорем сложения и умножения вероятностей являются формулы: полной вероятности, Байеса и Бернулли.

Напомним, что события H_1, H_2, …, H_n образуют полную группу несовместных событий, если $\forall i \neq j\ H_i \cdot H_j = \varnothing$ и

$$\sum_{i=1}^{n} H_i = \Omega.$$

Систему таких событий называют также *разбиением*, а сами события – *гипотезами.*

2.3.1. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ

Пусть $H = \{H_1, H_2, \ldots, H_n\}$ – разбиение.

Рассмотрим вероятность $P(A/H_i)$ события A, которое может наступить при условии наступления одного из несовместных событий из разбиения H.

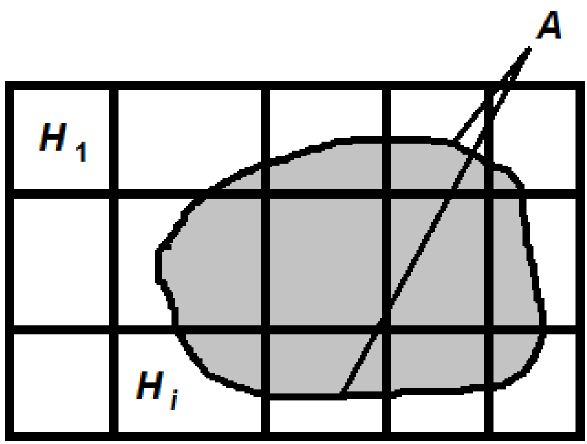


Рис. 2.3. Иллюстрация к формуле полной вероятности

Предполагается, что условная вероятность $P(A/H_i)$ наступления события A при наступлении события H_i может быть равна 0, 1, или находиться между ними:

$$\forall i 0 \le P(A/H_i) \le 1, \quad P(A/H_i) = \frac{S_{(A \cdot H_i)}}{S_{H_i}}.$$

То есть при наступлении каких-то H_i событие A становится невозможным, при каких-то достоверным, а при каких-то просто возможным.

Какова же будет вероятность наступления события A, если известны вероятности наступления гипотез $P(H_i)$ и условные вероятности $P(A/H_i)$.

Теорема 2.1. Вероятность события A, которое может наступить лишь при условии появления одного из попарно независимых событий H_1, H_2, …, H_n, образующих полную группу, равна сумме произведений вероятностей каждого из этих событий на соответствующую условную вероятность события A:

$$P(A) = P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) + \cdots + P(H_n) \cdot P(A/H_n) =$$

$$= \sum_{i=1}^{n} P(H_i) \cdot P(A/H_i). \tag{2.4}$$

Пример 2.1.

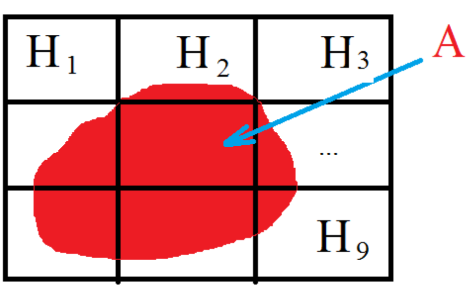


Рис. 2.4. Иллюстрация к формуле полной вероятности

Пример 2.2. В первой коробке находятся 10 патронов, 4 из которых являются холостыми, во второй коробке – 15 патронов, из которых 3 холостых. Найти вероятность того, что взятый наудачу патрон будет холостым.

Решение. Событие A – то, что взятый наудачу патрон будет холостым – может наступить при условии появления двух гипотез: патрон берется из первой коробки (обозначим эту гипотезу через H_1), патрон берется из второй коробки (H_2). Вероятность этих гипотез одинакова $P(H_1) = P(H_2) = 1/2 = 0{,}5$. Условная вероятность того, что из первой коробки будет извлечен холостой патрон $P(A/H_1) = 4/10 = 0{,}4$. Условная вероятность того, что холостой патрон будет извлечен из второй коробки $P(A/H_2) = 3/15 = 0{,}2$.

Вероятность события A рассчитывается по формуле полной вероятности:

$$P(A) = P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) =$$
$$= 0{,}5 \cdot 0{,}4 + 0{,}5 \cdot 0{,}2 = 0{,}2 + 0{,}1 = 0{,}3.$$

Пример 2.3. В сборочный цех завода поступает 40% деталей из I цеха и 60% – из II цеха. В I цехе производится 90% стандартных деталей, а во II – 95%. Найти вероятность того, что наудачу взятая сборщиком деталь окажется стандартной.

Решение. Взятие детали можно разбить на два этапа. Первый – это выбор цеха. Имеются две гипотезы: H_1 – деталь изготовлена I цехом, H_2 – II цехом. Второй этап – взятие детали. Событие A – взятая наудачу деталь стандартна. Очевидно, события H_1 и H_2 образуют полную группу, $P(H_1) = 0{,}4$, $P(H_2) = 0{,}6$. Числа 0,90 и 0,95 являются условными вероятностями события A при условии гипотез H_1 и H_2 соответственно, т.е. $P(A/H_1) = 0{,}90$ и $P(A/H_2) = 0{,}95$. По формуле (2.4) находим

$$P(A) = P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) = 0{,}4 \cdot 0{,}9 + 0{,}6 \cdot 0{,}95 = 0{,}93.$$

2.3.2. ФОРМУЛА БАЙЕСА

Следствием формулы (2.1) является формула Байеса или теорема гипотез. Она позволяет переоценить вероятности $P(H_i)$ гипотез H_i, принятых до опыта и называемых *априорными* («a priori», доопытные, лат.) по результатам проведенного опыта (в котором событие A произошло), т.е. найти условные вероятности $P(H_i/A)$, которые называют *апостериорными* («a posteriori», послеопытные).

До опыта	после опыта
$P(H_i)$	$P(H_i/A)$

Теорема 2.2. Пусть события $H_1, H_2, \ldots, H_n$ образуют полную группу событий (являются разбиением). Тогда условная вероятность события $H_i \quad (i=\overline{1,n})$ при условии, что событие A *произошло*, задается формулой

$$P(H_i/A)=\frac{P(H_i)\cdot P(A/H_i)}{P(A)}, \tag{2.5}$$

где $P(A)=\sum\limits_{i=1}^{n}P(H_i)\cdot P(A/H_i)$.

Выражение (2.5) называют формулой Байеса[10].

Доказательство.

Из

$$P(A\cdot H_i)=P(A)\cdot P(H_i/A)=P(H_i)\cdot P(A/H_i)$$

следует

$$P(H_i/A)=\frac{P(A\cdot H_i)}{P(A)}=\frac{P(H_i)\cdot P(A/H_i)}{P(A)}.$$

Пример 2.4. Рассчитать вероятности гипотез из примера 2.1.

---

[10] То́мас Ба́йес (Бейес, англ. Reverend Thomas Bayes) (1702 – 17 апреля 1761) – английский математик и пресвитерианский священник, член Лондонского королевского общества (1742).

Пример 2.5. Из примера 2.2: В первой коробке находятся 10 патронов, 4 из которых являются холостыми, во второй коробке – 15 патронов, из которых 3 холостых. Найти вероятность того, что патрон был взят из первой коробки, если он оказался холостым.

Решение. Если вытащенный патрон оказался холостым, то вероятность того, что он был взят из первой коробки $P(H_1/A)$ определяется по формуле Байеса:

$$P(H_1/A)=\frac{P(H_1)\cdot P(A/H_1)}{P(A)}=\frac{\frac{1}{2}\cdot\frac{2}{5}}{\frac{3}{10}}=\frac{\frac{1}{5}}{\frac{3}{10}}=\frac{2}{3}.$$

Пример 2.6. Из примера 2.3: В сборочный цех завода поступает 40% деталей из I цеха и 60% – из II цеха. В I цехе производится 90% стандартных деталей, а во II – 95%. Найти вероятность того, что наудачу взятая сборщиком деталь окажется сделанной во втором цехе, если она оказалась стандартной.

Решение. Определим вероятность гипотезы H_2 при условии, что событие A (взятая деталь стандартна) уже произошло, т.е. $P(H_2/A)$:

$$P(H_2/A)=\frac{P(H_2)\cdot P(A/H_2)}{P(A)}=\frac{0{,}6\cdot 0{,}95}{0{,}93}\approx 0{,}613.$$

2.3.3. ФОРМУЛА БЕРНУЛЛИ

Схема Бернулли

С понятием «независимых событий» связано понятие «независимых испытаний (опытов)».

Несколько опытов называются независимыми, если их исходы представляют собой независимые события (независимые в совокупности).

Другими словами, если проводится несколько испытаний, т.е. опыт выполняется при данном комплексе условий многократно (такое явление называется «последовательностью испытаний»), причем вероятность наступления некоторого события A в каждом испытании не зависит от исходов других испытаний, то такие испытания называются *независимыми*.

Примерами независимых испытаний могут служить: несколько (n раз) подбрасываний монеты; стрельба (n раз) по мишени без поправок на ранее допущенную ошибку при новом выстреле; несколько (n раз) выниманий из урны одинаковых на ощупь занумерованных шаров, если шары каждый раз после просмотра возвращаются назад в урну, и т.д.

При практическом применении теории вероятностей часто используется стандартная схема, называемая *схемой Бернулли* или схемой независимых испытаний.

Последовательность n независимых испытаний, в каждом из которых может произойти некоторое событие A (его называют успехом) с вероятностью $P(A) = p$ или противоположное ему событие $\neg A$ (его называют неудачей) с вероятностью $P(\neg A) = q = 1 - p$, называется *схемой Бернулли*[11].

Например, при стрельбе по мишени: событие A – попадание (успех), событие $\neg A$ – промах (неудача); при обследовании n изделий на предмет годности: событие A – деталь годная (успех), событие $\neg A$ – деталь бракованная (неудача) и т.д.

В каждом таком опыте Ω состоит только из двух элементарных событий (исходов), т.е.

$$\Omega = \{\omega_1, \omega_2\},$$

где ω_1 – неудача; ω_2 – успех.

При этом

$$A = \{\omega_2\}, \neg A = \{\omega_1\}.$$

[11] Я́коб Берну́лли (нем. Jakob Bernoulli, 27 декабря 1654, Базель, – 16 августа 1705, там же) – швейцарский математик, профессор математики Базельского университета (с 1687 года). Один из основателей теории вероятностей и математического анализа. Иностранный член Парижской Академии наук (1699) и Берлинской академии наук (1701).

Вероятности этих событий обозначают через p и q соответственно ($p + q = 1$).

Множество Ω элементарных исходов для n опытов состоит из 2^n элементов. Например, при $n = 3$, т.е. при трехкратном повторении опыта возможны следующие исходы

$$\Omega = \{\bar{A}\bar{A}\bar{A}, \bar{A}\bar{A}A, \bar{A}A\bar{A}, A\bar{A}\bar{A}, \ldots, AAA\}.$$

Вероятность каждого исхода определяется однозначно, например

$$P(\bar{A}\bar{A}A) = qqp.$$

В общем случае, успеху сопоставляют число 1, неудаче – число 0. Элементарным событием для n опытов будет последовательность из n нулей и единиц. Тройка чисел (0, 0, 0) означает, что во всех трех опытах событие A не наступило; тройка чисел (0, 1, 0) означает, что событие A наступило во 2-м опыте, а в 1-м и 3-м не наступило.

Формула Бернулли

Простейшая задача, относящаяся к схеме Бернулли, состоит в определении вероятности того, что в n независимых испытаниях событие A наступит m раз ($0 \le m \le n$). Обозначается искомая вероятность так: $P_n(m)$ или $P_{n,\,m}$.

Например, при бросании игральной кости 3 раза $P_3(2)$ означает вероятность того, что в 3-х опытах событие A – выпадение цифры 4 (можно любой другой цифры) – произойдет 2 раза.

Очевидно, это $\{H_6, H_5, H_3\} = \{(A, A, \bar{A}); (A, \bar{A}, A); (\bar{A}, A, A)\} = \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ из $\{H_0, \ldots, H_7\} = \{(0, 0, 0), \ldots, (1, 1, 1)\}$ с вероятностями ppq, pqp, qpp. Все эти вероятности равны p^2q.

Причем

$$P(A/H_6) = P(A/H_5) = P(A/H_3) = 1;$$

$$P(A/H_0) = P(A/H_1) = P(A/H_2) = P(A/H_4) = P(A/H_7) = 0.$$

Тогда по (3.1)

$P(A) = P(H_0) \cdot P(A/H_0) + \ldots + P(H_7) \cdot P(A/H_7) =$

$= P(H_0) \cdot 0 + P(H_1) \cdot 0 + P(H_2) \cdot 0 + P(H_3) \cdot 1 + P(H_4) \cdot 0 + P(H_5) \cdot 1 + P(H_6) \cdot 1 +$

$+ P(H_7) \cdot 0 = P(H_6) + P(H_5) + P(H_3) = ppq + pqp + qpp =$

$= 3 \cdot p^2q = 3 \cdot \left(\frac{1}{6}\right)^2 \cdot \frac{5}{6} = 0{,}069.$

Но выделенное $3 \cdot p^2q = C_3^2 \cdot p^2 \cdot q^{3-2}$.

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Это наблюдение истинно для любого $P_n(m)$ и можно доказать следующую теорему.

Теорема 2.3. Если производится n независимых испытаний, в каждом из которых вероятность появления события A равна p, а вероятность его непоявления равна $q = 1 - p$, то вероятность того, что событие A произойдет m раз, определяется *формулой Бернулли*

$$P_n(m) = C_n^m \cdot p^m \cdot q^{n-m}. \qquad (2.6)$$

Доказательство.

Вероятность одного сложного события, состоящего в том, что событие A в n независимых опытах появится m раз в первых m опытах и не появится $(n - m)$ раз в остальных опытах (это событие $\underbrace{A \cdot A \cdot A \cdot \ldots \cdot A}_{m \text{ раз}} \cdot \underbrace{\overline{A} \cdot \overline{A} \cdot \overline{A} \cdot \ldots \cdot \overline{A}}_{(n-m) \text{ раз}}$) по теореме умножения вероятностей равна $p^m \cdot q^{n-m}$. Вероятность появления события A снова n раз, но в другом порядке, например

$$\overline{A} \cdot \underbrace{A \cdot A \cdot A \cdot \ldots \cdot A}_{m \text{ раз}} \cdot \overline{A} \cdot \overline{A} \cdot \overline{A} \cdot \ldots \cdot \overline{A} \quad \text{или} \quad A\overline{A}A\overline{A} \cdot \ldots \cdot A\overline{A}$$

будет той же самой, т.е. $p^m \cdot q^{n-m}$.

Число таких сложных событий – в n опытах m раз встречается событие A в различном порядке – равно числу сочетаний из n по m, т.е. C_n^m .

Так как все эти сложные события несовместны, то по теореме сложения вероятностей искомая вероятность равна сумме вероятностей всех возможных сложных событий, т.е.

$$P_n(m) = \underbrace{p^m q^{n-m} + \cdots + p^m q^{n-m}}_{C_n^m \text{ слагаемых}} = C_n^m p^m q^{n-m}, \; m = 0, 1, 2, ..., n.$$

Пример. Производится 3 независимых выстрела по цели. Вероятности попадания при разных выстрелах одинаковы и равны p = 0,9.

Какова вероятность:

а) промаха;

б) одного попадания;

в) двух попаданий;

г) трех попаданий?

Теорема Пуассона

Использование формулы Бернулли (3.1) при больших значениях n и m вызывает большие трудности, так как это связано с громоздкими вычислениями. Так, при n = 200, m = 116, p = 0,72 формула Бернулли принимает вид

$$P_{200}(116) = C_{200}^{116} \cdot 0{,}72^{116} \cdot 0{,}28^{84}.$$

Подсчитать результат практически невозможно.

Вычисление $P_n(m)$ вызывает затруднения также *при малых значениях p и q*.

Возникает необходимость в отыскании приближенной формулы для вычисления $P_n(m)$, обеспечивающей необходимую точность.

Теорема 2.4. Если число испытаний неограниченно увеличивается ($n \to \infty$) и вероятность p наступления события A в каждом испытании неограниченно уменьшается ($p \to 0$), но так, что их произведение np является постоянной величиной ($np = a = \text{const}$), то вероятность $P_n(m)$ удовлетворяет предельному равенству

$$\lim_{n\to\infty} P_n(m) = \frac{a^m \cdot e^{-a}}{m!}.$$

Приведенное выражение называется *асимптотической формулой Пуассона*.

Но ею можно пользоваться, только если $p \to 0$, т.е. выше приведенный пример рассчитать по ней нельзя. Существует теорема (Локальная теорема Муавра–Лапласа):

Теорема 2.5. Если вероятность p наступления события A в каждом испытании постоянна и отлична от нуля и единицы, а число независимых испытаний достаточно велико, то вероятность $P_n(m)$ может быть вычислена по приближенной формуле

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ где } x = \frac{m - np}{\sqrt{npq}}.$$

Задание: проверить в маткаде!

2.4. АКСИОМАТИЧЕСКИЙ ПОДХОД К ТЕОРИИ ВЕРОЯТНОСТЕЙ

Приведенный материал «классической» теории вероятностей основан на «частотной» интерпретации вероятности. Ее основой является предположение о равновозможности элементарных исходов.

Советским математиком А. Н. Колмогоровым в 1933 г. предложен *алгебраический подход* к математическому аппарату теории вероятностей, основанный на *аксиоматике* абстрактных событий и *теории меры* (множеств).

Без привязки к частости событий и отсутствия требования равновозможности элементарных исходов!

Пусть Ω – множество элементов ω, которые называются элементарными исходами, а F – множество подмножеств Ω, называемых случайными событиями (или просто – событиями), а Ω – пространством элементарных исходов.

- ***Аксиома I*** (*алгебра событий*). F является алгеброй событий.
- ***Аксиома II*** (*существование вероятности событий*). Каждому событию x из F поставлено в соответствие неотрицательное вещественное число $P(x)$, которое называется вероятностью события x.
- ***Аксиома III*** (*нормировка вероятности*) $P(\Omega) = 1$.
- ***Аксиома IV*** (*аддитивность вероятности*). Если события x и y не пересекаются, то $P(x + y) = P(x) + (y)$.

Алгебра событий:

Пусть Ω – пространство элементарных исходов. Составим множество F из всех подмножеств Ω, т.е. булеан $F = 2^{\Omega}$. Пусть множество всех возможных событий F удовлетворяет следующим двум условиям:

1) если $A \subset F$ и $B \subset F \Rightarrow A \cup B \subset F, A \cap B \subset F$.
2) для любого события $A \in F$ имеет место включение $\bar{A} \in F$.

Класс F случайных событий, удовлетворяющих этим условиям, называется F-алгеброй событий.

Комментарий. В случае, если Ω конечно и содержит n элементарных исходов, F содержит 2^n событий, т.е. $|F| = 2^{|\Omega|}$.

Колмогоров А. Н. отказался от предположения равновозможности элементарных событий и распространил первое свойство F-алгебры на счетное число событий из σ-алгебры событий. Это дало ему возможность дать общее аксиоматическое определение вероятности события.

Свойства системы аксиом Колмогорова

1. Система аксиом Колмогорова непротиворечива, так как существуют реальные объекты, которые удовлетворяют одновременно всем аксиомам Колмогорова.

2. Система аксиом Колмогорова неполна. Это значит, что даже при одном множестве элементарных исходов Ω вероятности на множестве F могут быть выбраны многими различными способами.

Неполнота системы аксиом Колмогорова не является недостатком, а, наоборот, обеспечивает возможность ее широкого практического применения, так как позволяет в разных задачах рассматривать одинаковые множества случайных событий с различными вероятностями. Это можно проиллюстрировать известным **парадоксом Бертрана**[12]. Пусть для некоторой окружности случайным образом выбирается хорда. Найти вероятность того, что эта хорда длиннее стороны правильного треугольника, вписанного в данную окружность. Бертран утверждает, что эта вероятность определяется неоднозначно, т.е. различные методы приводят к разным результатам.

Первый метод:

Случайным образом (равномерно) в данном круге выбирается точка. Эта случайная точка определяет единственную хорду, серединой которой она является. Эта хорда длиннее стороны нашего правильного

[12] https://ru.wikipedia.org/wiki/%D0%9F%D0%B0%D1%80%D0%B0%D0%B4%D0%BE%D0%BA%D1%81_%D0%91%D0%B5%D1%80%D1%82%D1%80%D0%B0%D0%BD%D0%B0_(%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D1%8C)

треугольника тогда и только тогда, когда ее середина лежит внутри круга, вписанного в треугольник. Радиус этого круга равен половине радиуса исходного круга, следовательно, площадь вписанного круга составляет 1/4 площади исходного. Таким образом, вероятность того, что случайно выбранная точка лежит внутри вписанного круга, равна 1/4. Так что этот метод дает ответ 1/4.

Второй метод:

Исходя из соображений симметрии, можем считать, что одним концом хорды является произвольная фиксированная точка на окружности. Пусть этой точкой является вершина вписанного треугольника. Выберем другой конец случайно. Вершины треугольника делят окружность на три равные дуги, и случайная хорда длиннее стороны правильного треугольника, если она пересекает этот треугольник. Так что искомая вероятность теперь равна 1/3.

Получение разных результатов кажется парадоксальным, так как было убеждение, что слова «случайный выбор» однозначно определяют искомую вероятность. Парадокс показывает, что возможны различные способы выбора случайным образом, причем каждый способ выглядит по-своему «естественным». Фактически это означает, что в зависимости от того, что именно мы понимаем под словами «Случайным образом (равномерно)», вероятности могут быть выбраны многими различными способами.

3. ТЕОРЕТИКО-ВЕРОЯТНОСТНЫЕ ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

3.1. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Понятие случайной величины

Одним из важнейших понятий теории вероятностей (наряду со случайным событием и вероятностью) является понятие случайной величины[13].

Под *случайной величиной* понимают числовую величину, которая в результате опыта принимает то или иное значение, причем неизвестно заранее, какое именно.

Или: числовая величина, принимающая то или иное значение в результате реализации испытания случайным образом, называется *случайной величиной*.

Случайные величины (СВ) обозначаются прописными латинскими буквами X, Y, ... (или строчными греческими буквами ξ(кси), η(эта), θ(тэта), ψ(пси) и т.д.), а принимаемые ими значения соответственно малыми буквами x_1, x_2, x_3.

Например:

– при бросании игрального кубика выпадает число очков от 1 до 6;

– при взятии 6 карт из колоды можно получить от 0 до 4 тузов;

За определенный промежуток времени (скажем, день или месяц) в городе регистрируется то или иное количество преступлений, происходит какое-то количество дорожно-транспортных происшествий.

– рост случайно взятого человека.

– из орудия производится выстрел. Дальность полета снаряда принимает какое-либо значение случайным образом.

---

[13] Письменный Д. Т. Конспект лекций по теории вероятностей, математической статистике и случайным процессам. – М. : АЙРЕСС ПРЕСС 2010. – 288 с.

Типы случайных величин

Если случайная величина может принимать конечное или счетное множество значений, то она называется *дискретной* (дискретно распределенной).

Непрерывной случайной величиной называется такая случайная величина, которая может принимать все значения из некоторого конечного или бесконечного интервала на действительной числовой оси.

СВ может быть *скалярной* – одной или *векторной* (системой случайных величин) – когда имеет место совокупность нескольких СВ, совместно характеризующих рассматриваемое случайное явление (координаты точки попадания снаряда – случайный вектор).

Дадим теперь строгое определение СВ, исходя из теоретико-множественной трактовки основных понятий теории вероятностей.

Случайной величиной X называется функционал, определенный на пространстве элементарных исходов Ω, который каждому элементарному исходу ω ставит в соответствие число $X(\omega)$, т.е. $X = X(\omega)$, $\omega \in \Omega$ (или $X = f(\omega)$).

3.2. ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Если случайная величина может принимать конечное или счетное множество значений, то она называется *дискретной* (дискретно распределенной) (ДСВ).

Закон распределения дискретной СВ

Для задания случайной величины недостаточно перечислить ее всевозможные значения (даже если это возможно) потому, что вероятности, с которыми эти случайные величины принимают свои значения, будут совершенно разными.

Поэтому для задания дискретной случайной величины кроме перечня ее всех возможных значений нужно еще указать их вероятности.

Соответствие между возможными значениями случайной величины и их вероятностями называют *законом распределения дискретной случайной величины.*

Про СВ говорят, что «она подчиняется данному закону распределения».

Закон распределения можно задать в виде:

– таблицы,

– формулы,

– графически.

При табличном задании *закона распределения* в первой строке таблицы записываются возможные значения случайной величины, а во второй – соответствующие значениям вероятности:

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Такая таблица называется *рядом распределения дискретной случайной величины X*.

Так как случайная величина в результате испытания примет одно и только одно значение, то события: $X = x_1$, $X = x_2$, …, $X = x_n$ образуют полную группу несовместных событий. Следовательно, сумма вероятностей этих событий равна единице:

$$p_1 + p_2 + \cdots + p_n = \sum_{k=1}^{n} p_k = 1.$$

Для наглядности ряд распределения СВ можно изобразить графически. Для этого в прямоугольной системе координат по оси абсцисс *OX* будем откладывать значения случайной величины x_k, $k = 1, 2, \ldots, n$, а по оси ординат OY – соответствующие им вероятности p_k. Полученные точки соединяются отрезками прямых. Построенная таким образом фигура называется *многоугольником* (*полигоном*) *распределения*.

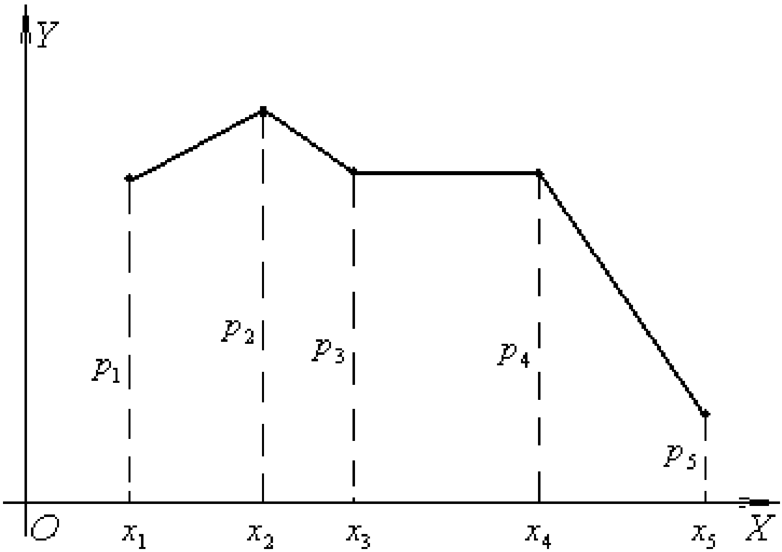


Рис. 3.1. Пример многоугольника распределения

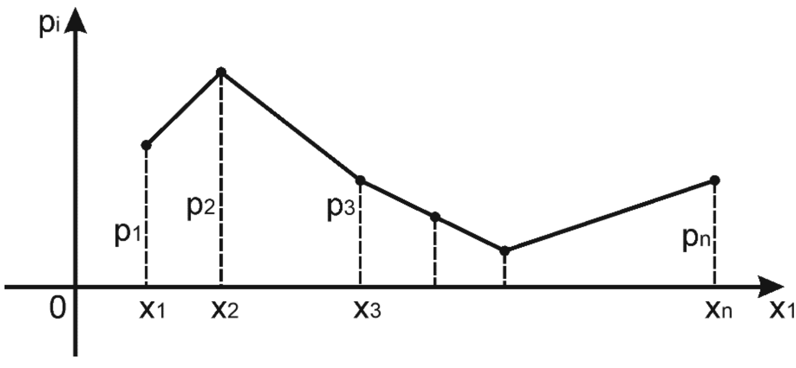


Рис. 3.2. Пример многоугольника распределения

Многоугольник распределения, так же как и ряд распределения, полностью характеризует случайную величину. Он является одной из форм закона распределения.

СВ задать можно и формулой

$$p_i = p(x_i).$$

Особенно это актуально, если ДСВ имеет счетное количество значений, так как ни таблицей, ни графически задать ее невозможно.

Пример. Для процессов гибели и размножения.

$$P_0 = \frac{1}{1+\sum_{j=1}^{\infty}\prod_{k=1}^{j}\frac{\lambda_{k-1}}{\mu_k}}; \quad P_j = P_0\prod_{k=1}^{j}\frac{\lambda_{k-1}}{\mu_k}, \forall j.$$

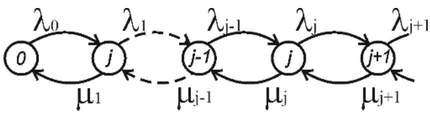
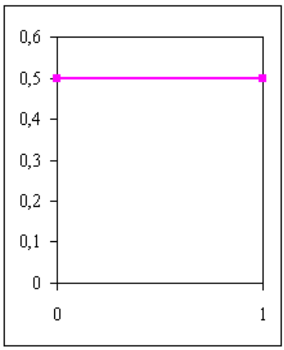


Рис. 3.3. Граф процесса «гибели и размножения»

Пример. Случайным образом бросается монета.

Построить ряд и многоугольник распределения числа выпавших гербов, написать формулу.

X	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$



Формула: $p_i = ½$,

Пример. Случайным образом бросаются две игральные кости.

Составить ряд распределения значений суммы чисел на верхних гранях игральных костей (2 – 15).

Операции над дискретными случайными величинами

Определим математические операции над дискретными СВ.

Суммой (*разностью*, *произведением*) ДСВ X, принимающей значения x_i с вероятностями $p_i = P\{X = x_i\}$, $i = 1, 2, ..., n$ и ДСВ Y, принимающей значения y_j с вероятностями $p_j = P\{Y = y_j\}$, $j = 1, 2, ..., m$ называется ДСВ $Z = X + Y$ ($Z = X - Y$, $Z = X \cdot Y$), принимающая значения $z_{ij} = x_i + y_j$ ($z_{ij} = x_i - y_j$, $z_{ij} = x_i \cdot y_j$) с вероятностями $p_{ij} = P\{X = x_i\ , Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j / X = x_i\}$ для всех указанных значений i и j.

В случае совпадения некоторых сумм $x_i + y_j$ (разностей $x_i - y_j$, произведений $x_i \cdot y_j$) соответствующие вероятности *складываются* (*всегда*, и для вычитаний и умножений).

Произведение ДСВ Y на число c называется ДСВ cX, принимающая значения cx_i с вероятностями $p_i = P\{X = x_i\}$.

Две ДСВ X и Y называются *независимыми*, если события $A_i = \{X = x_i\}$ и $B_i = \{Y = y_j\}$ независимы для любых $i = 1, 2, ..., n$; $j = 1, 2, ..., m$, т.е.

$$P\{X = x_i\,;\, Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j\}.$$

В противном случае СВ называются *зависимыми*. Несколько СВ называются *взаимно независимыми*, если закон распределения любой из них не зависит от того, какие возможные значения приняли остальные величины.

Пример.

$X =$	1	2	3
	0,1	0,4	0,5

$Y =$	2	3
	0,4	0,6

$Z = X + Y$	3	4	5	6
	0,04	0,22	0,44	0,3

3.3. ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Равномерный закон распределения

Случайная величина распределена по равномерному закону, если все ее значения равновероятны и вероятности равны $1/n$, где n – количество возможных значений ДСВ.

Биномиальный закон распределения

Пример. Монета бросается 4 раза. Построить многоугольник распределения СВ X – числа выпадений герба.

Пример. Стрелок производит три выстрела по мишени. Вероятность попадания при каждом выстреле равна 0,7. Построить ряд и многоугольник распределения числа попаданий в мишень.

Решение. Очевидно, данные примеры относятся к схеме Бернулли.

Ряд распределения для каждой задачи – это ряд вероятностей наступления события m раз из n опытов.

Каждая вероятность считается по формуле

$$P_n(m) = C_n^m \cdot p^m \cdot q^{n-m}, \tag{3.1}$$

где

$$C_n^m = \frac{n!}{m! \cdot (n-m)!}.$$

Биномиальным называется распределение вероятностей, определяемое формулой Бернулли (3.1).

По биномиальному закону распределена случайная величина X числа появлений события A при проведении n независимых испытаний, если вероятность появления события A в каждом испытании одинакова и равна p ($q = 1 - p$). В n испытаниях событие A может вообще не появиться, появиться 1 раз, 2 раза, 3 раза, …, n раз. Таким образом, возможные значения X таковы: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, …, $x_{n+1} = n$. А соответствующие им вероятности подсчитываются по формуле Бернулли (3.1). Ряд распределения в этом случае будет таким:

X	0	1	2	…	k	…	x_n
p	q^n	$C_n^1\ p^1 q^{n-1}$	$C_n^2\ p^2 q^{n-2}$	…	$C_n^k\ p^k q^{n-k}$	…	p^n

Геометрическое распределение

Пример. Стрелком производятся выстрелы по мишени *до первого попадания*. Вероятность попадания в каждом выстреле равна $p = 0{,}7$. Построить ряд распределения *количества произведенных выстрелов до первого попадания*.

Обозначим через X дискретную случайную величину – число произведенных выстрелов до первого попадания. Возможными значениями X являются натуральные числа $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, … . Множество значений X является бесконечным счетным множеством, как и ряд натуральных чисел.

Вероятность того, что случайная величина принимает значение $x_1 = 1$, т.е. попадание происходит при 1-м выстреле, по условию задачи равна $p_1 = p = 0{,}7$.

Вероятность принятия случайной величиной значения $x_2 = 2$ (попадание происходит при 2-м выстреле) подсчитывается как вероятность сложного события по теореме умножения вероятностей.

Непопадание в мишень в первом выстреле и попадание во втором – события независимые. Поэтому

$$p_2 = (1 - p)\cdot p = (1 - 0{,}7)\times 0{,}7 = 0{,}3\times 0{,}7 = 0{,}21.$$

Аналогичным образом находятся вероятность значения случайной величины $x_3 = 3$:

$$p_3 = (1 - p)\cdot(1 - p)\cdot p = (1 - 0{,}7)\times(1 - 0{,}7)\times 0{,}7 = 0{,}3\times 0{,}3\times 0{,}7 = 0{,}063 \text{ и т.д.}$$

Подсчитаем вероятность того, что случайная величина принимает значение $x_k = k$. Это означает, что попадание произошло лишь в k-м выстреле, а до этого был $k - 1$ промах. Так как все выстрелы независимы друг от друга, то по теореме умножения вероятностей $p_k = (1 - p)^{k-1}\cdot p$.

Ряд распределения будет характеризоваться уменьшением вероятности p_k с ростом k.

Можно заметить, что ряд вероятностей, соответствующих значениям случайной величины, образует бесконечную убывающую геометрическую прогрессию со знаменателем $q = 0{,}3$ и первым членом $p = 0{,}7$:

$$0{,}7;\ 0{,}7\times 0{,}3;\ 0{,}7\times 0{,}3^2;\ 0{,}7\times 0{,}3^3;\ 0{,}7\times 0{,}3^4;\ \ldots\ .$$

Поэтому распределение случайной величины в данном примере называют геометрическим.

Геометрическим называется распределение вероятностей случайной величины X, которое определяется следующим законом:

$$p_k = p \cdot q^{k-1},\ k \geq 1.$$

Здесь p – вероятность наступления события A, q – вероятность того, что событие A не произойдет: $q = 1 - p$. Случайная величина X определяется как число независимых испытаний, которые нужно произвести до первого появления события A.

Основные числовые характеристики дискретных случайных величин

Понятие числовой характеристики случайной величины

Закон распределения полностью характеризует случайную величину. Однако часто построение закона или ряда распределения представляет весьма трудоемкую задачу, либо закон распределения неизвестен вовсе.

На практике иногда бывает достаточно описать случайную величину «суммарно», *указав ее отдельные числовые параметры, до некоторой степени характеризующие существенные черты распределения случайной величины*.

К таким параметрам можно отнести *среднее значение*, около которого группируются возможные значения случайной величины; *число, характеризующее степень разбросанности значений случайной величины относительно среднего* и др.

Назначение таких характеристик – выразить компактно, в сжатой форме наиболее существенные особенности распределения. Все эти характеристики называются *числовыми характеристиками случайной величины*.

Числовые характеристики СВ не случайны! *Они закономерны*! Так, для полной характеристики успеваемости учащегося и прогнозирования получения им оценки в будущем можно построить ряд распределения его оценок. Однако достаточно часто успеваемость характеризуется лишь одной, средней оценкой.

Числовые характеристики играют большую роль в теории вероятностей, поскольку, оперируя ими, можно значительно упростить ряд практических вероятностных задач и получить важные результаты.

Например, в тех случаях, когда на численный результат эксперимента оказывают влияние отдельные случайные величины и их достаточно много, то *закон распределения результирующей случайной величины, оказывается, не будет зависеть от законов распределения составляющих величин*. В этих случаях для анализа результирующей величины необходимо лишь знать некоторые числовые характеристики отдельных случайных величин.

Рассмотрим наиболее важные числовые характеристики случайной величины.

Математическое ожидание дискретной случайной величины

Рассмотрим дискретную случайную величину X, принимающую значения $x_1, x_2, \dots, x_n$, с вероятностями $p_1, p_2, \dots, p_n$.

Сумма произведений всех возможных значений случайной величины на вероятности этих значений называется *математическим ожиданием* (МО) *случайной величины* и обозначается $M[X]$.

$$M[X] = x_1 \cdot p_1 + x_2 \cdot p_2 + \cdots x_n \cdot p_n = \sum_{k=1}^{n} x_k \cdot p_k. \qquad (3.2)$$

Пример. Найти МО ДСВ, ряд распределения которой задан в примерах:

1. Монета бросается 4 раза. Построить ряд распределения СВ X – числа выпадений герба.

2. Стрелок производит три выстрела по мишени. Вероятность попадания при каждом выстреле равна 0,7. Построить ряд распределения числа попаданий в мишень.

Вероятностный смысл этой числовой характеристики таков: математическое ожидание случайной величины приближенно равно среднему значению случайной величины. *Но это не означает, что это «среднее значение» будет появляться чаще всего*!!!

Математическое ожидание приближенно равно среднему арифметическому наблюдаемых значений случайной величины. Равенство будет тем точнее, чем больше число испытаний.

Математическое ожидание больше наименьшего и меньше наибольшего возможных значений. И может не совпадать ни с одним из значений ДСВ. Поэтому можно сказать, что математическое ожидание характеризует положение случайной величины на числовой оси, т.е. указывает некоторое среднее значение, около которого группируются все возможные значения случайной величины. *Такое среднее значение является «представителем» случайной величины и может замещать ее при грубых оценочных расчетах.*

Для системы стохастически взаимосвязанных СВ может быть рассчитано *условное МО* одной СВ при принятии конкретного значения другой СВ и функция МО одной СВ от значения другой СВ.

Свойства математического ожидания случайной величины:

1. Математическое ожидание постоянной величины равно самой величине:

$$M\ [C] = C.$$

2. Постоянный множитель можно выносить за знак математического ожидания:

$$M[C \cdot X] = C \cdot M[X].$$

3. Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий этих величин:

$$M[X + Y] = M[X] + M[Y].$$

4. Математическое ожидание произведения двух случайных величин равно произведению математических ожиданий этих величин плюс их ковариацию:

$$M[X \cdot Y] = M[X] \cdot M[Y] + \mathrm{cov}(X, Y),$$

$$\mathrm{cov}(X, Y) = M[(X - M[X]) \cdot (Y - M[Y])].$$

Математическое ожидание двух *независимых* случайных величин равно произведению математических ожиданий этих величин, так как их ковариация равна нулю:

$$M[X \cdot Y] = M[X] \cdot M[Y].$$

Две случайные величины называются независимыми, если закон распределения одной из них не зависит от того, какие возможные значения принимает другая величина.

Задача. Пусть имеется схема Бернулли с вероятностью наступления события p. Каково МО количества наступления событий при n опытах?

Математическое ожидание $M[X]$ числа появлений события A в n независимых испытаниях, в каждом из которых вероятность p появления события одинакова, равно произведению числа испытаний на вероятность появления события в одном испытании:

$$M_{б}[X] = np.$$

Задача. Для рассмотренных примеров бросания монеты 4 раза и трех выстрелов стрелка рассчитать МО.

Математическое ожидание геометрического распределения

$$M_{\text{г}}[X] = \frac{1}{p}.$$

Дисперсия дискретной случайной величины

Можно привести пример двух дискретных случайных величин X и Y, которые имеют различные возможные значения и при этом одинаковые математические ожидания. Рассмотрим следующие ряды распределения X и Y:

X	–100	100
p	0,5	0,5

Y	–1	1
p	0,5	0,5

Математические ожидания величин X и Y равны друг другу:

$$M[X] = -100\times0{,}5 + 100\times0{,}5 = -50 + 50 = 0,$$

$$M[Y] = -1\times0{,}5 + 1\times0{,}5 = -0{,}5 + 0{,}5 = 0.$$

Однако, возможные значения величин X и Y значительно отличаются. Таким образом, *зная лишь математическое ожидание случайной величины, нельзя судить ни о ее возможных значениях, ни о рассеянии значений около математического ожидания*.

Как можно задать величину разброса возможных значений величины?

На практике эта величина чрезвычайно важна. Например, ее необходимо знать, оценивая кучность поражения мишени при стрельбе из пистолета.

На первый взгляд, кажется, что необходимо проанализировать отклонение случайной величины от $M[X]$.

Отклонением называют разность между случайной величиной и ее математическим ожиданием: $X - M[X]$.

Отклонение само по себе является случайной величиной. Найдем ее МО

$$M[X - M[X]] = M[X] - M[M[X]] = M[X] - M[X] = 0.$$

То есть математическое ожидание отклонения случайной величины равно 0:

$$M[X - M[X]] = 0.$$

Это объясняется тем, что одни отклонения положительны, а другие – отрицательны. И в результате их взаимного сложения значение отклонение будет равно 0. Поэтому отклонение случайной величины нельзя использовать для оценки ее рассеяния. Для этого чаще всего вычисляют среднее значение квадрата отклонения, называемое *дисперсией*.

Дисперсией случайной величины называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D[X] = M[(X - M[X])^2]. \tag{3.3}$$

Рассмотрим дискретную случайную величину X, принимающую значения $x_1, x_2, \ldots, x_n$, с вероятностями $p_1, p_2, \ldots, p_n$.

Используя в выражении (3.3) определение математического ожидания (3.2), получим следующую формулу для вычисления дисперсии:

$$\begin{aligned} D[X] &= \sum\nolimits_{k=1}^{n} (x_k - M[X])^2 \cdot p_k = \\ &= (x_1 - M[X])^2 \cdot p_1 + (x_2 - M[X])^2 \cdot p_2 + \cdots + \\ &+ (x_n - M[X])^2 \cdot p_n. \end{aligned} \tag{3.4}$$

Раскроем квадрат разности в выражении (3.3)

$$D[X] = M[(X - M[X])^2] = M[X^2 - 2 \cdot X \cdot M[X] + M[X]^2].$$

Учитывая, что M[X] – это некоторое постоянное число, раскроем это равенство так:

$$D[X] = M[X^2] - M[2 \cdot X \cdot M[X]] + M[M[X]^2] =$$
$$= M[X^2] - 2 \cdot (M[X])^2 + (M[X])^2 = M[X^2] - (M[X])^2.$$

То есть

$$D[X] = M[X^2] - (M[X])^2. \qquad (3.5)$$

Или

$$D[X] = x_1^2 \cdot p_1 + x_2^2 \cdot p_2 + \ldots x_n^2 \cdot p_n - (M[X])^2.$$

Задача. Посчитать для ранее рассмотренных примеров дисперсию по выражению (3.5).

Свойства дисперсии случайной величины:

1. Дисперсия постоянной величины равна нулю:

$$D[C] = 0.$$

2. Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D[C \cdot X] = C^2 \cdot D[X].$$

Вопрос: как связаны $D[X]$ и $D[-X]$?

3. Дисперсия суммы двух *независимых* ($\mathrm{cov}(X, Y) = 0$) случайных величин равна сумме дисперсий этих величин:

$$D[X + Y] = D[X] + D[Y] + 2\mathrm{cov}(X, Y).$$

Вопрос: $D[X + C] = ?$, где $C = \mathrm{const}$.

4. Дисперсия разности двух независимых случайных величин равна сумме дисперсий этих величин:

$$D[X - Y] = D[X] + D[Y] - 2\mathrm{cov}(X, Y).$$

5. Дисперсия произведения двух *независимых* случайных величин равна:

$$D[XY] = D[X]D[Y] + m_x^2 D[Y] + m_y^2 D[X];$$

$$m_x = M[X]; \quad m_y = M[Y];$$

$$D[XY] = M[X^2] \cdot M[Y^2] - (M[X])^2 \cdot (M[Y])^2.$$

Если СВ независимы и еще $m_x = m_y = 0$, то $D[XY] = D[X] \cdot D[Y]$.

Дисперсия $D[X]$ числа появлений события A в n независимых испытаниях, в каждом из которых вероятность p появления события одинакова, равна произведению числа испытаний на вероятности появления и непоявления события в одном испытании:

$$D_б[X] = npq. \tag{3.6}$$

При каких значениях p дисперсия наибольшая?

Дисперсия геометрического распределения

$$D_г[X] = \frac{q}{p^2}.$$

Другие числовые характеристики случайной величины

Легко заметить, что в отличие от математического ожидания, *размерность дисперсии равна квадрату размерности случайной величины*.

Для характеристики рассеивания более удобно использовать другую величину, размерность которой совпадает с размерностью величины. Такой характеристикой является *среднее квадратическое отклонение*.

Средним квадратическим отклонением (СКО) случайной величины X называется квадратный корень из дисперсии:

$$\sigma(X) = \sqrt{D(X)}. \tag{3.7}$$

Мода M_0 дискретной случайной величины – это значение, обладающее наибольшей вероятностью

$$M_0 = \arg\max_{x \in X} P(x).$$

То есть мода ДСВ совпадает с одним или несколькими из ее значений.

Медианой M_P случайной величины называется такая величина, относительно которой равновероятно получение большего или меньшего значения случайной величины.

$$P(X > M_P) = P(X < M_P).$$

3.4. НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Понятие и закон распределения непрерывной случайной величины

Непрерывной случайной величиной (НСВ) называется такая случайная величина, которая может принимать все значения из некоторого конечного или бесконечного интервала на *действительной числовой оси*.

Исчерпывающей характеристикой *дискретной* случайной величины является *закон распределения дискретной случайной* величины – соответствие между возможными значениями случайной величины и их вероятностями. Он задается рядом распределения (таблично), полигоном распределения или формулой.

Но эти формы закона распределения не могут служить характеристикой непрерывной СВ. Таблично и полигоном нельзя задать НСВ потому, что у нее бесконечное количество значений, а кроме того, как будет показано ниже – *вероятность принятия НСВ конкретного значения равна* 0.

Определим вероятность того, что НСВ X примет значение, меньшее чем некоторое x. *Она будет некоторой неубывающей функцией от* x.

$$P(X < x) = F(x). \tag{3.8}$$

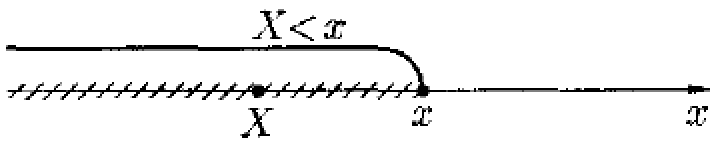


Рис. 3.4. Иллюстрация к определению функции распределения НСВ

Эта функция вероятности называется *функцией распределения НСВ* (ФР НСВ) и она является исчерпывающей характеристикой *непрерывной* случайной величины.

Рассмотрим интервал (a, b) в области значений НСВ X.

Появление НСВ в этом интервале означает реализацию случайного события $a \le X < b$. Оно является событием «*попадания НСВ в интервал* (a, b)».

Обозначим

B – событие, заключающееся в том, что $X < b$;

A – событие, заключающееся в том, что $X < a$;

C – событие, заключающееся в том, что $a \le X < b$.

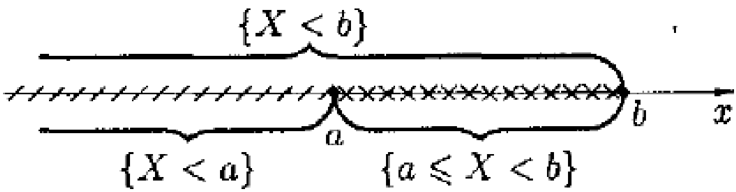


Рис. 3.5. Иллюстрация к выводу формулы вероятности принятия НСВ значения из заданного интервала

Тогда

$B = A + C$, т.е. B это – либо A либо C.

Поскольку события A = «$X < a$» и C = «$a \le X < b$» несовместны, то

$$P(B) = P(A) + P(C)$$

или

$$P(X < b) = P(X < a) + P(a \le X < b).$$

$$P(a \le X < b) = P(X < b) - P(X < a) +.$$

Используя функцию распределения $F(x)$ запишем

$$\boxed{\boldsymbol{P(a \le X < b) = F(b) - F(a).}} \quad (3.9)$$

То есть вероятность попадания НСВ в некоторый интервал равна приращению функции распределения на этом интервале.

Возможность по ФР НСВ определения вероятности попадания НСВ в любой интервал, включая (a, a) *– т.е. определение вероятности принятия НСВ значения «а» означает, что ФР НСВ* $F(X)$ *исчерпывающе характеризует НСВ.*

Таким образом *закон распределения НСВ задается ее ФР.*

Кстати, по (3.9) вероятность принятия НСВ значения a

$$P(X = a) = P\ (a \le X < a) = F(a) - F(a) = 0.$$

Можно ли задать дискретную СВ функцией распределения? Можно. Она будет иметь вид неубывающей ступенчатой функции со ступеньками в точках, соответствующих возможным значениям ДСВ, а величина «ступени», т.е. прироста ФР – вероятности принятия ДСВ этого значения.

Таким образом ФР задает закон распределения любой СВ и по ней можно сразу определить НСВ и ДСВ.

Можно дать более точное определение НСВ.

Случайную величину X называют непрерывной, если ее функция распределения непрерывна в любой точке и дифференцируема всюду, кроме, может быть, отдельных точек.

Пример. Дана ДСВ:

X	0	1	2	3
P	1/56	15/56	30/56	10/56

Тогда ее ФР

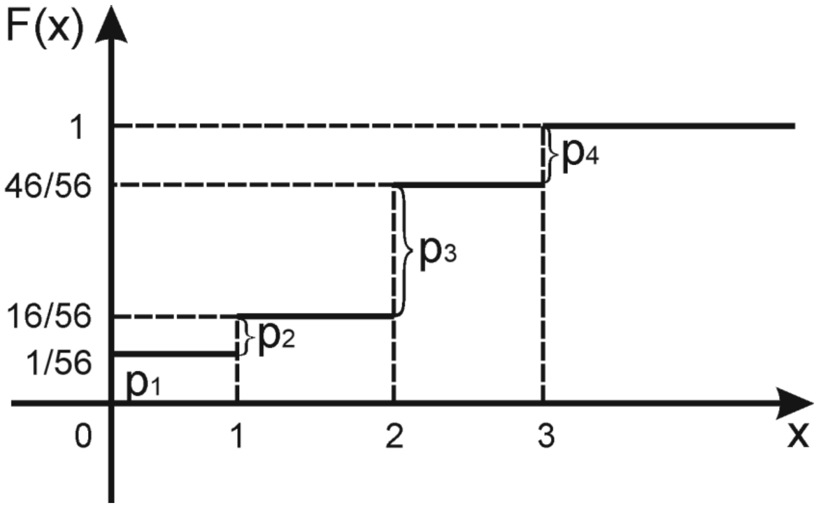


Рис. 3.6. Функция распределения дискретной случайной величины

Вопрос: что это за случайная величина?

$$F(x) = \begin{cases} 0, & x \le x_0, \\ 1, & x > x_0. \end{cases}$$

Свойства функции распределения непрерывной случайной величины

1. $F(x)$ ограничена, т.е.

$$0 \le F(x) \le 1.$$

2. $F(x)$ – неубывающая функция на R, т.е. если $x_2 > x$, то

$$F(x_2) \geq F(x_1).$$

3. $F(x)$ обращается в ноль на минус бесконечности и равна единице в плюс бесконечности, т.е.

$$F(-\infty) = 0, F(+\infty) = 1.$$

4. Вероятность попадания СВ X в промежуток $[a, b)$ равна приращению ее функции распределения на этом промежутке, т.е.

$$P\{a \leq X < 6\} = F(b) - F(a).$$

5. $F(x)$ непрерывна слева, т.е.

$$\lim_{x \to x_0 - 0} F(x) = F(x_0).$$

Задача.

Убедиться, что функция

$$F(x) = \begin{cases} 0, & \text{если } x \leq 0, \\ 1 - e^{-x}, & \text{если } x > 0. \end{cases}$$

является функцией распределния некоторой случайной величины. Найти $P\{0 \leq x < 1\}$ и построить график $F(x)$.

Плотность распределения непрерывной случайной величины

Важнейшей характеристикой непрерывной случайной величины (помимо функции распределения) является *плотность распределения вероятностей*.

Напомним, что СВ X называется непрерывной, если ее функция распределения непрерывна и дифференцируема всюду, кроме, быть может, отдельных точек.

Плотностью распределения вероятностей (плотностью распределения, плотностью вероятностей или просто плотностью) непрерывной случайной величины X называется *производная ее функции распределения*

$$f(x) = F'(x).$$

Функцию $f(x)$ называют также *дифференциальной функцией распределения*. Она является одной из форм закона распределения случайной величины, *существует только для непрерывных случайных величин*.

Установим вероятностный смысл плотности распределения. Из определения производной следует

$$f(x) = \lim_{\Delta x \to 0} \frac{\Delta F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

или

$$f(x) = \lim_{\Delta x \to 0} \frac{P\{x \le X < x + \Delta x\}}{\Delta x},$$

т.е. плотность распределения есть предел отношения вероятности попадания НСВ в промежуток $(x, x + \Delta x)$ к длине Δx этого промежутка, когда Δx стремится к нулю.

«Физический смысл» плотности – «*скорость*» *роста вероятности.*

Свойства плотности распределения непрерывной случайной величины

1. $f(x)$ неотрицательная, т.е.

$$f(x) \ge 0.$$

Это следует из того, что $F(x)$ – неубывающая, значит ее производная неотрицательная.

2. Вероятность попадания НСВ в промежуток $(a; b)$ равна определенному интегралу от ее плотности в пределах от a до b, т.е.

$$P\{x \le X \le b\} = \int_a^b f(x)dx. \quad (3.10)$$

Это следует из

$$\int_a^b f(x)dx = F(b) - F(a) = P\{a \le X < b\}.$$

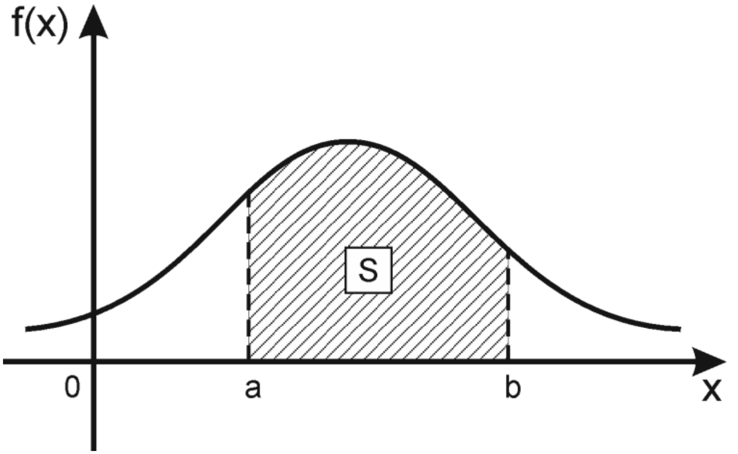


Рис. 3.7. Вероятность принятия НСВ значения в пределах от *a* до *b* как определенный интеграл от плотности распределения НСВ

3. Функция распределения НСВ может быть выражена через ее плотность вероятности по формуле

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

4. Условие нормировки: несобственный интеграл от плотности вероятности НСВ в бесконечных пределах равен единице, т.е.

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

Очевидно, что

$$P\{X = c\} = P\{c \le X \le c\} = \int_c^c f(x)dx = 0.$$

3.5. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Числовые характеристики НСВ удобно рассматривать в сравнении с аналогичными числовыми характеристиками для ДСВ.

Математическое ожидание НСВ

Математическим ожиданием (или *средним* значением) ДСВ X, имеющей закон распределения $p_i = P\{X = x_i\}$, $i = 1, 2, 3, \ldots, n$, называется число, равное сумме произведений всех ее значений на соответствующие им вероятности.

Таким образом, по определению

$$M\left[X\right] = \sum_{i=1}^{n} x_i \cdot p_i.$$

Математическим ожиданием НСВ X с плотностью вероятности $f(x)$ называется число, определяемое по выражению

$$M[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx.$$

Свойства МО НСВ аналогичны свойствам МО ДСВ:

1. Математическое ожидание постоянной равно самой этой постоянной, т.е.

$$M[C] = C.$$

2. Постоянный множитель выносится за знак МО т.е.

$$M[CX] = CM[X].$$

3. МО суммы НСВ равно сумме их МО, т.е.

$$M[X + Y] = M[X] + M[Y].$$

4. МО отклонения НСВ от ее МО равно нулю, т.е.

$$M[X – MX] = 0.$$

5. МО произведения независимых НСВ равно произведению их МО, т.е. если X и Y независимы, то

$$M[X \cdot Y] = M[X] \cdot M[Y].$$

Дисперсия и СКО НСВ

Дисперсией СВ X называется математическое ожидание квадрата ее отклонения от своего математического ожидания

$$D[X] = M[X - M[X]]^2.$$

Для ДСВ X

$$D[X] = \sum_i (x_i - M[X])^2 \cdot p_i;$$

для НСВ X

$$D[X] = \int_{-\infty}^{+\infty} (x - M[X])^2 \cdot f(x)dx$$

или

$$D[X] = \sum_i x_i^2 \cdot p_i - (M[X])^2,$$

$$D[X] = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx - (M[X])^2 \cdot$$

Свойства дисперсии НСВ

1. Дисперсия постоянной равна нулю, т.е.

$$D[C] = 0.$$

2. Постоянный множитель можно выносить за знак дисперсии, возведя его в квадрат, т.е.

$$D[C\cdot X] = C^2\cdot D[X].$$

3. Дисперсия суммы независимых СВ равна сумме их дисперсий, т.е. если X и Y независимы, то

$$D[X + Y] = D[X] + D[Y].$$

4. Дисперсия СВ не изменится, если к этой СВ прибавить постоянную, т.е.

$$D[X + C] = D[X].$$

5. Если СВ X и Y независимы, то

$$D[XY] = M[X^2]\cdot M[Y^2] - (M[X])^2\cdot(M[Y])^2.$$

Средним квадратическим отклонением СВ X называется квадратный корень из ее дисперсии

$$\sigma(X) = \sqrt{D[X]}.$$

Мода и медиана НСВ

Модой ДСВ X называется ее значение, принимаемое с наибольшей вероятностью по сравнению с двумя соседними значениями, обозначается через $M_o X$.

Для НСВ $M_o X$ – точка локального максимума плотности $f(x)$.

Если мода единственна, то распределение НСВ называется унимодальным, в противном случае – полимодальным (рис. 3.8).

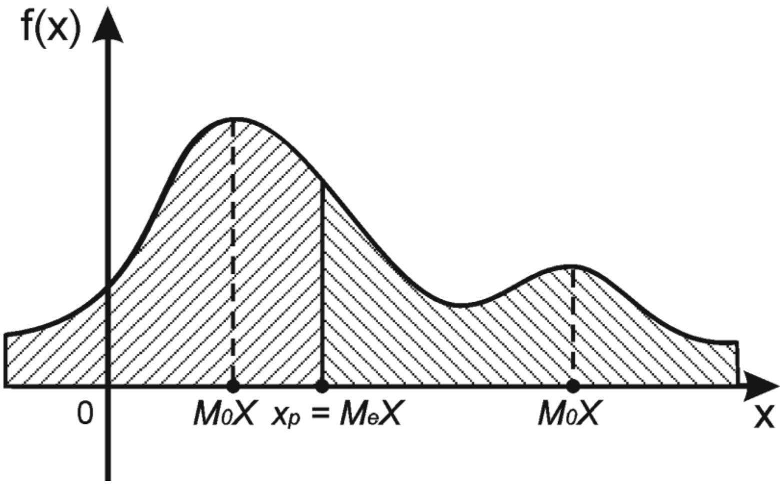


Рис. 3.8. Полимодальное распределение НСВ

Медианой M_eX НСВ X называется такое ее значение x_p, для которого

$$P(X < x_p) = P(X > x_p) = ½,$$

т.е. одинаково вероятно, окажется ли СВ X меньше x_p или больше x_p (см. рис. 3.8).

Кроме того, математическое ожидание и дисперсия являются частными случаями следующих более общих понятий – моментов СВ.

Начальным моментом порядка k СВ X называется МО k-й степени этой величины, обозначается, через α_k.

Таким образом, по определению

$$\alpha_k - M[X^k].$$

Для ДСВ начальный момент выражается суммой:

$$\alpha_k = \sum_i x_i^k \cdot p_i,$$

а для НСВ – интегралом:

$$\alpha_k = \int_{-\infty}^{+\infty} x^k \cdot f(x)dx.$$

В частности, $\alpha_1 = M[X]$, т.е. начальный момент 1-го порядка есть МО.

Центральным моментом порядка k СВ X называется МО величины $(X - M[X])^k$ обозначается через μ_k.

Таким образом, по определению

$$\mu_k = M[X - M[X]]^k.$$

В частности, $\mu_k = D[X]$, т.е. центральный момент 2-го порядка есть дисперсия; $\mu_1 = M[X - M[X]] = 0$ (см. свойство 4 МО НСВ).

Для ДСВ:

$$\mu_k = \sum_i (x_i - M[X])^k \cdot p_i,$$

а для НСВ:

$$\mu_k = \int_{-\infty}^{+\infty} (x - M[X])^k \cdot f(x)dx.$$

3.6. ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЙ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Равномерное распределение

Непрерывная СВ X имеет *равномерное распределение* на отрезке $[a, b]$, если ее плотность вероятности $f(x)$ постоянна на этом отрезке, а вне его равна нулю:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{при } x \in [a, b], \\ 0, & \text{при } x \notin [a, b], \end{cases}$$

т.е. $f(x) = c$ при $x \in [a, b]$, но $\int_{-\infty}^{+\infty} c\,dx = 1$, очевидно $c = \dfrac{1}{b-a}$.

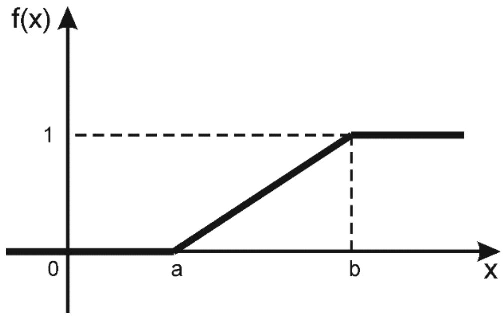


Рис. 3.9. Функция распределения вероятности равномерного распределения НСВ

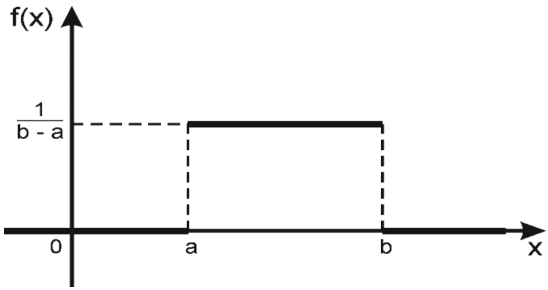


Рис. 3.10. Плотность вероятности равномерного распределения НСВ

$$F(x)=\int_a^x \frac{dt}{b-a}=\frac{t}{b-a}\Big|_a^x=\frac{x-a}{b-a},$$

$$M[X]=\frac{a+b}{2},\ D[X]=\frac{(b-a)^2}{12},$$

$$\sigma(X)=\frac{b-a}{2\sqrt{3}},$$

$$M[X]=\int_b^a x\cdot\frac{1}{b-a}dx=\frac{x^2}{2(b-a)}\Big|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{a+b}{2},$$

$$M[X^2]=\int_{-\infty}^{+\infty} x^2 f(x)dx=\int_b^a x^2\frac{1}{b-a}dx=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3},$$

Пояснение:

$$(b-a)\cdot(a^2+ab+b^2)=a^2b+ab^2+b^3-a^3-a^2b-ab^2=b^3-a^3,$$

$$D[X]=M[X^2]-(M[X])^2=\frac{(b-a)^2}{12}.$$

Доказательство:

$$\frac{a^2+ab+b^2}{3}-\left(\frac{a+b}{2}\right)^2=\frac{4(a^2+ab+b^2)-3(a+b)^2}{12}=$$

$$=\frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12}=\frac{a^2-2ab+b^2}{12}=\frac{(a-b)^2}{12}.$$

Вопрос: Чему равна $P(x_1 < X < x_2)$, если $x_1 < a$, $x_2 > b$.

Кстати:

Дискретная случайная величина X имеет равномерное распределение, если она принимает целочисленные значения 1, 2, 3, ..., n с вероятностью $p_m = P\{X = m\} = 1/n$, где $m = 1, 2, 3, ..., n$.

В этом случае $M[X] = (1 + n)/2$, $D[X] = (n^2 - 1)/12$. Так, при $n = 5$, многоугольник распределения имеет вид: $M[X] = 3$.

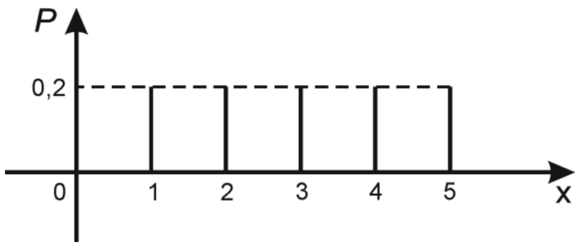


Рис. 3.11. Многоугольник равномерного распределения ДСВ с пятью последовательными возможными значениями

В общем случае:

Пусть случайная величина X принимает n значений с вероятностями $p_1, p_2, \ldots, p_n$. Данная случайная величина называется равномерно распределенной случайной величиной, если $p_i = \frac{1}{n},\ i = \overline{1, n}$ для всех $x \in X = \{x_1, x_2, ..., x_n\}$.

В этом случае:

– ряд распределения

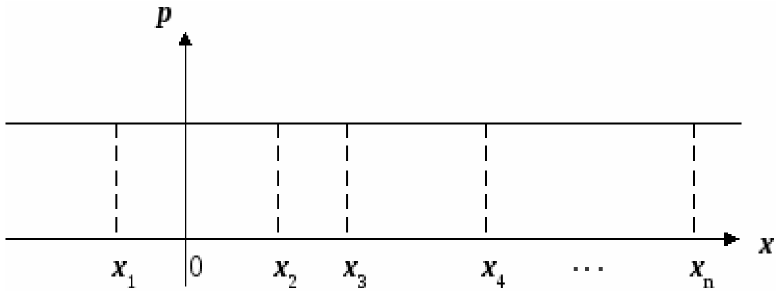


Рис. 3.12. Многоугольник равномерного распределения ДСВ, имеющей *n* произвольных возможных значений

– функция распределения

$$F(x_i) = \frac{1}{n} \cdot \sum_{k<i} k,\ \ k = 0, 1,\ 2, ..., n.$$

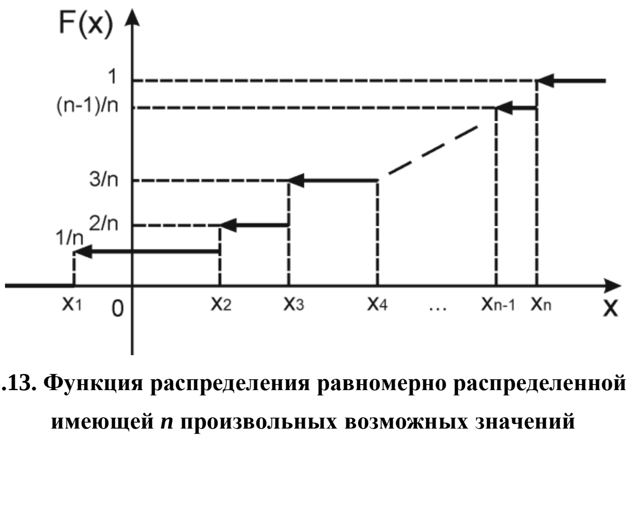


Рис. 3.13. Функция распределения равномерно распределенной ДСВ, имеющей *n* произвольных возможных значений

– математическое ожидание

$$M[X]=\frac{1}{n}\cdot\sum_{i=1}^{n}x_i;$$

– дисперсия

$$D[X]=\frac{1}{n}\cdot\sum_{i=1}^{n}x_i^2+(M(X))^2.$$

Показательное (экспоненциальное) распределение

Непрерывная случайная величина X имеет *показательный* (или *экспоненциальный*) закон распределения, если ее плотность вероятности имеет вид

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, \text{ при } x\ge 0,\\ 0, \quad\quad \text{ при } x<0,\end{cases} \tag{3.11}$$

где $\lambda > 0$ – параметр распределения.

График плотности $f(x)$ приведен на рис. 3.14.

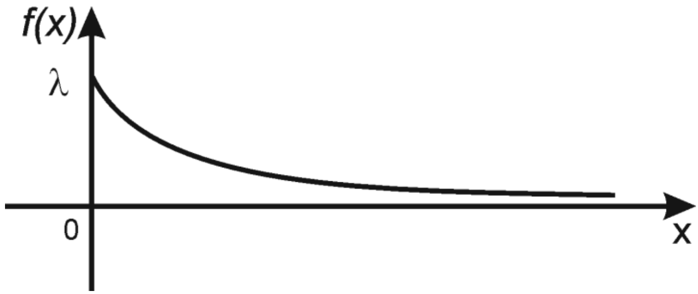


Рис. 3.14. Плотность экспоненциального распределения

Функция распределения показательного распределения имеет вид

$$F(x)=\begin{cases}1-e^{-\lambda x}, \text{ при } x\ge 0,\\ 0, \quad\quad \text{ при } x<0.\end{cases} \tag{3.12}$$

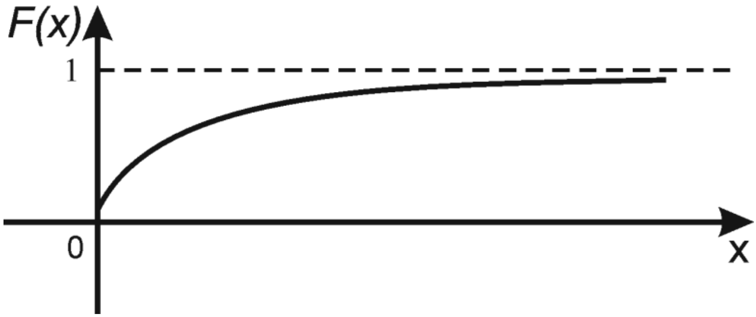


Рис. 3.15. Функция распределения экспоненциального распределения

$$MX = \frac{1}{\lambda},\ DX = \frac{1}{\lambda^2},\ \sigma_X = \frac{1}{\lambda}$$

Таблица 3.1

Обозначение	Exp(λ)
Параметры	$\lambda > 0$ – интенсивность или обратный *коэффициент масштаба*
Область определения	$x \in [0; \infty)$
Плотность вероятности	$\lambda e^{-\lambda x}$
Функция распределения	$1 - e^{-\lambda x}$
Математическое ожидание	λ^{-1}
Медиана	$\ln(2)/\lambda$
Мода	0
Дисперсия	λ^{-2}

Показательное распределение используется в приложениях теории вероятностей, особенно в теории массового обслуживания (ТМО), в физике, в теории надежности. Оно используется для описания распределения случайной величины вида: длительность работы прибора до первого отказа, длительность времени обслуживания в системе массового обслуживания и т.д.

Нормальное распределение

Нормальный закон («закон Гаусса») играет исключительную роль в теории вероятностей. Главная особенность закона Гаусса состоит в том, что он является предельным законом, к которому приближаются при определенных условиях другие законы распределения. Нормальный закон наиболее часто встречается на практике.

НСВ X распределена по нормальному закону с параметрами a и $\sigma > 0$, если ее плотность распределения имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \ x \in R. \tag{3.13}$$

Функция распределения

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt, \tag{3.14}$$

где a – МО; σ – СКО.

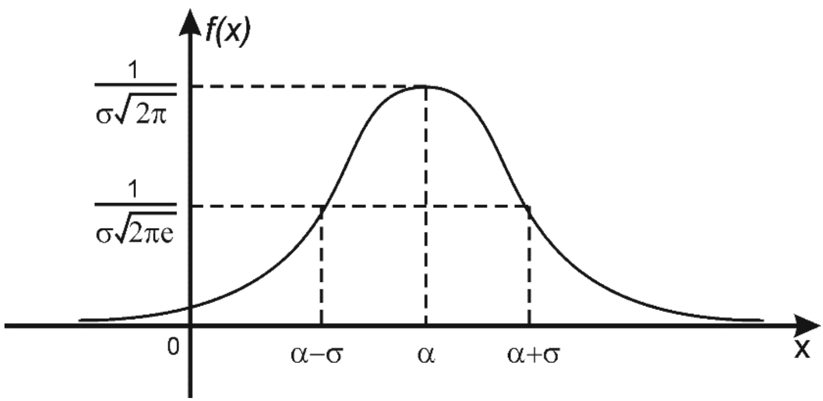


Рис. 3.16. Плотность нормального распределения

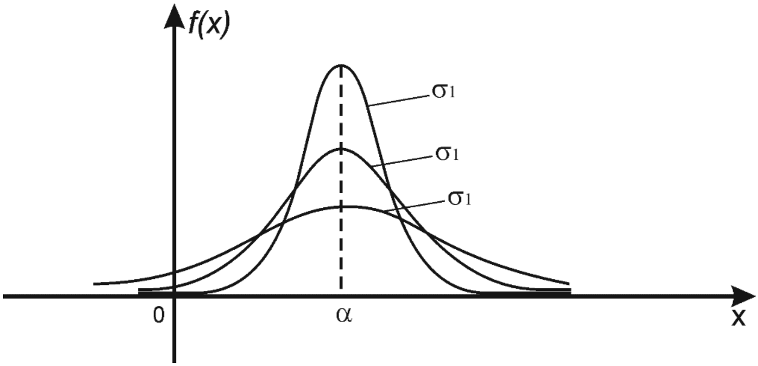


Рис. 3.17. Плотность нормального распределения при различных СКО

$$\sigma_1 < \sigma < \sigma_2.$$

Если $\alpha = 0$ и $\sigma = 1$, то нормальное распределение с такими параметрами называется *стандартным*. Плотность стандартной случайной величины имеет вид

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{3.15}$$

Его функция распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \tag{3.16}$$

называется «*функция Лапласа*».

Известна также табличная нечетная *нормированная функция Лапласа* (см. ниже).

Через функцию Лапласа $\Phi_0(x)$ выражается и функция распределения $F(x)$ нормально распределенной НСВ X.

$$F(x) = \frac{1}{2} + \Phi_0\left(\frac{x-a}{\sigma}\right). \tag{3.17}$$

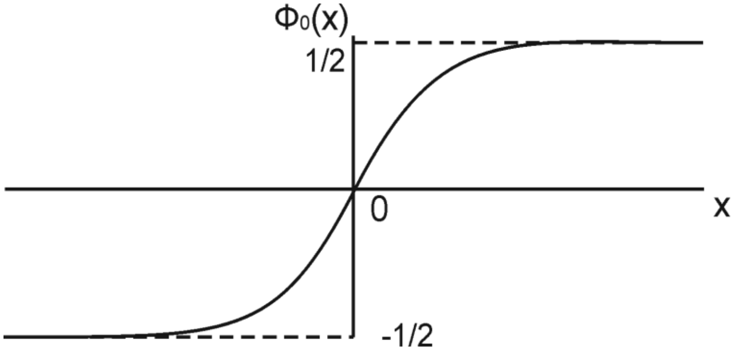


Рис. 3.18. Табличная нечетная нормированная функция Лапласа

Также используя функцию Лапласа можно найти

$$P\{\alpha < X < \beta\} = \Phi_0\left(\frac{\beta - a}{\sigma}\right) - \Phi_0\left(\frac{\alpha - a}{\sigma}\right). \tag{3.18}$$

Нормальному закону подчиняются ошибки измерений, величины износа деталей в механизмах, рост человека, ошибки стрельбы, вес клубней картофеля, величина шума в радиоприемном устройстве, колебания курса акций и т.д.

3.2. Значение функции $\Phi_0(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0040	0080	0112	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0754
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1814	1879

Продолжение табл. 3.2

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2258	2291	2324	2357	2389	2422	2454	2486	2518	2549
0,7	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2996	3023	3051	3079	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3438	3461	3485	3508	3531	3553	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	1394	4406	4418	4430	4441
1,6	4452	4463	4474	4485	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4700	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4762	4767
	Десятые доли x									
x	0	1	2	3	4	5	6	7	8	9
2,	4773	4821	4861	4893	4918	4938	4953	4965	4974	4961
3,	4987	4990	4993	4995	4997	4998	4998	4999	4999	5000[1]

[1] Точное значение отличается меньше, чем на $5\cdot10^{-5}$.

Усеченное нормальное распределение

Область определения нормального распределения – вся действительная числовая ось $(-\infty, +\infty)$. Иногда это создает определенные трудности в его применении.

Пусть T, $T > 0$ – среднее время работы устройства. Реальное время t работы устройства – НСВ, которая, в теории может иметь и отрицательное значение.

Если $T \gg \sigma$ ($T > 3\sigma$), то вероятностью этого события можно пренебречь, но если T и σ соизмеримы ($T > 3\sigma$), то такое пренебрежение приводит к погрешности. На рисунке ниже приведен пример с $T = 1{,}5\sigma$. Из него видно, что «отрицательная наработка» имеет существенную вероятность.

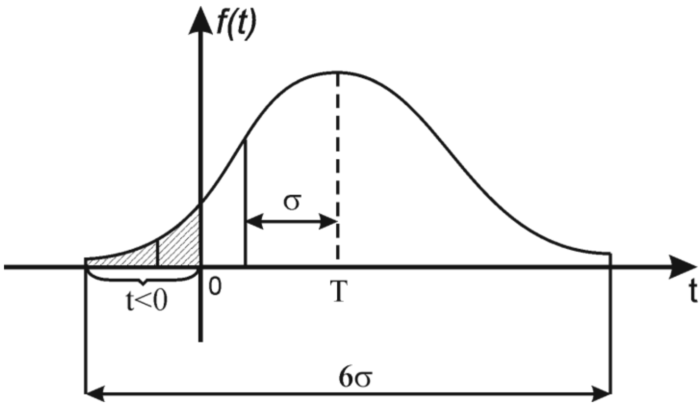


Рис. 3.19. Отрицательная наработка

Усеченным нормальным распределением (УНР) называется распределение, получаемое из классического нормального, при ограничении интервала возможных значений НСВ.

В общем случае усечение может быть:

– левым – (t_1; $+\infty$), в частности $t_1 = 0$;

– двусторонним – (t_1, t_2).

Плотность УНР описывается формулой (для левого усечения)

$$f_{\text{ус}}(t) = \begin{cases} 0, & t \le t_1, \\ c \cdot f(t), & t \ge t_1, \end{cases}$$

где $f(t)$ – плотность нормального распределения, см. (3.13); c – константа, значение которой находят исходя из

$$\int_{t_1}^{\infty} f_{\text{ус}}(t) = 1,$$

$$\int_{t_1}^{\infty} f_{\text{ус}}(t) = \int_{t_1}^{\infty} c \cdot f(t) = c\int_{t_1}^{\infty} f(t) = 1, \tag{3.19}$$

$$c = \frac{1}{\int_{t_1}^{\infty} f(t)} = \frac{1}{1 - F(t_1)}.$$

С использованием функции Лапласа

$$c = \frac{1}{1 - F(t_1)} = \frac{1}{1 - \left(0{,}5 + \Phi_0\left(\frac{t_1 - T}{\sigma}\right)\right)} = \frac{1}{0{,}5 - \Phi_0\left(\frac{t_1 - T}{\sigma}\right)}, \tag{3.20}$$

где T – математическое ожидание.

Плотность УНР для двустороннего усечения описывается формулой

$$f_{\text{ус}}(t) = \begin{cases} 0, & t \le t_1, \\ c \cdot f(t), & t_1 < t < t_2, \\ 0, & t \ge t_2, \end{cases} \tag{3.21}$$

где $f(t)$ – плотность нормального распределения; c – константа, значение которой находят исходя из

$$\int_{t_1}^{t_2} f_{\text{ус}}(t) = 1,$$

$$\int_{t_1}^{t_2} f_{\text{ус}}(t) = \int_{t_1}^{t_2} c \cdot f(t) = c\int_{t_1}^{t_2} f(t) = 1, \tag{3.22}$$

$$c = \frac{1}{\int\limits_{t_1}^{t_2} f(t)}.$$

С использованием функции Лапласа

$$c = \frac{1}{\Phi_0\left(\frac{t_2 - T}{\sigma}\right) - \Phi_0\left(\frac{t_1 - T}{\sigma}\right)}. \qquad (3.23)$$

Кривая $f_{\text{ус}}(t)$ выше, чем $f(t)$, так как площади под кривыми $f_{\text{ус}}(t)$ и $f(t)$ одинаковы и равны единице.

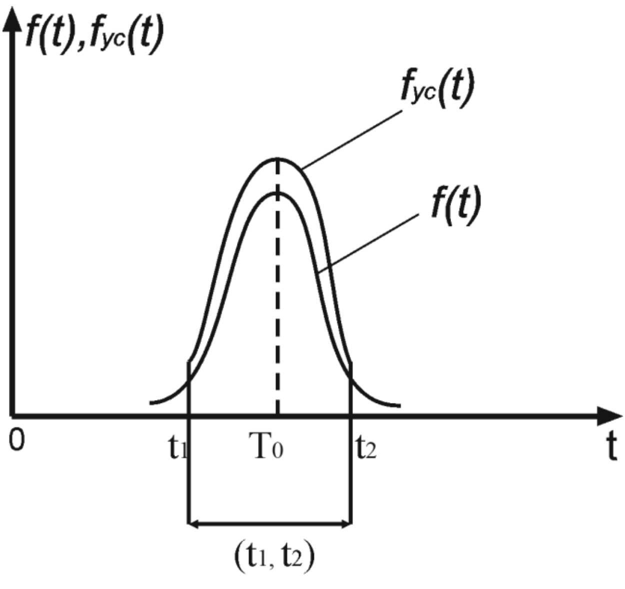


Рис. 3.20. Плотности нормального и усеченного нормального распределений

Ниже представлена зависимость коэффициента c для левого усечения от соотношения T/σ.

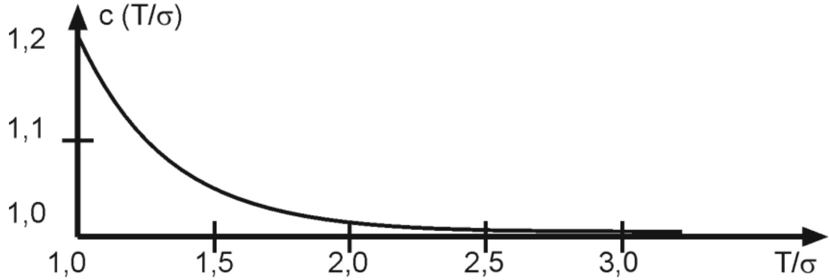


Рис. 3.21. Зависимость коэффициента *c*
для левого усечения от соотношения *T*/σ

Из рисунка видно, что

$\max c(T/\sigma) = 1{,}2$ при $T/\sigma = 1$, т.е. $T = \sigma$;

при $T/\sigma > 3$, т.е. $T > 3\sigma$, $c = 1$ и $f_{\text{ус}}(t) = f(t)$.

4. ОСНОВЫ ТЕОРИИ ЭФФЕКТИВНОСТИ И ПРИНЯТИЯ РЕШЕНИЙ

Большинство математических постановок задач (МПЗ) научных исследований формулируется в виде оптимизационной задачи.[14]

В связи с этим целесообразно рассмотреть имеющие к этому отношение вопросы эффективности, качества, показателей и критериев, оптимизации и принятия решений. Это тем более актуально, что в настоящее время имеют место неоднозначности в определениях этих терминов в различных научных школах и системах терминов.

4.1. СИСТЕМА ОСНОВНЫХ ПОНЯТИЙ ТЕОРИИ ЭФФЕКТИВНОСТИ

Термины «эффективность», «качество» и им сопутствующие применяются к РАЗЛИЧНЫМ категориям объектов, но в то же время имеют общие свойства, и чтобы рассматривать их СИСТЕМНО целесообразно рассмотреть системообразующее понятие – ОПЕРАЦИЮ.

Операция есть система целенаправленных действий, объединенных общим замыслом и единой целью.

Понятие операции включает по меньшей мере три определяющих момента:

– управляющая деятельность;
– активные средства;
– влияющие факторы, условия (другие средства).

1. *Управляющая деятельность человека* (органа управления, распорядительного центра, лица, принимающего решения (ЛПР)), организующего операцию на основе выбора рационального *способа использования активных средств для достижения цели операции – стратегии* в условиях воздействия влияющих факторов.

[14] Подиновский В. В. Математическая теория выработки решений в сложных ситуациях. Учебник. – М. : МО СССР, 1981. – 211 с.

2. *Активные средства* (технические системы, ресурсы), находящиеся в распоряжении управляющего органа (ЛПР), и используемые в операции в соответствии с выбранным способом (стратегией) применения.

3. *Влияющие факторы, условия* (*другие средства*) – процессы, явления, функционирующие технические и эргасистемы, *неподвластные ЛПР, взаимодействующие с его активными средствами и влияющие на исход операции,* к которым обычно относят:

– объекты воздействия активных средств;

– средства, находящиеся в распоряжении других распорядителей в операции (их активные средства);

– объекты, не находящиеся ни в чьем распоряжении;

– условия, в которых используются активные средства.

Это могут быть внутренние (по отношению к системе) и внешние, определенные и неопределенные (стохастические и нестохастические), целенаправленно-действующие (активные средства других распорядителей в операции) и действующие стихийно.

В общем плане эти три момента отражают ответы на вопросы *как действовать, чем действовать и на что воздействовать для успешного достижения поставленной цели операции*.

Управляющая деятельность

Цель операции выступает основным системообразующим фактором, как способ интеграции различных действий в единую последовательность.

Цель есть идеальное представление (предвосхищение) в сознании ЛПР желаемого результата операции. Она определяет способы и формы действий, их характер и системную упорядоченность, а также средства достижения и выступает как определенный механизм интеграции различных действий в систему «цель–средство–результат».

В формализованном виде цель выражается набором – *вектором, кортежем* (далее – *вектором*) некоторых параметров целеполагания Y^0 (в частном случае всего один параметр – *скаляр* может отражать цель операции).

Результат Y операции зависит (при равных исходных данных) от выбранной стратегии u и описывается *вектором тех же параметров, что и цель операции*. В общем случае функция $Y(u)$ может быть вектором, компоненты которого характеризуют частные результаты операции, или результаты частных операций, направленных на достижение частных целей, на которые декомпозируется общая цель операции.

Обычно в вектор параметров целеполагания Y^0 и результат $Y(u)$ операции включают полезный эффект q, затраченные ресурсы C и время T[15]:

$$Y^0 = Y^0(q^0, C^0, T^0),$$

$$Y(u) = Y\,(q(u), C(u), T(u)).$$

Формально-математически вектор параметров целеполагания Y^0 и результат $Y(u)$ операции задают векторами

$$Y^0 = \left(y_1^0, y_2^0, ..., y_N^0\right),$$

$$Y(u) = \left(y_1(u), y_2(u), ..., y_N(u)\right),$$

Функцию целеполагания может осуществлять старшая система, в состав которой входит исследуемая. В этом случае старшая система ставит задачу, т.е. осуществляет руководство исследуемой системой.

Реальный результат $Y(u)$ операции (фактический или ожидаемый ее конечный итог) может не совпадать с желаемым Y^0.

[15] Время T также является ресурсом и может быть включено в C.

В [20, 22] под ***эффективностью*** *варианта целенаправленной деятельности* (*стратегии*) *понимается степень соответствия ее результата* $Y(u)$ *требуемому* (*желаемому*) *результату* Y^0.

Особенностями данного подхода к определению эффективности являются простота сопоставления и сравнения различных стратегий по эффективности с одной стороны, и необходимость построения *метрического* пространства на множестве возможных значений кортежа параметров, характеризующих результат реализации стратегии $Y(u)$, что не всегда возможно. Требуется скалярная численная *мера* этой степени соответствия $p(Y(u), Y^0)$.

Возможные альтернативные пути (стратегии) достижения цели в общем случае обладают разной эффективностью.

Целесообразен выбор (синтез) стратегии, обеспечивающей *наилучшую*, или *достаточную* эффективность операции. Такая стратегия называется *оптимальной*. В этом контексте корректными будут выражения «более эффективный …», «менее эффективный …», «самый эффективный …» и некорректными – «более оптимальный», «самый оптимальный …».

Следует обратить внимание на две задачи:

1. Оценка эффективности стратегии – задача анализа.
2. Определение (выбор, синтез) оптимальной стратегии – это задача синтеза, задача принятия решений (ЗПР).

Вторая задача решается обоснованием (выбором) и применением правила, определяющего предпочтительность стратегии – *критерия эффективности*.

Существуют научные школы, в частности [Дзялошинский И. М. Проблема эффективности пропагандистской деятельности, осуществляемой с помощью СМИ / И. Дзялошинский // http://www.dzyalosh.ru/01-comm/statii/dzyalosh-01/poiski.html], где под *эффективностью операции понимают только ее целевой эффект, а под оптимальностью –*

как достигнутый эффект, так и затраченные на это ресурсы и влияние на надсистему (для общественной деятельности – полезность и моральную оценку этой деятельности). В этом контексте выражения слова «более», «менее» или «самый» к термину «оптимальный» вполне корректны.

Или еще «эффективность – это продуктивность использования ресурсов в достижении какой-либо цели (экономическая эффективность, эффективность производства)».

ГОСТ Р ИСО 9000–2015 СИСТЕМЫ МЕНЕДЖМЕНТА КАЧЕСТВА. ОСНОВНЫЕ ПОЛОЖЕНИЯ И СЛОВАРЬ (стр. 17):

3.7.10 **эффективность** (efficiency): Соотношение между достигнутым результатом и использованными ресурсами.

3.7.11 **результативность** (effectiveness): Степень реализации запланированной деятельности и достижения запланированных результатов.

Еще из этого ГОСТа:

3.5.12 **стратегия** (strategy): План достижения долгосрочной или общей цели (3.7.1).

3.7.1 **цель** (ob*j*ective): Результат, который должен быть достигнут.

Семантика термина «эффективность»

Обратим внимание на семантику термина «эффективность»:

Эффект – результат какого-либо действия, характеристика действия.

Эффективный – дающий эффект, результативный, действенный. Отсюда – *эффективность – это свойство действия,* свойство операции давать эффект.

Термин «эффективность» может быть применен к оценке свойств алгоритма, методики, стратегии.

Так как же быть с часто используемым выражением «эффективность технической системы»?

А словосочетание «эффективность технической системы» в общем случае некорректно, так как техническая система не есть действие.

Эффективность применения технических систем

Активные средства. Пусть операция как целенаправленная деятельность имеет некоторую цель, *однако* в ней в качестве инструмента достижения цели применяются технические системы, ресурсы, эргасистемы[16] – далее ТС.

Говоря об эффективности технической системы, имеют в виду *эффективность использования ее в качестве активного средства в заранее* (*до принятия решения*) *известной или типовой операции, для проведения которой предназначена эта система*.

Техническая система как активное средство может быть использовано в операции различными способами, т.е. различными стратегиями, и результат использования (эффективность) также может быть различным (-ной).

Под эффективностью ТС в общем случае понимают ее ***потенциальную эффективность***, т.е. эффективность при наилучшем способе применения.

ТС может применяться в качестве активного средства в нескольких операциях, цели которых могут не совпадать. В таком случае рассчитывают эффективность всех операций. Может оказаться, что в одной операции исследуемая техническая система обладает высокой эффективностью, т.е. является высокоэффективным средством достижения первой цели, тогда как в другой операции эффективность низкая и эта техническая система является недостаточно эффективным средством достижения цели второй операции. Проиллюстрируем сказанное. Пусть имеются

[16] Эргатическая система – схема производства, одним из элементов которой является человек или группа людей.

три ТС: C^1, C^2, C^3, которые характеризуются двумя параметрами S^1 и S^2 (рис. 4.1). ТС – генераторы электроэнергии, характеризующиеся рабочей температурой (S^1) и выходной мощностью (S^2). Генератор C^1 предназначен для работы в условиях жаркого климата, C^2 – умеренного, C^3 – холодного арктического. Генератор C^2 имеет наибольшую мощность, C^3 – наименьшую, C^1 – среднюю.

Пусть следует выбрать из имеющихся трех мощный генератор для работы в жарком климате. Эту цель иллюстрирует вектор O^1.

Опустим на этот вектор перпендикуляры из трех точек пространства характеристик C^1, C^2, C^3 и оценим пригодность генераторов по длине векторов – проекций. Из рисунка видно, что лучшим будет первый генератор C^1, пригодным – C^2, непригодным – C^3.

Теперь пусть следует выбрать из имеющихся трех мощный генератор для работы в холодном климате. Эту цель иллюстрирует вектор O^2.

Опустим теперь перпендикуляры из трех точек пространства характеристик C^1, C^2, C^3 на этот вектор и оценим пригодность генераторов по длине векторов – проекций.

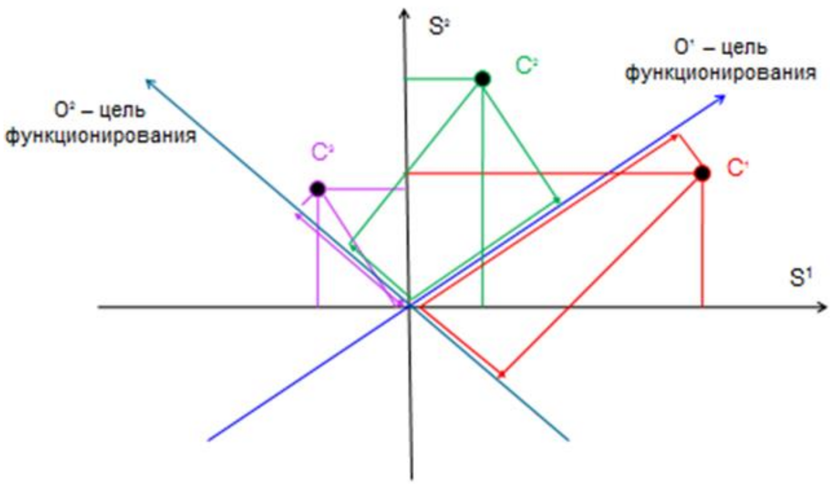


Рис. 4.1. Пример соответствия параметров ТС целям ее функционирования

Из рисунка видно, что лучшим будет третий генератор C^3, пригодным – C^2, крайне непригодным – C^1.

Таким образом, имеются два специализированных генератора, предназначенные для работы в экстремальных климатических условиях и один универсальный, предназначенный для работы в умеренном климате.

4.2. КАЧЕСТВО ТЕХНИЧЕСКИХ СИСТЕМ

Понятие качества

Применяемые в качестве активных средств технические системы обладают рядом свойств, часть из которых в той или иной степени способствуют достижению цели, часть не влияют на ее достижение, а часть – препятствуют.

Чтобы ТС можно было использовать в качестве активного средства в операции, она должна обладать определенными свойствами.

Свойство объекта (ТС) – его объективная особенность, проявляющаяся при его создании, эксплуатации (потреблении) *при взаимодействии с другими объектами материального мира*.

Свойства могут характеризоваться вербально, численно, в виде таблиц, графиков и т.д.

Каждый объект находится в бесконечных взаимосвязях с другими объектами и обладает бесконечным множеством свойств. Поэтому попытки перечислить и охарактеризовать все его свойства уводят в бесконечность.

Однако, с точки зрения использования объекта в качестве активного средства в операции, не все его свойства имеют одинаковую значимость: одни из них могут быть важнейшими, другие второстепенными, а третьи вообще могут не отражаться на эффективности использования данного объекта. Здесь все зависит от того, какая задача выполняется с помощью данного объекта.

Качество объекта – это совокупность его свойств, обусловливающих его пригодность удовлетворять определенные потребности в соответствии с его назначением (по ГОСТ 15467–79).

По ГОСТ Р ИСО 9000–2015:

3.6.2 **качество (quality):** Степень соответствия совокупности присущих *характеристик объекта требованиям*.

Таким образом ***качество*** **ТС** – это кортеж ее свойств, существенных с точки зрения возможности и целесообразности ее использования при реализации стратегии.

Категория качества объекта не сводится к отдельным его свойствам, даже важнейшим. Она выражает целостную характеристику функционального единства существенных свойств объекта, его определенности, его отличия от других объектов или сходства с ними.

Таким образом, характеризуя активное средство в операции (ТС) следует говорить не о его эффективности, а о его КАЧЕСТВЕ.

Методический аппарат исследования качества

Проблема качества имеет много аспектов, в зависимости от целей исследований.

В самом широком философском смысле качество характеризует *неотделимую от бытия объекта его существенную определенность, благодаря которой он является именно этим, а не иным объектом, или, иначе, качество – это наличие существенных признаков, свойств* (*атрибутов*), *отличающих один объект от другого*.

Философскую категорию «качество» исследовал еще Аристотель: «*… тот пребывающий видовой признак, который отличает данную сущность в ее видовом своеобразии от другой сущности, принадлежащей к тому же роду*».

В узком смысле *качество объекта* – это совокупность его свойств, обусловливающих его пригодность удовлетворять определенные потребности в соответствии с его назначением.

Качество является предметом науки – **КВАЛИТОЛОГИИ**.

От «*gualitas*» – лат. «качество» (квалитет).

«Квалификация» – «качество специалиста».

КВАЛИМЕТРИЯ – отрасль науки, изучающая и реализующая методы количественной оценки качества.

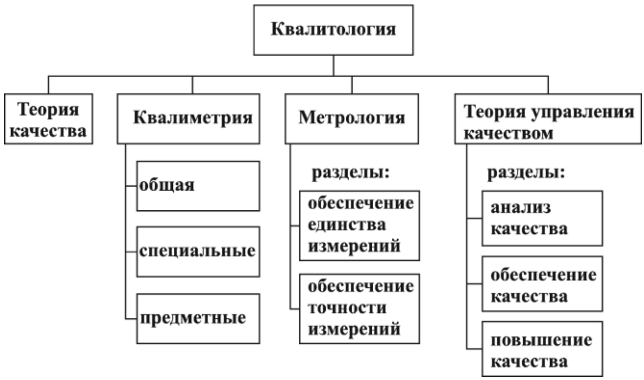


Рис. 4.2. Структура квалитологии

Рис. 4.3. Системы теорий квалитологии

МЕТРОЛОГИЯ – отрасль науки, изучающая и реализующая методы измерения качества.

ТАКСОНОМИЯ – теория классификации и систематизации сложноорганизованных объектов, имеющих обычно иерархическое строение (классификация и систематизация показателей и свойств, объектов оценки и т.д.).

4.3. ПОКАЗАТЕЛИ КАЧЕСТВА И ЭФФЕКТИВНОСТИ

Определение показателя качества и эффективности

Чтобы можно было сравнивать, сопоставлять, оценивать свойства объектов они должны быть *измеримыми*.

Измеримая характеристика свойства объекта, входящего в состав его качества, называется *показателем качества*[17].

Показателями качества технической системы являются ее измеримые свойства, характеристики, существенно влияющие на исход операции с применением этой ТС.

Следует обратить внимание на то, что набор показателей, составляющих качество ТС, для различных операций может быть различен. В этом случае, перед оценкой качества должно быть оговорено, для какой операции оценивается качество. Либо, без предварительной оговоренности, качество может оцениваться для типовой операции, для которой и была создана эта ТС.

Эффективность операции (реализации определенной стратегии в определенных условиях), как уже рассматривалось выше, оценивается вектором (кортежем) показателей $Y(u) = (y_1(u), y_2(u), \ldots, y_N(u))$.

[17] По ГОСТ 15467–79: это *количественная* характеристика одного или нескольких свойств продукции, входящих в ее качество, рассматриваемая применительно к определенным условиям ее создания и эксплуатации или потребления.

Показатели, существенным образом характеризующие результат операции, называются показателями эффективности операции.

То есть эффективность – это «качество» операции. Показатели эффективности операции суть показатели ее качества.

Скалярные и векторные показатели

Иногда качество объекта (эффективность операции) может быть охарактеризовано одним свойством, т.е. достаточно одного показателя. В этом случае говорят о ***скалярном показателе.***

Если же качество (эффективность операции) характеризуется некоторым множеством свойств, то отражаются они кортежем, *вектором показателей,* или *векторным показателем.*

Сложные ТС и результаты операции, как правило, характеризуются вектором показателей.

Выбор скалярного и каждой компоненты векторного показателя осуществляется исходя из следующих *требований*:

Соответствие – показатель должен соответствовать смыслу (существу) характеризуемого свойства.

Критичность – показатель должен быть достаточно чувствительным к изменению значений характеризуемого свойства.

Содержательность – показатель должен иметь «физический смысл», что упрощает анализ и интерпретацию полученных результатов и формулировку рекомендаций ЛПРу.

Вычислимость – значение показателя должно быть вычислимо (измеримо).

Требования к вектору показателей

Полнота – набор (вектор) из n компонент считается полным, если качество ясно и четко характеризуется совокупностью значений показателей и *введение дополнительных показателей не приводит к изменению оценки качества*;

Отсутствие избыточности, минимальность:

– исключение любого показателя существенно влияет на оценку качества.

– вектор должен иметь наименьшую возможную размерность. Это достигается исключением дублирования оценивания одного и того же свойства различными показателями.

Способы сведения векторных показателей к скалярным

Сравнение вариантов стратегий, ТС, если они характеризуются вектором показателей, сопряжено с определенными трудностями. Одним из способов преодоления этих трудностей является сведение векторного показателя к скалярному (свертку). Рассмотрим основные способы сведения векторного показателя к скалярному.

Если показатели однородны, то иногда возможно сведение векторного показателя к скалярному путем введения ***комплексного скалярного показателя*:**

Аддитивный комплексный скалярный показатель:

$$K(g) = \sum_{i=1}^{n} a_i K_i(g), \quad \sum_{i=1}^{n} a_i = 1,$$

где g – вариант; $K_i(g)$ – значение i-го частного показателя варианта g; a_i – нормированный весовой коэффициент, характеризующий относительную значимость i-го частного показателя.

К достоинствам способа следует отнести простоту и наглядность. Но он имеет ряд особенностей, которые следует учитывать. Частные показатели должны быть однородными (например, оценки, полученные обучающимся g на экзаменах по различным учебным дисциплинам). В исключительных случаях это правило может быть нарушено путем нормирования частных показателей, или введения размерности коэффициентов a_i, обратных размерностям частных показателей. Другой особенностью способа аддитивной свертки показателей является возможность компенсации малого значения одного частного показателя большим значением другого. Например, двое обучающихся сдавали по два экзамена. Первый обучающийся получил обе оценки «удовлетворительно», средний балл – 3,0. Второй обучающийся получил отличную и неудовлетворительную оценки, средний балл – 3,5. У второго обучающегося средний балл выше, чем у первого, хотя он не смог сдать второй экзамен, в то время как первый обучающийся успешно сдал оба экзамена. Если по условиям задачи компенсация значений частных показателей допустима, то этот способ может быть рекомендован к применению.

Если подобная компенсация не допустима, рекомендуется использовать мультипликативную свертку показателей.

Мультипликативный комплексный скалярный показатель:

$$K(g) = \prod_{i=1}^{n} K_i^{W_i}(g),$$

где W_i – весовые коэффициенты (не нормированные – натуральные числа больше единицы).

Частные показатели $K_i(g)$ должны быть однородными и больше единицы. Недопустимые значения частных показателей должны быть равны нулю. Тогда при наличии нулевого значения хотя бы у одного частного показателя значение комплексного мультипликативного показателя также будет равно нулю.

Недостатком рассмотренного способа свертки является его ненаглядность.

Если показатели неоднородны, и раскрывают свойства ТС, характеризующие качественно различные физические процессы и стороны системы, то

$$K_{\text{общ}}(g) = \sum_{i=1}^{n} \delta_i(g)$$

или

$$K_{\text{общ}}(g) = \prod_{i=1}^{n} \delta_i(g),$$

где δ_i – функции, определяемые по табличной или иной схеме преобразования. Здесь нет весовых коэффициентов, так как предполагается, что значимость частных показателей учтена в функции δ_i.

Аддитивный комплексный скалярный показатель с условием

Способ свертки [13] сочетает наглядность аддитивной свертки с возможностью учета недопустимых значений и сочетаний значений частных показателей. Может быть применен как для однородных, так и для неоднородных частных показателей

$$K(g) = P\big(k_1(g),\, ...,\, k_i(g),\, ...,\, k_n(g)\big) \cdot \sum\nolimits_i a_i k_i(g), \;\; \sum\nolimits_i a_i = 1,$$

$$K(g) = P\big(k_1(g),\, ...,\, k_i(g),\, ...,\, k_n(g)\big) \cdot \sum\nolimits_i \delta_i.$$

Предикат

$$P\big(k_1(g), ..., k_i(g), ..., k_n(g)\big) = \begin{cases} 0, \;\; \exists k_i(g) \in NK_i, \\ 0, \;\; \exists \otimes k_i(g) \in \otimes NK_i, \\ 1, \;\; \forall k_i(g) \in K_i \wedge \forall \otimes k_i(g) \in \otimes K_i, \end{cases}$$

где NK_i – множество недопустимых значений i-го частного показателя; K_i – множество допустимых значений i-го частного показателя; $\otimes k_i(g)$ – комбинация значений i-х частных показателей; $\otimes NK_i$ – множество недопустимых комбинаций значений i-х частных показателей; $\otimes K_i$ – допустимая комбинация значений i-х частных показателей.

Предикат $P(k_1(g))$ может быть также задан алгоритмически с помощью логических операторов.

Шкалы показателей качества и эффективности

Шкалы показателей качества (эффективности) (ПКиЭ) упорядочены по степени совершенства.

Введем функцию $\varphi(K(g))$, g – объект, $K(g)$ – ПКиЭ объекта g.

Определение. Если $\varphi(K(g))$, вновь оказывается показателем, измеряющим тот же признак, то это *допустимое преобразование показателя* $K(g)$.

Определение. Шкала считается тем более *совершенной*, чем у́же множество допустимых преобразований.

Шкалы показателей:

Номинальная (*классификационная*) *шкала* – указывается только, одинаковые варианты или нет (пример: различные анкеты). Множество допустимых преобразований – все взаимнооднозначные функции

$$\Phi_н = \{\varphi \mid \forall x \neq y \Rightarrow \varphi(x) \neq \varphi(y)\}.$$

Порядковая шкала – оценки вариантов сравниваются только на «больше-меньше». Сравнивать какие-либо интервалы (расстояния) между оценками бессмысленно. Пример: твердость материалов. Множество допустимых преобразований – все монотонно возрастающие функции:

$$\Phi_п = \{\varphi \mid \forall x > y \Rightarrow \varphi(x) > \varphi(y)\}.$$

Шкала интервалов – в ней сохраняются отношения длин интервалов (отношения расстояний между оценками). Пример: дата некоторого события – нужно начало отсчета и масштаб – годы, дни, минуты и т.д. Множество допустимых преобразований (здесь и выше – количественные показатели) – все линейные числовые функции вида:

$$\Phi_{\text{и}} = \{\varphi \,|\, \varphi\,(x) = kx + l,\ k > 0,\ l \neq 0\},$$

$$\frac{K(a)-K(b)}{K(c)-K(d)} = \frac{\left(kK(a)+l\right)-\left(kK(b)+l\right)}{\left(kK(c)+l\right)-\left(kK(d)+l\right)} = \text{const}.$$

Шкала отношений – сохраняется отношение оценок. В этой шкале измеряются все физические величины, имеющие размерность.

Пример: длина ракеты, высота полета, ускорение, сила – можно менять единицу измерения.

$$\Phi_{\text{о}} = \{\varphi \,|\, \varphi\,(x) = kx,\ k > 0\}.$$

Абсолютная шкала – в ней измеряется абсолютное количество объектов. Единственное допустимое преобразование – тождественное:

$$\Phi_{\text{а}} = \{\varphi \,|\, \varphi\,(x) \equiv x\}.$$

Классификация показателей качества ТС

Признаки классификации:

– по характеризуемым свойствам;
– по способу выражения;
– по количеству характеризуемых свойств;
– по применению для оценки;
– по стадии определения.

1. По характеризуемым свойствам:

1.1. Показатели назначения.

1.2. Показатели надежности (безотказности, долговечности, сохраняемости, ремонтопригодности).

1.3. Эргономические показатели.

1.4. Эстетические показатели.

1.5. Показатели технологичности.

1.6. Показатели транспортабельности.

1.7. Показатели стандартизации и унификации.

1.8. Патентно-правовые показатели.

1.9. Экологические показатели.

1.10. Показатели безопасности.

1.11. Экономические показатели.

1.12. Показатели утилизируемости.

2. По способу выражения: характеризуются шкалами показателей, в частности, измеряемые в натуральных единицах (кг, м, с), безразмерные, в частности – баллы, стоимостные единицы, абсолютные и относительные и т.д.

Сюда же можно отнести *детерминированные* и *стохастические*.

3. По количеству характеризуемых свойств:

– *единичные,*

– *комплексные*, характеризующие несколько свойств объекта.

Пример: коэффициент готовности $K_г$, равный вероятности работоспособности ИС, характеризует сразу безотказность и ремонтопригодность:

$$K_г = \frac{T}{T + T_в},$$

где T – наработка на отказ; $T_в$ – время восстановления.

– *интегральные,* относятся к комплексным, отражают соотношение суммарного полезного эффекта от использования к суммарным затратам на его разработку, изготовление и эксплуатацию;

– *определяющие*: те, по которым принимаются решения о качестве, как правило, комплексные, но могут быть и единичные, отражающие наиболее важные свойства.

4. По применению для оценки:

– *базовые*: эталонные, гипотетические;

– *относительные*, относительно к эталону.

5. По стадии определения:

– прогнозируемые;

– проектные;

– производственные;

– эксплуатационные.

Иерархическая структура показателей качества. В квалитологии качество рассматривается как некоторая иерархическая совокупность свойств, расположенных на различных уровнях Оценка каждого свойства на любом уровне определяется совокупностью оценок соответствующих свойств более низкого уровня.

Показатели качества

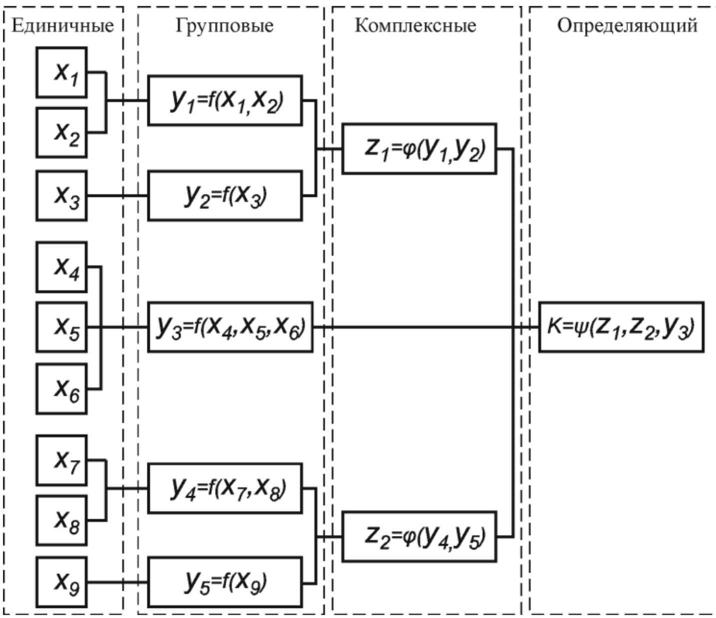


Рис. 4.4. Иерархическая структура показателей качества

Методы определения величины показателей качества

Методы определения величины показателей качества зависят от конструкторских, технологических и эксплуатационных особенностей продукции:

– инструментальные, с использованием различных измерительных и контрольных приборов;

– расчетно-аналитические, путем расчета показателей и установления взаимосвязи между ними;

– опытные, позволяющие путем испытаний установить, а в отдельных случаях и проверить, значение показателей, найденных другими методами (например, испытание автомобилей на полигоне, ускоренные испытания двигателей и т.д.);

– лабораторные, служащие для определения показателей с помощью анализов и испытаний;

– органолептические, заключающиеся в определении показателей с помощью органов чувств (например, контроль окраски, наличие царапин и т.д.);

– социальные, позволяющие определить качество путем анкетного опроса потребителей;

– экспертные, с использованием экспертов в анкетных опросах, с целью получения более точных значений величины показателя.

4.4. ОСНОВЫ ТЕОРИИ ПРИНЯТИЯ РЕШЕНИЙ

Предмет теории принятия решений

При проектировании сложных технических систем, создании сложных промышленных комплексов и управлении ими, а также в других сферах деятельности человека *возникают задачи принятия наилучших решений*. Для решения этих прикладных задач были созданы и в настоящее время разрабатываются специальные математические модели и методы, которые принято объединять под названием «Теория принятия решений». Таким образом, *ТПР* – это наука, изучающая методы принятия оптимальных решений в различных областях целенаправленной деятельности человека[18].

Теория принятия решений – область научных исследований, предметом изучения которой являются закономерности выбора людьми путей решения проблем и задач, а также способов достижения желаемого результата.

Принятие решения заключается в выборе (синтезе) варианта (действия, ТС, ее режима функционирования и т.д.) из множества возможных.

Википедия выделяет следующие этапы принятия решений:

1. Ситуационный анализ (анализ проблемной ситуации).
2. Идентификация проблемы и постановка цели.
3. Поиск необходимой информации.
4. Формирование множества возможных решений.
5. Формирование критериев оценки решений.
6. Разработка индикаторов и критериев для мониторинга реализации решений.
7. Проведение оценки решений.

[18] Теория принятия решений [Электронный ресурс]: конспект лекций по дисциплине «Теория принятия решений» для студентов бакалавриата очной формы обучения направления подготовки 27.03.04 Управление в технических системах / М-во образования и науки Рос. Федерации, Нац. исследоват. Моск. гос. строит. ун-т, каф. электротехники и электропривода; сост.: С.В. Шилкина, О.Л. Широкова. Электрон. дан. и прогр. (1,1 Мб). Москва: НИУ МГСУ, 2015. Учебное сетевое электронное издание. Режим доступа: http://lib.mgsu.ru/ Scripts /irbis64r_91/cgiirbis_64.exe?C21COM=F&I21DBN=IBIS&P21DBN=IBIS. Загл. с титул. экрана.

8. Выбор наилучшего решения.
9. Планирование.
10. Реализация.
11. Мониторинг реализации.
12. Оценка результата.

Математическая модель проблемной ситуации. Классификация задач принятия решений. Модель проблемной ситуации

Проблема принятия решения связана с выбором варианта действий для достижения целей операции.

Такой выбор осуществляет ЛПР – лицо, принимающее решение, которое наделено определенными правами и полномочиями и несет всю полноту ответственности за последствия принимаемых решений.

Переходным этапом от проблемы к постановке формальных задач является *проблемная ситуация,* в ходе которой вербальная цель разбивается на подцели – задачи, устанавливаются общие ограничения, условия решения.

Для того чтобы охватить проблему принятия решения в целом, представить ее основные элементы, которые необходимо сформировать для получения окончательного решения о стратегии проведения операции, применяют математическую *модель проблемной ситуации* (*МПС*). Эта модель отображает взаимосвязи основных элементов процесса выработки решения и последовательность формирования частных задач.

Результатом анализа модели проблемной ситуации является *формальная постановка задачи* исследования, которая указывает, каких результатов, в каких условиях и к какому сроку требуется достичь.

Рассмотрим общую схему МПС

Для достижения поставленных целей ЛПР имеет в своем распоряжении некоторый запас *активных средств, ресурсов*.

Эти ресурсы могут быть использованы различными способами, которые называются *стратегиями*. Реализация стратегии приводит к некоторому результату, который называется *исход*. Исход, в общем случае, зависит также и от независимых от ЛПР-факторов.

Эти факторы могут быть *определенные*, т.е. их значения известны, и *неопределенные,* т.е. их значения до реализации стратегии неизвестны.

Неопределенные факторы могут быть *случайными* (закон их распределения известен или неизвестен), или неопределенные *нестохастического характера* (*природные* – из-за недостатка изученности каких-либо явлений или *противодействующие* (*стратегические*), обусловленные наличием людей, преследующих свои цели).

При различных исходах цели, преследуемые ЛПР, достигаются в различной степени, поэтому, с его точки зрения, исходы различаются по предпочтительности, т.е., *на совокупности исходов имеет место некоторая система предпочтений ЛПР*.

При наличии лишь случайных факторов с известными распределениями каждая стратегия может привести к различным исходам с соответствующими вероятностями, т.е. каждой стратегии соответствует некоторое распределение вероятностей исходов. Эти распределения также имеют систему предпочтений ЛПР, отражающую *отношение ЛПР к риску*.

С учетом всего сказанного *математическая модель проблемной ситуации,* которая должна отражать формально все перечисленные компоненты, может быть представлена в виде кортежа:

$$D = < U, G, \Lambda, \Psi, P, \Theta >, \tag{4.1}$$

где U – множество стратегий, $|U| \geq 2$, иначе не было бы возможности выбора стратегии; Λ – множество возможных значений влияющих факторов, $|\Lambda| \geq 1$; G – множество исходов, $2 \leq |G| \leq |U| \times |\Lambda|$, иначе все исходы были бы эквивалентны, равноценны; Ψ – отображение, ставящее в соответствие фиксированной стратегии u и значению влияющего фактора λ некоторый исход g:

$$g = \Psi\ (u, \lambda), \text{ т.е. } \Psi\colon U \times \Lambda \longrightarrow G, \tag{4.2}$$

P – модель предпочтений ЛПР на множестве G; Θ – вся остальная имеющаяся информация о проблемной ситуации.

Классификация задач принятия решений (ЗПР)

По признаку *кто выступает в роли ЛПР*:

– индивидуального решения (ЛПР – отдельный человек);
– группового решения (ЛПР – коллектив людей):
– с совпадающими целями;
– с несовпадающими целями:
– с одним, компромиссным решением;
– с индивидуальными решениями;

По признаку *характера обстановки*:

– принятие решения в условиях *определенности* – неопределенные факторы отсутствуют;

– принятие решения в условиях *риска* (стохастической неопределенности) – имеются только случайные факторы с известными законами распределения;

– принятие решения в условиях *неопределенности* (поведенческой неопределенности) – имеются только случайные факторы с неизвестными законами распределения или природные неопределенные факторы;

– принятие решения в условиях *противодействия* (стратегической неопределенности) – при наличии активно действующих противников;

– принятие решения в условиях *конфликта* (ЛПР – групповой с несовпадающими целями его элементов).

Описание предпочтений

Предположим, что имеется совокупность (множество) D некоторых объектов и требуется сравнить их между собой по предпочтительности для ЛПР.

Формально задача ставится следующим образом.

Пусть дано произвольное множество D.

Под *предпочтением* понимается любая форма упорядочения элементов множества D.

Выражение (оценивание) предпочтений может осуществляться следующими способами:

В *балльных шкалах* часто производятся экспертные оценки. В качестве шкалы, как правило, используется ряд целых чисел.

Сортировка – разбиение множества элементов на непересекающиеся классы (по степени предпочтительности), в пределах которых элементы равноценны.

Ранжирование – это представление объектов в виде последовательности в соответствии с убыванием их предпочтительности. При этом допускается наличие нескольких равноценных объектов (в противном случае – *строгое ранжирование*).

Попарное сравнение состоит в указании более предпочтительного объекта в каждой паре объектов. При этом часть объектов может быть равноценными или несравнимыми. Результаты сравнения удобно представлять в виде таблицы.

Здесь не требуется выполнение транзитивного закона. Поэтому такая оценка составляется проще чем предыдущие две, и может быть составлена, когда предыдущие невозможны.

	a	b	c
a	1/2	1	–
b	0	1/2	1
c	–	0	1/2

Отношения предпочтения

Отношение *строгого предпочтения* aPb ($a \succ b$).

Отношение *безразличия* aIb ($a \sim b$).

Отношение *нестрогого предпочтения* aRb ($a \succcurlyeq b$).

Отношение *несравнимости* aNb.

Не следует путать отношения безразличия и несравнимости: безразличие означает одинаковость по предпочтению ЛПР, несравнимость предполагает невозможность сравнить варианты.

В теории принятия решений отношения предпочтений строят на множествах:

– исходов операций G;
– стратегий U, приводящих к исходам g, $g \in G$;
– показателей качества K технических систем.

В ваших работах надо, имея отношение предпочтений на множестве исходов, получить отношение предпочтений на множестве стратегий. Здесь следует различать задачи: – нахождения одного (возможно – из нескольких) решения; – нахождения всего множества решений; – разработки методики для нахождения первого или второго.

4.5. КРИТЕРИИ ОПТИМАЛЬНОСТИ

С понятием качества ТС и эффективности их применения решаются две группы задач:

Задачи *анализа* – где определяется качество объекта (или эффективность его применения – стратегия);

Задачи *синтеза* – где из множества вариантов выбираются наиболее предпочтительные, лучшие, оптимальные, наиболее эффективные или просто пригодные технические объекты (или стратегии).

Из определения качества объекта следует, что он предназначен для выполнения какой-либо функции (задачи, операции), или предполагается его целевое применение.

Из этого следуют выводы:

1. При выполнении различных задач перечень свойств, входящих в качество может быть различен, и, соответственно, вектор показателей тоже может быть различным.
2. Различающиеся по качеству объекты (по эффективности – стратегии) могут более или менее предпочтительными для ЛПР.
3. При решении различных задач предпочтительными могут быть разные объекты (стратегии).

Различные варианты целенаправленной деятельности (стратегии), а также варианты технических систем при использовании их в операции (далее – варианты), могут быть более или менее предпочтительными (приоритетными) с точки зрения ЛПР на выбор варианта стратегии или варианта технической системы.

Определение. *Правило,* формально определяющее, какой вариант (стратегии, технической системы, исхода) более предпочтителен (приоритетен), *называется критерием* (эффективности, оптимальности, качества).

Между кортежем показателей и критерием существует непосредственная связь: в критерии определяются требуемые, допустимые или предпочтительные значения показателей или их комбинаций. То есть в критерии не должно быть требований к характеристикам вариантов, не входящим в кортеж показателей их качества (эффективности), а в кортеж качества (эффективности) не имеет смысла включать показатели, к значениям которых не предъявляются требования в критерии.

Наряду с приведенной системой терминов («показатель – характеристика», «критерий – правило») существует система терминов [21, 22, 23]: «критерий – характеристика», «принцип – правило». В научном труде следует придерживаться одной из этих систем и избегать их смешения. В этом отношении частой ошибкой является называние «критерием» и характеристик, и правил приоритета.

Существуют две концепции выбора рациональных решений: концепция *пригодности* и концепция *оптимальности*.

В контексте концепции *пригодности* решением оптимизационной задачи является *допустимый* (*пригодный*) *вариант* u^* из множества возможных вариантов U. Допустимым, по определению, считается вариант u^*, значения *всех* показателей качества (эффективности) которого принадлежат множествам допустимых значений этих показателей. Допустимых вариантов может быть несколько, они образуют множество допустимых вариантов U^*, $U^* \subseteq U$, $u^* \in U^*$, множество решений. Формально критерий может быть записан следующим образом:

$$U^* = \left\{ u_i^* \middle| \forall i, j \;\; u_i^* \in U, \, K_j(u_i^*) \in K_j^0 \right\}, \qquad (4.3)$$

где $K_j(u_i)$ – значение j-го показателя i-го варианта; K_j^0 – множество допустимых значений j-го показателя варианта.

Для непрерывных множеств

$$U^* = \left\{ u^* \middle| \forall j \;\; u^* \in U, \, K_j(u^*) \in K_j^0 \right\}. \tag{4.4}$$

В контексте концепции *оптимальности* решением является *наилучший,* наиболее предпочтительный в определенном смысле вариант u^* из множества возможных вариантов U. Этот «наилучший» вариант называется *оптимальным* («наилучших» – оптимальных вариантов, в принципе, может быть множество).

Существенным вопросом при математической постановке задачи (МПЗ) является вопрос *существования решения сформулированной задачи*. Существует распространенное заблуждение, что при концепции *пригодности* возможность существования решения гораздо больше, чем при концепции *оптимальности*. В последнем случае, даже если существует всего одна альтернатива, решение существует – эта самая альтернатива. То есть решение существует. Всегда. В случае концепции пригодности наличие решений и их количество зависят от системы ограничений, требований к значениям показателей. В любом коллективе (даже если он состоит из одного человека) можно найти самого здорового члена коллектива, но даже в большом коллективе может не оказаться ни одного человека, годного по состоянию здоровья для принятия в отряд космонавтов. Поскольку в большинстве практических случаев в МПЗ так или иначе используются ограничения на значения показателей альтернатив, необходимо следить за наличием решений, обосновывать их существование при данной системе ограничений.

В рассмотренных выше концепциях выбор решения осуществляется однократно. Но существует широкий класс ЗПР, в которых выбор решения осуществляется многократно, периодически или непрерывно, в зависимости от изменяющейся обстановки: изменения множества возможных вариантов, допустимых значений показателей, приоритетов и т.п. – *это задачи управления*. Эти критерии иногда выделяют в отдельную концепцию – *адаптивизации*. В этом случае следует помнить, что *адаптивизации* может быть использована и в концепции *пригодности* и в концепции *оптимальности*. Ниже приведено примерное выражение для критерия в концепции пригодности с учетом адаптивизации.

$$U^*(t) = \left\{u_i^*(t) \middle| \forall i, j \;\; u_i^*(t) \in U(t),\, K_j(u_i^*(t), t) \in K_j^0(t)\right\}.$$

Таким образом, можно выделить четыре концепции выбора рациональных решений:

1. Концепция *пригодности* в однократном выборе решения;
2. Концепция *оптимальности* в однократном выборе решения;
3. Концепция *пригодности* в многократном выборе решения;
4. Концепция *оптимальности* в многократном выборе решения выборе решения.

Вернемся к концепции *пригодности*.

Для скалярного показателя

$$U^* = \left\{u_i^* \middle| \forall i \;\; u_i^* \in U,\, K(u_i^*) \in K^0\right\}.$$

Для векторного

$$U^* = \left\{u_i^* \middle| \forall i, j \;\; u_i^* \in U,\, K_j(u_i^*) \in K_j^0\right\}.$$

А как в концепции *оптимальности*?

Если показатель *скалярный*, то

$$U^* = \left\{u^* \middle| K(u^*) = extr_{u \in U} K(u)\right\}, \qquad (4.5)$$

или

$$U^* = \left\{u^* \middle| u^* = \arg\max_{u \in U} K(u)\right\},$$

$$U^* = \left\{u^* \middle| u^* = \arg\min_{u \in U} K(u)\right\}, \qquad (4.6)$$

А если показатель *векторный*?

Можно ли записать

$$U^* = \left\{ u^* \mid \forall j \; K_j(u^*) = extr_{u \in U} K_j(u) \right\} \qquad (4.7)$$

Решением в большинстве случаев будет $U^* = \varnothing$.

Здесь возникает задача «многокритериальной» оптимизации.

Если используется система терминов: «критерий–характеристика», «принцип–правило», то под «многокритериальной оптимизацией» понимается факт того, что вариант характеризуется не одним – скалярным, а несколькими параметрами – «критериями».

Если используется применяемая в настоящем материале система терминов «показатель–характеристика», «критерий–правило», то под «многокритериальной оптимизацией» понимается факт того, что действительно имеется множество независимых правил того, что считать «лучшим» вариантом.

В этом случае практически неизбежно возникает конфликт решений, так как по одному из правил («критериев») лучшим будет один из возможных вариантов, по другому правилу («критерию») – другой вариант[19].

Выходом из этой ситуации является согласование требований критериев и формулирование единого комплексного критерия, представляющего собой компромисс требований частных критериев и нахождение решения по этому комплексному критерию.

Способов сведения множества критериев в единый критерий существует достаточно много, по ним имеется доступная литература [1, 12, 13, 18, 20 – 25].

---

[19] В тех редких ситуациях, когда решения по различным критериям всегда совпадают, достаточно использовать один критерий, так как найденное по нему решение будет решением и по другим критериям.

Наиболее часто применяемые:

1. Сведение векторного показателя к скалярному и решение задачи скалярной оптимизации.

2. Формирование критерия «главного показателя».

Критерий главного показателя (*субоптимизация*):

$$U^* = \left\{ u^* \middle| K(u^*) = extr_{u \in U} K(u),\ \forall j \neq i\ K_j(u^*) \in K_j^0 \right\}.$$

3. Лексикографическая оптимизация. Здесь критерии ранжируются по значимости. Находится решение, оптимальное по наиболее значимому критерию. Если таких окажется несколько, то среди них выбирается оптимальный вариант по второму по значимости критерию и т.д. Здесь же уместны и критерии уступок.

4. *Критерии уступок*. Предположим, что компоненты вектора показателей эффективности (качества) расположены в порядке убывающей важности: k_1, k_2, … . Будем считать, что предпочтительными будут наибольшие значения каждого из них (если это не так, достаточно изменить знак показателя).

Процедура построения векторного критерия оптимальности сводится к следующему.

Сначала ищется решение, обращающее в максимум значение первой компоненты k_1. Затем назначается, исходя из практических соображений и точности, с какой известны исходные данные (а часто она бывает небольшой), некоторая «уступка» Δk_1, которую ЛПР согласно допустить для того, чтобы обратить в максимум значение второй компоненты k_2. Налагаем на значение k_1 ограничение, чтобы оно было не меньше, чем $k_1^* - \Delta k_1$, где k_1^* – максимально возможное значение k_1, и при этом ограничении ищем решение, обращающее в максимум k_2. Далее снова назначается «уступка» в показателе k_2, ценой которой можно максимизировать k_3, и т.д.

Такой способ построения предпочтительного решения хорош тем, что здесь сразу видно, ценой какой «уступки» в одном показателе приобретается выигрыш в другом. Надо сказать, что свобода выбора решения, приобретаемая ценой даже незначительных «уступок», может оказаться существенной, так как в районе максимума обычно эффективность решения меняется очень слабо.

Различают следующие виды уступок: *абсолютные, относительные,* «*справедливые*».

В методе «справедливых уступок» сумма снижений значений одних показателей в абсолютных значениях не должна превышать суммы прироста значений других показателей.

Одним из частных случаев метода «справедливых уступок» является применение принципа равенства, т.е. когда значимость всех компонент вектора показателей одинакова.

5. *Критерий Парето*: предназначен для того, чтобы проверить, улучшает ли предложенное изменение в экономике общий уровень благосостояния. Основывается на правиле: «Следует считать, что любое изменение, которое никому не причиняет убытков и которое приносит некоторым людям пользу, является улучшением». Он применяется при решении таких задач, когда оптимизация означает улучшение одних показателей при условии, чтобы другие не ухудшались.

Определение. Вариант u_2 лучше (предпочтительнее) варианта u_1:

$$u_2 \succ u_1 \Leftrightarrow \forall i\, K_i(u_2) \geq K_i(u_1),\ \exists j\, K_j(u_2) > K_j(u_1).$$

Критерий. Вариант u^*, оптимальный по Парето:

$$u = u^* \Leftrightarrow \nexists\, u_i,\ u_i \succ u^*.$$

В общем виде критерий оптимальности может быть записан следующим образом. Пусть вариант u решения характеризуется кортежем показателей $P(u) = \langle p_1(u), \ldots, p_i(u), \ldots, p_I(u)\rangle$. Тогда вариант u^* является оптимальным (решением оптимизационной задачи), если

$$u^* = \arg\max_{u \in U} J_0(P(u)), \tag{4.8}$$

$$J_n(P(u^*)) \in J_n^0, \ n = \overline{1, N}, u^* \in U, \tag{4.9}$$

где $J_0(P(u))$ – некоторый функционал от кортежа параметров $P(u)$, а система (4.9) определяет ограничения на допустимые значения и совокупности значений параметров $P(u^*)$. Следует отметить, что функции

$$J_n(P(u)), \ n = \overline{1, N}$$

не обязательно должны быть числовыми, и даже не обязаны быть функционалами. А $J_0(P(u))$ может быть не функционалом, а функцией, значения которой принадлежат упорядоченному множеству.

Применение концепции оптимизации накладывает на исследователя одно важное обязательство, которое не всегда им выполняется. Если исследователь предлагает методику (алгоритм) нахождения решения u^* для выражения (4.8), то он обязан доказать, что при выполнении ограничений (4.9) не существует решения u', такого, что $J_0(P(u')) > J_0(P(u^*))$. Часто же, исследователь предлагает решение, лучшее по сравнению с существующими $u_{\text{сущ}}$, но не гарантирует, что никакой другой исследователь никогда не сможет предложить методику (алгоритм), дающий результат еще лучше. Такие решения приемлемы, но неприемлемо применение (4.8), а следует применять отношение предпочтения «$\prec$», «$\succ$». Например,

$$u^* \succ u_{\text{сущ}} \Leftrightarrow J_0(P(u^*)) > J_0(P(u_{\text{сущ}})). \tag{4.10}$$

Следует обратить внимание на то, что в одних исследованиях необходимо найти оптимальное решение Y^* для конкретного заданного набора исходных данных $X = \langle x_1, \ldots, x_i, \ldots, x_I\rangle$,

$$X \longrightarrow Y^*, \tag{4.11}$$

а в других (таких большинство) разработать методику M или алгоритм A нахождения оптимальных решений для задаваемых исходных данных

$$M: X \longrightarrow Y^*, \quad A: X \longrightarrow Y^*.$$

Здесь частой ошибкой является применение выражения (4.11) для задачи разработки методики (алгоритма).

Задача разработки методики (алгоритма) тоже может являться оптимизационной, например, разработка алгоритма решения некоторой задачи, обеспечивающей требуемую точность решения при минимальной вычислительной сложности.

Принятие решений или оптимизация?

При МПЗ часто возникает вопрос: решаемая задача относится к «*задаче оптимизации*» или к «*задаче принятия решений*». Оба термины семантически близки, так как задача оптимизации в определенном смысле является задачей принятия решений и, наоборот, и в настоящее время не существует их единого четкого разграничения.

Например, в [25] дано определение «*Принятие решений*» в профессиональном отношении представляет собой особый вид человеческой деятельности, который состоит в обоснованном выборе наилучшего в некотором смысле варианта или нескольких предпочтительных вариантов из имеющихся возможных. По-английски этот термин звучит как decision making, т.е. буквально означает «делание» или создание решения…».

Про задачи оптимизации, они же «экстремальные задачи» в [1] написано «Задачи на отыскание наибольших и наименьших величин являются актуальными на протяжении всей истории человечества. Особенное значение они приобретают в настоящее время, когда возрастает важность наиболее эффективного использования природных богатств, людских ресурсов, материальных и финансовых средств. Все это приводит к необходимости отыскивать наилучшее, или, как говорят, *оптимальное* решение того или иного вопроса».

Анализ источников позволяют сделать вывод о том, что критерием различия рассматриваемых задач является вид функционала $J_0(P(u))$ в выражении (4.8), степень формальности его математического описания, «математической строгости». Если он задан аналитически, еще лучше, если в виде числовой дифференцируемой функции, и допускает решение математическими методами, позволяющими получить строго оптимальные решения (градиентными, методами математического программирования и т.п.), то речь идет о задаче оптимизации. Если же $J_0(P(u))$ представлен алгоритмически, вычислительно-сложным алгоритмом, с дискретным и, возможно, относительно небольшим множеством значений u, или с помощью экспертных оценок – речь идет в большей мере о задаче принятия решений.

Выводы:

1. Правильность и сама возможность решения задачи выбора оптимального решения в существенной степени зависит от обоснованности выбора показателей эффективности и критериев оптимальности.
2. Не существует единого алгоритма решения ЗПР, однако существуют алгоритмы решения частных задач оптимизации.
3. Решение ЗПР предполагает значительный объем вычислений, поэтому целесообразна автоматизация этого процесса, т.е. использование ЭВМ.

4.6. КРИТЕРИИ ПРИНЯТИЯ РЕШЕНИЙ В УСЛОВИЯХ ОПРЕДЕЛЕННОСТИ И РИСКА

Принятие решений в условиях определенности

Постановка задачи. Задача выработки решений в условиях определенности характеризуется отсутствием случайных и неопределенных факторов.

Поэтому каждая стратегия u, $u \in U$ приводит к вполне определенному исходу $g = \psi\ (u)$ и схема модели проблемной ситуации приобретает вид:

$$< G, U, \Psi, P >, \qquad (4.12)$$

где U – множество стратегий; G – множество исходов; Ψ – отображение, ставящее в соответствие фиксированной стратегии u некоторый исход g, $g = \Psi(u)$; P – структура предпочтений ЛПР на G.

Структура предпочтений ЛПР представляется отношениями *нестрогого предпочтения* R, *строгого предпочтения* P и *безразличия* I на множестве исходов G.

Но, поскольку каждой стратегии соответствует некоторый исход, то отношения предпочтений переносятся и на множество стратегий U:

$$uRv \text{ при } \Psi(u)R\Psi(v),$$

$$uPv \text{ при } \Psi(u)P\Psi(v),$$

$$uIv \text{ при } \Psi(u)I\Psi(v).$$

Принятие решений в условиях риска

Постановка задачи. Задача выработки решения в условиях риска (стохастической неопределенности) характеризуется наличием случайных факторов с известными законами распределения вероятностей. Сведения об этих законах составляют информацию Θ.

Математическая модель проблемной ситуации в этом случае может быть представлена в виде кортежа:

$$D = < U, \Lambda, G, \Psi, P, \Theta >, \qquad (4.13)$$

где U – множество стратегий; Λ – множество значений влияющих факторов; G – множество исходов; Ψ – отображение, ставящее в соответствие стратегии u и значению влияющего фактора λ некоторый исход g, $g = \Psi(u, \lambda)$ или Ψ: $U \times \Lambda \longrightarrow G$; P – модель предпочтений ЛПР на множестве G; Θ – информация о законе распределения вероятностей значений λ, $\lambda \in \Lambda$.

Поскольку на результат операции помимо выбранной стратегии влияют случайные факторы, ее исход будет случайным.

Поэтому на практике для решения задач оптимизации в условиях риска широко используются специальные критерии оптимальности, заключающиеся в том, что случайная функция $K_c(u)$ тем или иным путем преобразуется в неслучайную, при помощи которой все стратегии сопоставляются по предпочтительности и среди них выделяется оптимальная.

Эти критерии задают стратегии, *оптимальные в стохастическом смысле*, т.е. они не гарантируют каких-либо результатов в каждой отдельно взятой реализации, а эффективны в среднем при многократном применении.

Рассмотрим построение основных из них.

Критерий стохастического доминирования

Знание законов распределения случайных факторов позволяет *каждой стратегии поставить в соответствие распределение вероятностей на множестве исходов*.

Пример: если $G = \{g_1, g_2, ..., g_n\}$,
тогда $P_i(u)$ – вероятность появления i-го исхода при стратегии u.

$$P(u) = \{P_1(u), P_2(u), \ldots, P_n(u)\}.$$

Если каждый исход характеризуется показателем эффективности $K(g)$, то он является случайной функцией стратегии $K_c(u)$.

Для любой выбранной стратегии u распределение случайной величины $K_c(u)$ известно. Его можно записать при помощи *функции распределения*

$$F(x, u) = P(K_c(u) < x).$$

Если K_c (u) – дискретная, то F(x, u) задается рядом распределения. Если непрерывная, то используют *функцию распределения*.

Предположим, что для двух стратегий u и v:

$$F(x, u) \le F(x, v) \mid \forall x \in (-\infty, +\infty).$$

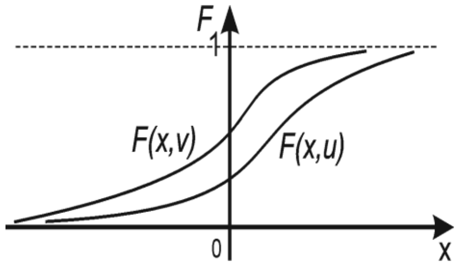


Рис. 4.5. Пример отношения стохастического доминирования стратегий

В соответствии с определением функции распределения

$$P(K_c\,(u) \ge x) > P(K_c\,(v) \ge x)\ \forall\ x.$$

Если $\forall$ x: $F(x, u) < F(x, v)$, то вводим строгое предпочтение P: uPv.

Определение. Отношение P называют *отношением стохастического доминирования*. Оптимальной будет недоминируемая стратегия.

К сожалению, для произвольных стратегий u и v графики F(x, u) и F(x, v) обычно пересекаются и стратегии становятся несравнимы по отношению P.

Критерий вероятностно гарантированного результата

Для выбранной стратегии u распределение $K_c(u)$ известно. Зададимся числом β и определим вероятность

$$P(K_c\,(u) \ge \beta) = \gamma\ (\beta, u).$$

Эта вероятность есть функция от β и u.

Наоборот, зададимся числом γ: $0 < \gamma < 1$. Тогда существует число $\beta(\gamma, u)$, такое что

$$P(K_c(u) \geq \beta) \geq \gamma.$$

Критерий *вероятностно гарантированного результата* заключается в том, что случайная функция $K_c(u)$ заменяется неслучайной функцией одним из следующих способов:

– задается «требуемый» уровень результата β и вводится числовая функция

$$K_\beta(u) = \gamma\,(\beta, u);$$

– задается «требуемая» степень гарантии (вероятность) γ и вводится числовая функция

$$K_\gamma(u) = \beta(\gamma, u).$$

Выражения для критерия:

При заданном β найти $u^* = \arg\max_{u \in U} K_\beta(u)$, $K_\beta(u)$ – это вероятность превышения β. Или при заданном γ найти $u^* = \arg\max_{u \in U} K_\gamma(u)$, $K_\gamma(u)$ – это число при заданной вероятности γ.

Критерии среднего результата и кучности

Рекомендуются к применению в случаях, когда выбираемые действия (операции, стратегии) предполагается производить многократно, причем *трудно, или невозможно производить вычисления непосредственно с законами распределений*, но известны числовые характеристики законов распределений.

Критерий заключается в нахождении стратегий, обеспечивающих экстремум соответствующей числовой характеристики закона распределения. При этом, на значения других характеристик, как правило, накладываются ограничения (см. критерий главного показателя).

Так, средний результат – МО $K_c(u)$ – $M[K_c(u)]$, кучность – дисперсия $K_c(u)$ – $D[K_c(u)]$.

И две формы критерия, в зависимости от того, что оптимизируется, а на что накладывается ограничение, МО или дисперсия:

$$u^* = \arg\max_{u \in U} M[K_c(u)],\ D[K_c(u)] \le D^0$$

или

$$u^* = \arg\max_{u \in U} D[K_c(u)],\ M[K_c(u)] \in M^0\ (example\ M^0 = 0).$$

Критерий приемлемого результата

Пусть $K_c(u) \in \{0,1\}$, т.е. все множество исходов делится на два подмножества: приемлемые исходы и неприемлемые.

В этом случае можно использовать выше изученные критерии, положив $\beta = 1$ и использовав *критерий вероятностно гарантированного результата*

$$u^* = \mathrm{argmax}_{u \in U} K_1(u),\ K_1(u) –$$

это вероятность получения $\beta = 1$, или использовав критерий *среднего результата*

$$u^* = \arg\max_{u \in U} M[K_c(u)],\ D[K_c(u)] \le D^0,$$

здесь $M[K_c(u)] \ne 1$, $0 \le M[K_c(u)] \le 1$.

Критерий Неймана–Пирсона

Пусть X, $X \in \boldsymbol{X}$ – множество исходных данных, по которым по некоторому критерию f принимается решение о верности одной из гипотез H_0 или H_1

$$f\colon X \to H_i,\ i \in \{0, 1\}.$$

Принято считать, что нулевая гипотеза H_0 соответствует состоянию «по умолчанию» (естественному, наиболее ожидаемому положению вещей) – например, что обследуемый человек здоров, или что проходящий через рамку металлодетектора пассажир не имеет запрещенных металлических предметов.

Соответственно, альтернативная гипотеза H_1 обозначает противоположную ситуацию, которая обычно трактуется как менее вероятная, неординарная, требующая какой-либо реакции.

Тогда

$$f\colon X \to H,\ H \in \{\ H_0\ ,\ \overline{H_0}\ \}.$$

Ошибки первого и второго рода

При этом при применении решающего правила f возможны ситуации:

1. Исходные данные соответствуют гипотезе H_0, и она точно определена критерием f, т.е. $f(X) = H_0$.
2. Исходные данные соответствуют гипотезе H_0, но она неверно отвергнута критерием f, т.е. $f(X) = H_1$. Это *ошибка первого рода* «*ложная тревога*».
3. Исходные данные соответствуют гипотезе H_1, и она точно определена критерием f, т.е. $f(X) = H_1$.
4. Исходные данные соответствуют гипотезе H_1, но она неверно отвергнута критерием f, т.е. $f(X) = H_0$. Это *ошибка второго рода* «*пропуск цели*».

Ошибку первого рода часто называют *ложной тревогой*, *ложным срабатыванием* или *ложноположительным срабатыванием* – например, анализ крови показал наличие заболевания, хотя на самом деле человек здоров, или металлодетектор выдал сигнал тревоги, сработав на металлическую пряжку ремня. Слово «положительный» в данном случае не имеет отношения к желательности или нежелательности самого события.

		Верная гипотеза	
		H_0	H_1
Результат применения критерия	H_0	H_0 верно принята	H_0 неверно принята (Ошибка *второго* рода)
	H_1	H_0 неверно отвергнута (Ошибка *первого* рода)	H_0 верно отвергнута

Рис. 4.6. Анализ гипотез

Ошибку второго рода называют *пропуском события* или *ложноотрицательным срабатыванием* – человек болен, но анализ крови этого не показал, или у пассажира имеется холодное оружие, но рамка металлодетектора его не обнаружила (например, из-за того, что чувствительность рамки отрегулирована на обнаружение только очень массивных металлических предметов). Слово «отрицательный» в данном случае не имеет отношения к желательности или нежелательности самого события.

Формулировка критерия Неймана–Пирсона

Пусть α – вероятность ошибки первого рода (ложной тревоги) называемая *уровнем значимости*;

β – вероятность ошибки второго рода (пропуска события).

Тогда критерием *Неймана–Пирсона* будет такое решающее правило f^*, которое обеспечивает минимум вероятности ошибки второго рода (пропуска события) при заданном ограничении на вероятность ошибки первого рода (ложной тревоги)

$$\begin{cases} \alpha(f^*) \le \alpha^0, \\ \beta(f^*) = \min_{f \in F} \beta(f). \end{cases} \qquad (4.13)$$

4.7. КРИТЕРИИ ПРИНЯТИЯ РЕШЕНИЙ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Постановка задачи

Задача выработки решений в условиях неопределенности (поведенческой неопределенности) характеризуется наличием неопределенных факторов *нестохастической природы*. Это означает, что:

1. Исход g зависит не только от выбранной стратегии u, но и от значения неопределенного фактора $\lambda \in \Lambda$.

2. Неопределенный фактор нестохастической природы означает, что либо неизвестен закон распределения его значений, либо принятие решения осуществляется для действия, которое будет выполняться однократно, или незначительное количество раз, когда законы статистики еще не могут работать.

Если исходы оцениваются скалярным показателем, то он является функцией двух аргументов $K(u, \lambda)$. В частности, если множества U и Λ конечны, то значения $K(u, \lambda)$ можно задать таблицей (платежная матрица):

	λ_1	λ_j	λ_m
u_1	$K(u_1, \lambda_1)$	$K(u_1, \lambda_j)$	$K(u_1, \lambda_m)$
u_i	$K(u_i, \lambda_1)$	$K(u_i, \lambda_j)$	$K(u_i, \lambda_m)$
u_n	$K(u_n, \lambda_1)$	$K(u_n, \lambda_j)$	$K(u_n, \lambda_m)$

Для решения ЗПР в данном случае предложен ряд критериев. Рассмотрим основные.

Критерий гарантированного результата (максимин, Вальда)

Каждая стратегия характеризуется *гарантированным уровнем* – значением показателя, соответствующим наименее благоприятному состоянию неизвестного фактора:

$$K_m(u) = \min_{\lambda \in \Lambda} K(u, \lambda).$$

Уровень является гарантированным в том смысле, что при любом значении неизвестного фактора при данной стратегии значение показателя $K(u, \lambda)$ не меньше этого уровня.

Критерий гарантированного уровня (*максиминный*, *Вальда*) предполагает оптимальной ту стратегию, при которой гарантированный уровень наибольший:

$$K_m(u^*) = \max_{u \in U} \min_{\lambda \in \Lambda} K(u, \lambda).$$

Недостаток: критерий учитывает только наихудшие значения неизвестного фактора и не учитывает благоприятные его значения.

Критерий минимаксного сожаления (Сэвиджа)

Предложен Сэвиджем как усовершенствование критерия Вальда. Его удобнее рассмотреть на примере.

u	λ_1	λ_2
u_1	0	1000
u_2	1	1

Пусть $\lambda = \lambda_1$. Но ЛПР об этом не знает. Если он выберет стратегию u_2, то он об этом не пожалеет, а если выберет u_1, то слегка пожалеет.

Пусть, теперь, $\lambda = \lambda_2$. Если ЛПР выберет стратегию u_1, то он об этом не пожалеет, а если выберет u_2, то его сожаление будет большим.

Критерий Сэвиджа предполагает считать оптимальной ту стратегию, максимальное значение сожаления при которой наименьшее

Составляем таблицу сожалений

	λ_1	λ_2
u_1	0	1000
u_2	1	1

	$\Delta\lambda_1$	$\Delta\lambda_2$
u_1	1	0
u_2	0	1000 – 1

Аналитические выражения для данного критерия такие

$$K^{MC}(u) = \max_{\lambda \in \Lambda}(\max_{v \in U} K(v, \lambda) - K(u, \lambda)),$$

где $K^{MC}(u)$ – максимальное сожаление при данной стратегии.

Тогда

$$u^* = \arg\min_{u \in U} K^{MC}(u) = \arg\min_{u \in U} (\max_{\lambda \in \Lambda} (\max_{v \in U} K(v, \lambda) - K(u, \lambda))).$$

Недостаток критерия.

При добавлении новой, даже неоптимальной стратегии, оптимальная может измениться.

Пример. В таблицу добавим еще одну стратегию, очевидно не оптимальную. Оптимальная изменится.

	λ_1	λ_2
u_1	0	1000
u_2	1	1
u_3	1000	–1

	$\Delta\lambda_1$	$\Delta\lambda_2$
u_1	1000	0
u_2	1000 – 1	1000 – 1
u_3	0	1000 + 1

Критерий пессимизма-оптимизма (Гурвица)

Рассмотренные выше критерии Вальда и Сэвиджа пессимистичны в том смысле, что рассматривают только наихудшие для каждой стратегии значения неопределенного фактора.

Использовать взвешенное среднее наилучшего и наихудшего значений предложил Гурвиц.

Определим

$$K_m(u) = \min_{\lambda \in \Lambda} K(u, \lambda) \quad \text{и} \quad K^M(u) = \max_{\lambda \in \Lambda} K(u, \lambda).$$

Возьмем произвольное число α: $0 < \alpha < 1$, называемое *показателем пессимизма-оптимизма*.

Тогда показатель

$$K^{\alpha}(u) = \alpha\, K_m(u) + (1 - \alpha)\, K^M(u),$$

называется «альфа-показатель стратегии».

Критерий Гурвица предполагает считать оптимальной стратегию, «альфа-показатель» которой наибольший:

$$u^* = \arg\max_{u \in U} K^{\alpha}(u).$$

Задание показателя α осуществляется произвольно, исходя из склонности ЛПР к оптимизму $\alpha \to 0$ или к пессимизму $\alpha \to 1$.

Принцип недостаточного основания (Бернулли).

Этот принцип (не критерий!!!) сформулирован Яковом Бернулли еще в 17 веке. Применяется, когда неизвестен закон распределения случайной величины (т.е. нет оснований считать какое-либо событие более вероятным, чем другие), а необходимо с ним работать. Бернулли предлагает считать распределение равномерным, а все события (исходы) равновероятными.

Критерий Байеса–Лапласа

Каждому значению влияющего фактора приписывается нормированный весовой коэффициент p_j, интерпретируемый, в частности, как вероятность этого значения.

Критерий предполагает максимум взвешенной суммы значений показателя при различных значениях влияющего фактора

$$K(u) = \sum_{j=1}^{J} p_j \cdot K(u, \lambda_j),\ \sum_{j=1}^{J} p_j = 1,$$

$$u^* = \arg\max_{u \in U} K(u).$$

Равновесный вариант критерия: если вероятности возникновения той или иной ситуации λ_j неизвестны, по принципу Бернулли считают их равновероятными. Тогда для каждой строки матрицы подсчитывается среднее арифметическое значение оценок. Оптимальному решению будет соответствовать такое решение, которому соответствует максимальное значение этого среднего арифметического, т.е.

$$\forall p_j = 1/J.$$

Критерий Гермейера

Также каждому значению влияющего фактора приписывается нормированный весовой коэффициент p_j, интерпретируемый, в частности, как вероятность этого значения.

Платежная матрица интерпретируется как расходы и заполняется отрицательными значеиями. Если это не так, то от каждого элемента матрицы отнимается одно и то же положительное число, выбираемое исходя из условия, чтобы в результирующей матрице все элементы стали отрицательными. Решение может зависеть этого выбранного числа!!!

Каждый элемент полученной матрицы умножается на соответствующую вероятность.

По полученной матрице решается задача по критерию Вальда.

Критерий Ходжа–Лемана

1. Для каждой стратегии определяются показатели по Вальду $K_{\text{В}}(u)$ и Байесу–Лапласу $K_{\text{БЛ}}(u)$.
2. Выбирается число v, $0 \le v \le 1$.
3. Для каждой стратегии вычисляется показатель $K_{\text{ХЛ}}(u)$

$$K_{\text{ХЛ}}(u) = v \cdot K_{\text{БЛ}}(u) + (1 - v) \cdot K_{\text{В}}(u).$$

4. Выбирается оптимальная стратегия по правилу

$$u^* = \arg\max_{u \in U} K_{\text{ХЛ}}(u).$$

4.8. КРИТЕРИИ ПРИНЯТИЯ РЕШЕНИЙ В УСЛОВИЯХ АКТИВНОГО ПРОТИВОБОРСТВА

Выше изучались критерии принятия решений в условиях, когда влияющие факторы принимали значения случайным образом, нецеленаправленно. Однако, влияющими на качество информационных систем факторами *могут быть целенаправленные деструктивные воздействия*.

Понятийный аппарат теории игр. Классификация игр

Модель проблемной ситуации характеризуется наличием нескольких участников, интересы которых не совпадают, а исход операции зависит от того, какие стратегии они выберут.

Проблемная ситуация называется конфликтной ситуацией, конфликтом.

Формальная модель конфликта называется игрой.

Математическая теория принятия решений в конфликтной ситуации называется *теорией игр* (*ТИ*).

Участники конфликта в модели называются *игроками*.

Показатели эффективности стратегий игроков – *функции выигрышей*.

Значения показателей, характеризующих исход конфликта, – *выигрыши*.

Конкретная реализация игры – *партия*.

Задачей ТИ является выработка рекомендаций для рационального поведения разумных игроков в конфликтной ситуации. ТИ должна ответить на следующие вопросы:

Какому игроку какие надо выбирать критерии? В чем заключается оптимальное решение?

Реализуем ли для данного класса игр выбранный критерий оптимальности?

Найти оптимальное решение.

Классификация игр

По количеству стратегий:

– *нестратегические* – в них выбирается одна стратегия, общая для всех игроков;

– *стратегические* – каждый игрок выбирает свою стратегию.

По числу игроков: *парные* – 2, *множественные* > 2.

По характеру взаимоотношений игроков:

– *кооперативные* – игроки могут вступать в переговоры;

– *некооперативные* – игроки не могут вступать в переговоры;

– *коалиционные* – игроки могут вступать в коалиции;

– *бескоалиционные* – игроки не могут вступать в коалиции.

По мощности множеств стратегий:

– *конечные* и *бесконечные*.

Будем рассматривать *стратегические бескоалиционные парные игры*.

Такая игра может быть представлена в форме:

$$< U, V, K_1, K_2 >,$$

где U, V – множества стратегий первого и второго игрока соответственно; K_1, K_2 – функции выигрышей первого и второго игроков, определенные на множестве $U \times V$.

Элемент (u, v) множества $U \times V$ называется ситуацией.

Предпочтения игроков: ситуация (u, v) с точки зрения первого игрока не менее предпочтительна чем (u', v'), если $K_1(u, v) \geq K_1(u', v')$ и

$K_2(u, v) \geq K_2(u', v')$ для второго игрока соответственно.

Важным частным случаем стратегической парной бескоалиционной игры является *антагонистическая игра* (*игра с нулевой суммой*).

Стратегическая парная бескоалиционная игра называется *антагонистической*, если $K_1 = -K_2$, т.е. интересы игроков противоположны. В этом случае можно рассматривать только

$$K_1 = K, \text{ т.е. } <U, V, K>.$$

Если мы не знаем намерений противника, то как наихудший случай должны предполагать антагонистическую игру. В этом заключается практическая значимость данного вида игр.

Решение игр в чистых стратегиях

Рассмотрим антагонистическую игру $\Gamma = <U, V, K>$.

Первый игрок стремится максимизировать свой выигрыш, т.е. максимизировать $K(u, v)$ а второй – максимизировать свой выигрыш, т.е. минимизировать ту же функцию $K(u, v)$.

Каким критерием следует руководствоваться первому игроку?
Каким критерием следует руководствоваться второму игроку?

u	v_1	v_j	v_m
u_1	$K(u_1, v_1)$	$K(u_1, v_j)$	$K(u_1, v_m)$
u_i	$K(u_i, v_1)$	$K(u_i, v_j)$	$K(u_i, v_m)$
u_n	$K(u_n, v_1)$	$K(u_n, v_j)$	$K(u_n, v_m)$

Первому игроку следует выбрать *максиминную стратегию* u^*

$$K_m (u^*) = \max_{u \in U} K_m (u) = \max_{u \in U} \min_{v \in V} K(u, v),$$

где $K_m(u) = \min_{v \in V} K(u, v)$.

Эта стратегия обеспечит ему максимальный гарантированный выигрыш.

Тогда для второго игрока оптимальной будет *минимаксная стратегия* v^*

$$K^M (v^*) = \min_{v \in V} K^M (v) = \min_{v \in V} \max_{u \in U} K(u, v),$$

где $K^M (v) = \max_{u \in U} K(u, v)$.

В теории игр доказывается, что

$$K_m (u^*) \leq K^M (v^*),$$

которые называются *нижним* и *верхним* значениями игры.

Важным случаем является случай, когда

$$K_m (u^*) = K^M (v^*) = K = K(u^*, v^*) \text{ – \textit{значение игры}}.$$

Ситуация (u^*, v^*) в этом случае называется ситуацией равновесия и имеет место двойное неравенство

$$K(u \neq u^*, v^*) \leq K(u^*, v^*) \leq K(u^*, v \neq v^*).$$

Это неравенство показывает, что игрокам нет необходимости искать какие-либо другие стратегии, отличные от u^* и v^*. Эти стратегии будут *оптимальными* для обоих игроков.

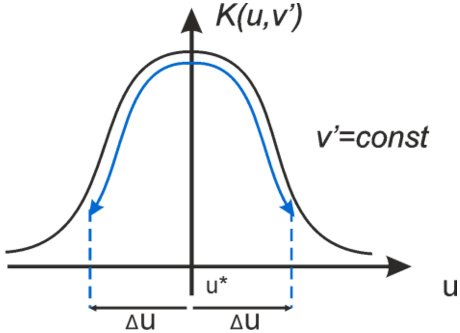


Рис. 4.7. Зависимость значения игры *K*(*u, v*’) от стратегии *u* первого игрока при фиксированной стратегии *v*’ второго игрока

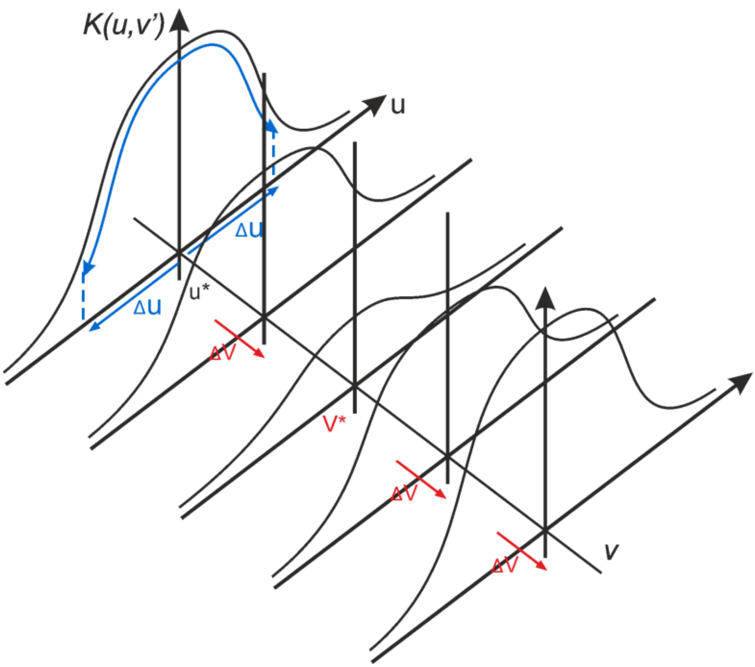


Рис. 4.8. Зависимость значения игры *K*(*u, v*) от стратегии *v* второго игрока

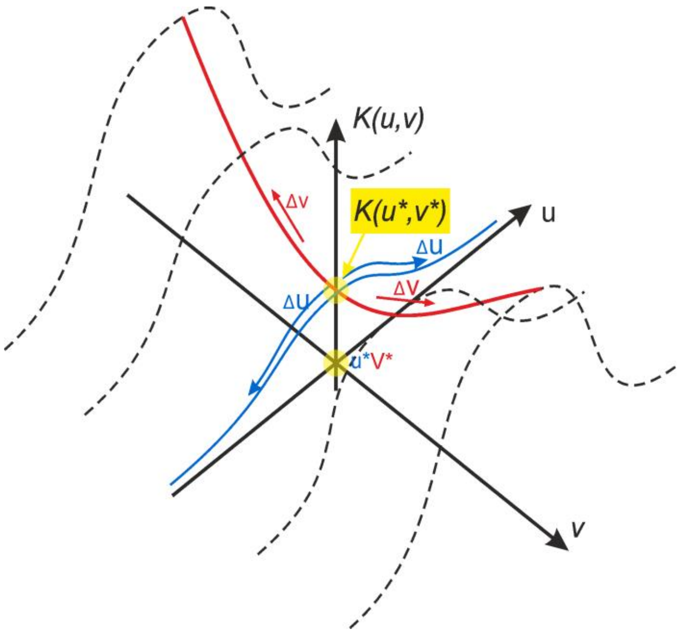


Рис. 4.9. Решение игры – седловая точка *K*(*u, *v**)**

Матричные игры

Def. *Матричной* называется конечная антагонистическая игра, т.е. в ней множества U и V конечны.

Такую игру можно полностью задать функциями выигрыша в форме матрицы вида

$$A = \begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Vmatrix}.$$

Первый игрок выбирает номер строки i, второй номер столбца j. Тогда a_{ij} – есть выигрыш первого игрока и, соответственно, проигрыш второго.

Если существует значение игры K, т.е. существует ситуация равновесия, то игроки имеют все основания применять образующие ее стратеги.

Ситуация равновесия, если она существует, характеризуется тем, что элемент матрицы $a_{i^*j^*}$ является наибольшим в j^*-*м* столбце и наименьшим в i^*-*й* строке.

Стратегии i^* и j^* являются оптимальными и должны быть выполнены игроками.

Пример: $\begin{Vmatrix} 2 & 3 & 1 \\ 3 & 4 & 3 \\ 1 & 5 & 3 \end{Vmatrix}$.

Здесь, очевидно, что для 1-го игрока оптимальна 2-я стратегия, для второго 1-я и 3-я и нет оснований их не придерживаться.

То есть для первого игрока, независимо от того, какую стратегию выберет второй игрок, оптимальной будет одна и та же стратегия.

Если ситуации равновесия не существует, то поведение игроков должно быть совсем иным.

Понятие смешанной стратегии

Пусть дана матрица $\begin{Vmatrix} 3 & 1 \\ 2 & 4 \end{Vmatrix}$.

Здесь от знания первым игроком стратегии второго игрока зависит выбор его стратегии. И нет однозначно предпочтительной стратегии.

В этом случае можно применить *смешанные стратегии*.

Определение. *Смешанной стратегией первого игрока* называется вероятностное распределение на множестве его чистых стратегий, т.е. вектор p:

$$p = (p_1, p_2, ..., p_n) \mid p_i = P(u_i), \quad \sum_i p_i = 1.$$

Аналогично смешанная стратегия второго игрока:

$$q = (q_1, q_2, ..., q_m) \mid q_j = Q(v_j), \quad \sum_j q_j = 1.$$

Очевидно, что «чистые стратегии» есть частный случай смешанных при

$$\exists\, p_k = 1, \ \forall\ i \neq k,\ \ p_i = 0.$$

Применение смешанной стратегии заключается в случайном выборе одной чистой стратегии из множества стратегий в соответствии с вероятностным распределением.

Введем МО выигрыша для ситуации (p, q):

$$M(p,\, q) = \sum_i \sum_j a_{ij}\, p_i q_j. \qquad (4.14)$$

Здесь *p* и *q* – распределения!!!

Максиминная смешанная стратегия для первого и второго игроков:

$$M_m(p^*) = \max_{p \in P} M_m(p), \text{ где } M_m(p) = \min_{q \in Q} M(p, q),$$

$$M^M(q^*) = \min_{q \in Q} M^M(q), \text{ где } M^M(q) = \max_{p \in P} M(p, q).$$

И обязательно $M_m(p^*) = M^M(q^*)$ и существует хотя бы одна пара стратегий (ситуация), это обеспечивающая.

То есть смешанная игра обязательно имеет значение.

Основная теорема матричных игр – каждая матричная игра имеет хотя бы одну ситуацию равновесия.

И имеет место система неравенств:

$$M(p, q^*) \le M(p^*, q^*) \le M(p^*, q).$$

$$K_m \le M^* \le K^M.$$

Очевидно, что в играх со смешанными стратегиями игрокам нет никакого резона извещать противника о своих намерениях.

Как на практике реализовать смешанную стратегию?

Иногда возможно физическое смешивание стратегий.

Пример. При стрельбе по самолету можно использовать зажигательные и бронебойные пули. Пусть теория рекомендует эти две стратегии с вероятностью 1/3 и 2/3. Тогда реально можно снарядить ленту так, что на одну зажигательную пулю в среднем будет 2 бронебойных.

Если нет возможности смешать стратегии, то выводы математики применимы при многократных играх случайным, в соответствии с полученным распределением, выбором стратегии.

Если же игра будет реализована однократно, то результат может быть любым, в том числе хуже (для первого игрока) чем K_m.

Биматричные игры

Конечная бескоалиционная игра двух игроков называется биматричной игрой.

Пусть первый игрок имеет m стратегий, а второй – n стратегий.

Выигрыши первого и второго игроков задаются матрицами

$$A = [a_{ij}]_{m\times n} \quad \text{и} \quad B = [b_{ij}]_{m\times n}\ .$$

Если первый игрок применяет стратегию i, а второй стратегию j, то первый игрок выигрывает a_{ij}, а второй – b_{ij}.

Пара стратегий (i_0, j_0) является ситуацией равновесия (в чистых стратегиях) в биматричной игре, если выполняются следующие неравенства

$$a_{i_0,\, j_0} \ge a_{i,\, j_0},\ i = 1, \ldots, m,$$

$$b_{i_0,\, j_0} \ge b_{i_0,\, j},\ j = 1, \ldots, n.$$

Все ситуации равновесия в чистых стратегиях можно найти следующим образом:

1) в каждом столбце матрицы *A* отмечаем звездочкой максимальные элементы:

2) в каждой строке матрицы *B* отмечаем звездочкой максимальные элементы;

3) все пары стратегий (i, j), такие, что оба элемента a_{ij} и b_{ij} отмечены звездочкой являются ситуациями равновесия.

Смысл ситуации равновесия состоит в том, что каждому игроку невыгодно односторонне отступать от входящей в ситуацию стратегии.

Понятие ситуации равновесия обобщает понятие седловой точки матричной игры. Это – равновесие по Нэшу.

Пример.

$$A = \begin{pmatrix} 8 & 4 & 6 \\ 7 & 8 & 9 \\ 2 & 1 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 4 & 5 \\ 3 & 7 & 6 \\ 5 & 6 & 9 \end{pmatrix}.$$

Позиции максимумов в столбцах матрицы *A* – (1, 1), (2, 2), (2, 3). Позиции максимумов в строках матрицы *B* – (1, 3), (2, 2), (3, 3). Пересечение этих двух множеств – (2, 2). Таким образом, есть ситуация равновесия в чистых стратегиях

$$A = \begin{pmatrix} 7 & 5 & 1 \\ 6 & 3 & 2 \\ 4 & 0 & 9 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 2 & 5 \\ 0 & 6 & 7 \\ 1 & 6 & 4 \end{pmatrix}.$$

Позиции максимумов в столбцах матрицы *A* – (1, 1), (1, 2), (3, 3). Позиции максимумов в строках матрицы *B* – (1, 3), (2, 3), (3, 2). Соответствующие значения обведены кружками. Эти два множества позиций не пересекаются, следовательно, ситуации равновесия в чистых стратегиях нет.

Выводы:

1. Конфликтные ситуации являются распространенными ситуациями в ЗПР.

2. Оптимальным при решении ЗПР в этих условиях является максиминный (минимаксный) критерий.

3. Если в задаче существует ситуация равновесия, то стратегии, входящие в нее, являются оптимальными и их следует придерживаться.

4. Не каждая задача ТИ имеет значение в чистых стратегиях, но каждая задача имеет значение в смешанных стратегиях.

5. Реализация смешанной стратегии осуществляется либо физическим смешением чистых стратегий, либо при многократной реализации.

4.9. ОСНОВАНИЯ МЕТОДА АНАЛИЗА ИЕРАРХИЙ

Метод анализа иерархий (МАИ), предложенный американским ученым Т. Саати, является одним из самых известных методов решения практических многокритериальных векторных задач принятия решений самого различного характера и сложности. Он позволяет решать задачу выбора варианта из конечного множества возможных вариантов по множеству показателей (как качественных, так и количественных), имеющих иерархическую структуру, с учетом их приоритетов (значимости).

Метод состоит в декомпозиции проблемы на все более простые составляющие части и сведении экспертной оценки к последовательности суждений в форме попарных сравнений по специальной шкале (шкале отношений).

МАИ основывается на следующих принципах.

1. Принцип декомпозиции. Данный принцип предусматривает структурирование проблемы в виде иерархии, что является первым этапом применения МАИ.

Используется *доминантная*[20] иерархия («перевернутое дерево с основой в вершине»).

[20] Кроме них существуют *холлархии* – доминантные иерархии с обратной связью, *китайский ящик* (или модулярные иерархии), который растет в размерах от простейших элементов или компонент (внутренние ящики) ко все более крупным совокупностям (внешние ящики). В биологии интерес представляют *неогенетические иерархии*, в которых новые верхние уровни возникают последовательно в течение эволюции.

Предполагается полная иерархия, т.е. считается, что каждый элемент заданного уровня связан со всеми элементами последующего уровня.

Простейшая полная иерархия проблемы выбора включает в себя три уровня (рис. 4.10): цель, показатели, альтернативные варианты.

2. Принцип сравнительных суждений. Чтобы установить приоритеты показателей и получить оценки для альтернативных решений, в МАИ используется метод парных сравнений – строятся матрицы парных сравнений $A = ||a_{i,j}||$, где $a_{i,j} = w_i / w_j$, w_i – «вес» i-го элемента иерархии. Очевидно, что $a_{ii} = 1$, $a_{ij} = 1/a_{ji}$, (т.е. диагональные элементы матрицы равны 1, матрица является обратносимметричной). По каждой матрице определяется вектор локальных приоритетов и вычисляется индекс согласованности мнений эксперта.

3. Принцип синтеза приоритетов.

Итак, будем считать, что:

1) построены матрицы парных сравнений: одна для второго уровня иерархии, а на каждом последующем уровне – столько матриц парных сравнений, сколько элементов содержит предшествующий уровень иерархии (в каждой матрице – результаты сравнения по одному из показателей);

2) вычислены векторы локальных приоритетов по каждой матрице.

Приоритеты синтезируются снизу–вверх. Локальные приоритеты альтернатив умножаются на приоритеты соответствующих показателей следующего уровня и суммируются по каждому элементу в соответствии с показателями. Таким образом, итоговой оценкой альтернативы в методе парных сравнений является вес альтернативы, вычисляемый как свертка весовых коэффициентов локальных показателей всех уровней иерархии.

4.10. АЛГОРИТМ МЕТОДА АНАЛИЗА ИЕРАРХИЙ

Алгоритм МАИ включает в себя следующие шаги:

1) формирование иерархии целей;
2) определение приоритетов;

3) расчет локальных векторов приоритетов;

4) проверка экспертных оценок на непротиворечивость (вычисление индекса согласованности);

5) расчет приоритетов целей и мероприятий для иерархии в целом на основе синтеза локальных приоритетов.

Рассмотрим его подробно.

Первым этапом применения МАИ является структурирование проблемы выбора в виде иерархии (*Принцип декомпозиции*).

В наиболее элементарном виде иерархия строится с вершины (для реализации целей управления), через промежуточные уровни (показатели, от которых зависят последующие уровни) к самому низкому уровню (который обычно является перечнем альтернатив).

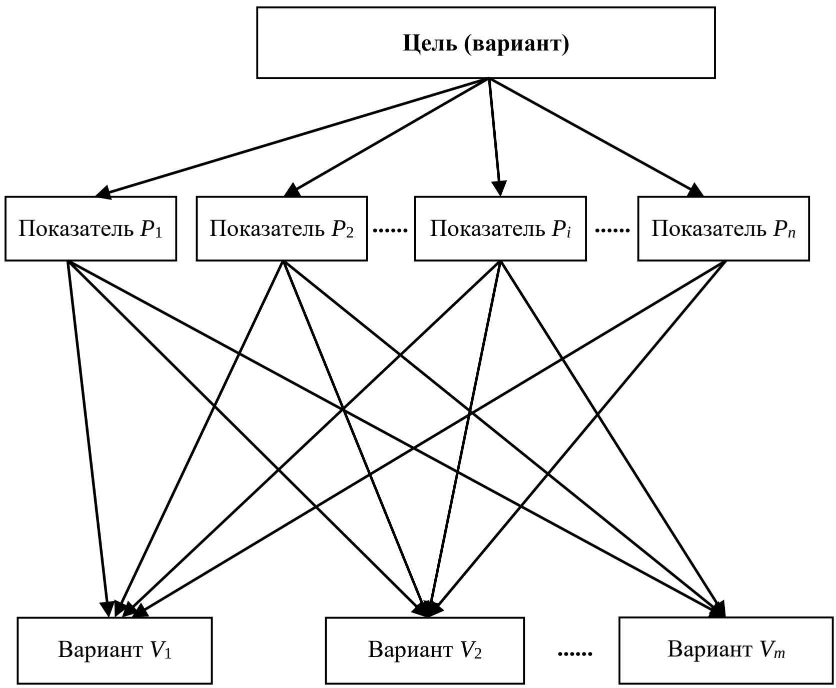


Рис. 4.10. Принцип декомпозиции

Выбор факторов (показателей) на втором и ниже уровнях является весьма ответственной процедурой и должен быть проверен и подтвержден.

Итак, пусть декомпозиция выполнена:

Второй шаг заключается в установлении приоритетов показателей.

Как уже говорилось, МАИ *состоит в … сведении экспертной оценки к последовательности суждений в форме попарных сравнений по специальной шкале* (*шкале отношений*).

Для этого строится квадратная матрица (табл. 4.1) размера $n \times n$ следующего вида:

Таблица 4.1

	P_1	P_2	…	P_i	…	P_n
P_1						
P_2						
…						
P_i						
…						
P_n						

Заполнение матрицы осуществляется путем попарных сравнений важности вариантов с использованием специальной шкалы – *рекомендуемой шкалы относительной важности*.

Рекомендуемая шкала относительной важности

Для проведения субъективных парных сравнений разработана шкала, описанная в табл. 4.2. Эта шкала оказалась эффективной не только во многих приложениях, ее правомочность доказана теоретически при сравнении со многими другими шкалами.

4.2. Шкала относительной важности

Интенсивность относительной важности	Определение	Объяснения
1	Равная важность	Равная важность двух вариантов
3	Умеренное превосходство одного над другим	Опыт и суждения дают легкое превосходство одного варианта над другим
5	Существенное или сильное превосходство	Опыт и суждения дают сильное превосходство одного варианта над другим
7	Значительное превосходство	Одному варианту дается настолько сильное превосходство, что оно становится практически доминирующим
9	Очень сильное превосходство	Очевидность полного превосходства одного варианта над другим не подвергается никакому сомнению
2, 4, 6, 8	Промежуточные решения между двумя соседними суждениями	Применяются в компромиссном случае
Обратные величины приведенных выше чисел	Если при сравнении одного варианта с другим получено одно из вышеуказанных чисел (например 3), то при сравнении второго варианта с первым получим обратную величину (т.е. 1/3)	

Матрица (табл. 4.3) заполняется следующим образом:

Начинать следует с левого варианта и задать вопрос; насколько он важнее, чем вариант вверху? При сравнении варианта с самим собой отношение равно единице. Если первый вариант важнее, чем второй, то используется целое число из рекомендуемой шкалы относительной важности, в противном случае используется обратная величина. В любом случае обратные друг к другу отношения заносятся в симметричные позиции матрицы. Поэтому матрицы получаются положительными обратносимметричными, и необходимо произвести только $n(n - 1)/2$ суждений, где n – общее число сравниваемых вариантов.

При этом не предполагается, что суждения людей полностью согласованы и согласовывать суждения требуется, только если они очень сильно рассогласованы, что проверяется далее специальными методами.

Однако, требование обратной симметричности должно выполняться неукоснительно!

Заполненная матрица имеет вид (табл. 4.3).

Таблица 4.3

	P_1	P_2	…	P_i	…	P_n
P_1	a_{11}	a_{12}	…	a_{1i}	…	a_{1n}
P_2	a_{21}	a_{22}		a_{2i}		a_{2n}
…	…	…	…	…	…	…
P_i	a_{i1}	a_{i2}	…	a_{ii}	…	a_{in}
…	…	…	…	…	…	…
P_n	a_{n1}	a_{n2}		a_{ni}		a_{nn}

Примечание: $\forall i\ a_{ii} = 1, \quad \forall i, j\ a_{ij} = 1/a_{ji}$.

Для следующего уровня иерархии строятся дополнительно n матриц (по количеству показателей) размером $m \times m$ и заполняются аналогично табл. 4.3, с помощью табл. 4.2, при этом в ячейках попарно сравниваются и оцениваются варианты (третий уровень иерархии) по соответствующему показателю.

Если какой-либо P_i показатель измеряется в количественной шкале (масса, объем, мощность, удельный расход топлива и т.п.), то нет необходимости прибегать к экспертным оценкам.

Если показатель характеризует какой-либо полезный эффект (т.е. чем больше его значение, тем «важнее» вариант), то в ячейку таблицы записывают дробь, в числителе которой стоит значение показателя варианта-строки, в знаменателе значение показателя варианта-столбца.

Таблица 4.4

Матрица № i: Приоритеты вариантов по показателю P_i					
	V_1	…	V_j	…	V_m
V_1	b_{11}	...	b_{1j}	...	b_{1m}
...	...	...	...	...	...
V_j	b_{j1}	...	b_{jj}	...	b_{jm}
…	...	...	...	...	...
V_m	b_{m1}	...	b_{mj}	...	b_{mm}

Примечание: $\forall j\ b_{jj} = 1, \quad \forall i, j\ b_{ij} = 1/b_{ji}$.

Примечание: то же значение вектора приоритетов вариантов можно получить проще – нормированием их.

Если показатель характеризует какие-либо затраты (т.е. чем меньше его значение, тем «важнее» вариант), то в ячейку таблицы записывают дробь, в числителе которой стоит значение показателя варианта-столбца, в знаменателе значение показателя варианта-строки.

На этом этап сбора информации заканчивается.

Этап второй – обработка результатов экспертного опроса.

Относительная сила, величина, ценность, желательность каждого отдельного объекта находится через «решение» матриц.

Для этого для каждой матрицы необходимо:

1. Вычислить множество собственных чисел, из них выбрать наибольшее λ_{max};

2. Найти собственный вектор, соответствующий наибольшему собственному числу λ_{max};

3. Нормировать найденный собственный вектор, получив тем самым вектор приоритетов.

Множество собственных значений (чисел) матрицы A есть множество корней ее характеристического полинома $f(x)$, который определяется как определитель матрицы $A - xE$, где E – единичная матрица

$$f(x) = \det(A - xE) \text{ или}$$

$$\det(A - xE) = \begin{pmatrix} a_{11} - x & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - x & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} - x & a_{3n} \\ a_{n1} & a_{n2} & \cdots & a_{nn} - x \end{pmatrix}.$$

Прежде, чем использовать найденный вектор приоритетов, *следует выяснить степень согласованности оценок* в табл. 4.3 и в тех (табл. 4.4), которые получены экспертным путем.

Для этого вычисляется индекс согласованности (ИС), который дает информацию о степени нарушения численной (кардинальной, $a_{ij}\,a_{jk} = a_{ik}$) и транзитивной (порядковой) согласованности

$$\text{ИС} = \frac{\lambda_{\max} - n}{n - 1}, \quad (4.15)$$

где n – число сравниваемых элементов.

Для обратносимметричной матрицы всегда $\lambda_{max} \geq n$. Чем меньше λ_{max}, тем лучше согласована матрица. Для матриц, содержащих отношения численных значений показателей, $\lambda_{max} = n$ и ИС = 0.

Однако, ИС зависит от размерности матрицы n.

Показателем согласованности оценок, не зависящим от размерности матрицы, является *отношение согласованности* (ОС), рассчитываемое как

$$\text{ОС} = \frac{\text{ИС}}{\text{СС}(n)}, \tag{4.16}$$

где СС(n) – случайная согласованность (СС), полученная усреднением значений ИС для матриц размера n, заполненных случайным образом по шкале 1/9, 1/8, 1/7, ..., 1, 2, ..., 9, но образовании обратносимметричной матрицы.

Ниже даны средние согласованности для случайных матриц разного порядка.

Таблица 4.5

Размер матрицы	1	2	3	4	5	6	7	8	9	10
Случайная согласованность (СС)	0	0	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49

Величина ОС должна быть порядка 10% или менее, чтобы быть приемлемой. В некоторых случаях можно допустить 20%, но не более. Если ОС выходит из этих пределов, то участникам нужно исследовать задачу и проверить свои суждения.

Нахождение множества собственных чисел матрицы есть решение алгебраического уравнения n-й степени.

Вычисление множества собственных значений – не является сложной задачей для современного специализированного программного обеспечения, однако ее решение может быть осуществлено с помощью табличного процессора Excel. Разработаны несложные и эффективные пути получения хорошего приближения к точным оценкам, одним из которых является *геометрическое среднее*.

Рассмотрим «решение» на примере первой матрицы (табл. 4.3).

Для каждой *i-й* строки матрицы находим среднегеометрическое a_i по строке

$$a_i = \sqrt[n]{\prod_{j=1}^{n} a_{ij}} \qquad (4.17)$$

и записываем в дополнительном столбце справа матрицы

Таблица 46

	P_1	P_2	…	P_j	…	P_n		
P_1	a_{11}	a_{12}	…	a_{1j}	…	a_{1n}	a_1	$\bar{a}_1$
P_2	a_{21}	a_{22}	…	a_{2j}	…	a_{2n}	a_2	$\bar{a}_2$
…	…	…	…	…	…	…	…	…
P_i	a_{i1}	a_{i2}	…	a_{ij}	…	a_{in}	a_i	$\bar{a}_i$
…	…	…	…	…	…	…	…	…
P_n	a_{n1}	a_{n2}	…	a_{nj}	…	a_{nn}	a_n	$\bar{a}_n$

Затем средние геометрические строк нормируются

$$\bar{a}_i = \frac{a_i}{a_1 + \ldots + a_i + \ldots + a_n} = \frac{a_i}{\sum_{i=1}^{n} a_i}. \qquad (4.18)$$

Тогда

$$\sum_{i=1}^{n} \bar{a}_i = 1. \qquad (4.19)$$

Согласованность оценок в матрице 2.5 также может быть рассчитана приближенно.

	P_1	P_2	…	P_i	…	P_n		
P_1	a_{11}	a_{12}	…	a_{1i}	…	a_{1n}	a_1	$\bar{a}_1$
P_2	a_{21}	a_{22}	…	a_{2i}	…	a_{2n}	a_2	$\bar{a}_2$
…	…	…	…	…	…	…		…
P_i	a_{i1}	a_{i2}	…	a_{ii}	…	a_{in}	a_i	$\bar{a}_i$
…	…	…	…	…	…	…		…
P_n	a_{n1}	a_{n2}	…	a_{ni}	…	a_{nn}	a_n	$\bar{a}_n$
	A_1	A_2	…	A_i	…	A_n		

$$A_i = \sum_j a_{ji}. \tag{4.20}$$

	P_1	P_2	…	P_i	…	P_n			
P_1	a_{11}	a_{12}	…	a_{1i}	…	a_{1n}	a_1	$\bar{a}_1$	$\bar{a}_1 \cdot A_1$
P_2	a_{21}	a_{22}	…	a_{2i}	…	a_{2n}	a_2	$\bar{a}_2$	$\bar{a}_2 \cdot A_2$
…	…	…	…	…	…	…		…	…
P_i	a_{i1}	a_{i2}	…	a_{ii}	…	a_{in}	a_i	$\bar{a}_i$	$\bar{a}_i \cdot A_i$
…	…	…	…	…	…	…		…	…
P_n	a_{n1}	a_{n2}	…	a_{ni}	…	a_{nn}	a_n	$\bar{a}_n$	$\bar{a}_n \cdot A_n$
	A_1	A_2	…	A_i	…	A_n			$\lambda_{\max}$

$$\lambda_{\max} = \sum_{i=1}^{n} \bar{a}_i A_i. \tag{4.21}$$

То есть сумма первого столбца умножается на величину первой компоненты нормированного вектора приоритетов, сумма второго столбца – на вторую компоненту и т.д. Затем полученные числа суммируются. Таким образом, можно получить величину $\lambda_{\max}$.

Последний этап – синтез глобальных приоритетов.

Теперь обратимся к синтезу глобальных приоритетов вариантов W_i. Приоритеты синтезируются, начиная со второго уровня вниз. Локальные нормированные приоритеты $\bar{b}_{ij}$ i-х вариантов по j-м показателям умножаются на приоритеты j-х показателей $\bar{a}_j$ и складываются.

	P_1	…	P_j	…	P_n	
	$\bar{a}_1$		$\bar{a}_j$		$\bar{a}_n$	
V_1	$\bar{b}_{11}$		$\bar{b}_{1j}$		$\bar{b}_{1n}$	W_1
…						
V_i	$\bar{b}_{i1}$		$\bar{b}_{ij}$		$\bar{b}_{in}$	W_i
…						
V_m	$\bar{b}_{m1}$		$\bar{b}_{mj}$		$\bar{b}_{mn}$	W_m

$$W_i = \sum_{j=1}^{n} \bar{b}_{ij} \cdot \bar{a}_j = \bar{b}_{i1} \cdot \bar{a}_1 + \bar{b}_{i2} \cdot \bar{a}_2 + \ldots + \bar{b}_{in} \cdot \bar{a}_n. \qquad (4.22)$$

$$V^* = \arg\max_{\forall V} W(V). \qquad (4.23)$$

Оптимальным будет считаться вариант V^*, имеющий наибольшее значение глобального приоритета (4.23).

4.11. КРИТИКА МЕТОДА АНАЛИЗА ИЕРАРХИЙ В. В. ПОДИНОВСКИМ

В ряде научных публикаций звучит критика МАИ, его «некорректность».

Одной из таких публикаций является статья – *В. В. Подиновский, О. В. Подиновская*. О некорректности метода анализа иерархий // Проблемы управления. № 1. 2011. С. 8 – 13, в которой приводится пример, в котором решение, предложенное МАИ, не согласуется с очевидным интуитивным решением.

Приведем этот пример.

Дано: трехуровневая иерархия (цель – показатели – альтернативы), два равноважных показателя P_1 и P_2, имеющих одинаковые численные шкалы, и четыре альтернативы, имеющие численные значения данных показателей.

	V_1	V_2	V_3	V_4
P_1	9	1	3	9
P_2	3	9	3	1

Интуитивным будет решение: $V^* = V_1$, далее следуют равные V_2 и V_4, и последний – V_3.

Нормируем значения показателей

	V_1	V_2	V_3	V_4
P_1	0,409	0,045	0,136	0,409
P_2	0,188	0,563	0,188	0,063

Поскольку показатели «равноважные», их весовые коэффициенты равны 0,5.

По формуле (4.22)

$$W_i = \sum_{j=1}^{n} \bar{b}_{ij} \cdot \bar{a}_j = \bar{b}_{i1} \cdot \bar{a}_1 + \bar{b}_{i2} \cdot \bar{a}_2 + \ldots + \bar{b}_{in} \cdot \bar{a}_n.$$

Получим

V_1	V_2	V_3	V_4
(0,409 + 0,188)/2 = = 0,298	(0,045 + 0,563)/2 = = 0,304	(0,136 + 0,188)/2 = = 0,162	(0,409 + 0,063)/2 = = 0,236

Согласно расчету, решением будет V_2 со значением глобального приоритета, равным 0,304, за ним следует V_1 с 0,298, затем V_4 с 0,236 и замыкает V_3 с 0,162, что очевидным образом не совпадает с интуитивным решением.

Причиной расхождения решений заключается в аддитивности свертки частных показателей при использовании шкалы отношений в рамках каждого показателя.

Исключить некорректность можно итерационным путем, когда на каждой итерации исключается наихудший вариант и пересчитывается вектор глобальных приоритетов до тех пор, когда останутся всего два варианта, из которых вычисляется лучший.

5. ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

В научных исследованиях важнейшее место занимает информация. Собственно, научные исследования и есть процесс получения и осмысления информации. Тема главы – информация.

5.1. ОПРЕДЕЛЕНИЕ ТЕРМИНА «ИНФОРМАЦИЯ»

Как и каждая отрасль науки, наука об информации должна оперировать своим понятийным аппаратом, лежащем в ее основе.

Однако, информация – категория особенная. Термин «информация» используется людьми давно и им не нужно было размышлять, что это такое, чтобы знать, когда это слова можно употреблять, а когда нет.

Дайте определение термину – *информация*.

С учетом ответов, в частности, на следующие вопросы:

– Есть ли информация во вчерашней газете?

– Есть ли информация в камне с отпечатком следа динозавра, лежащем под слоем песка, и который никто никогда не увидит?

– Циркулирует ли информация в САУ, в частности, при управлении уровнем воды в резервуаре, температурой или давлением в котле, количеством оборотов вала?

С развитием кибернетики, различных управляющих систем, потребовалось точно сформулировать понятие – информация и производные от него, в частности – «сообщение», «данные», в чем измерять «количество информации».

Известный польский ученый Мариан Мазур в 1972 году писал[21]:

Известны три группы публикаций, касающихся термина «информация»:

[21] Мазур М. Качественная теория информации. Перевод с польского О. И. Лочмеля. М.: Мир, 1974. 238 с.

1. Публикации, в которых информация отождествляется с количеством информации.

2. Авторы употребляют выражение «информация» без всяких объяснений, в таких, например, выражениях, как «перенос информации», «передача информации», «информация, заключенная в символах» и т.д., как бы подразумевая, что дело идет о понятии, не вызывающем сомнений.

3. Авторы пытаются объяснить термин, но с помощью таких же неопределенных терминов, как «сведения», «содержание», либо анализируют трудности в формулировании точного определения, сопоставляют взгляды авторов, чтобы в конце концов оставить вопрос открытым и предоставить самим читателям в хаосе неясностей и сомнений выработать собственные взгляды на этот предмет.

Примеры:

1. Винер Н.: «информация – обозначение содержания, черпаемого нами из внешнего мира в процессе нашего приспособления к нему и приведения в соответствие с ним нашего сознания» (здесь информация определяется через «мы», через «содержание», не как самостоятельный объект, а как атрибут «нашего приспособления»; и не определяется, есть ли информация в кибернетических системах?).

2. Куфиньяль:

– информация – сочетание носителя с семантикой;

– семантика – психический эффект информации;

– носитель – физическое явление, связанное с семантикой для образования информации.

С тех пор минуло много лет, но ясности в терминологии по-прежнему нет.

Юзвишин[22] И. И., 1996 г.: «информация это – генерализационно-фундаментальная субстанция единого кодово-сотового пространства Вселенной, включающего воздух, воду, землю, солнечные и другие светоносные лучи, поля, их следы и весь спектр космических излучений,

[22] Юзвишин И. И. Информациология: монография. 3-е изд., испр. и доп. М.: Радио и связь, 1996,

материализованных и дематериализованных сред, и выражающаяся через массу, скорость, энергию и другие формы, проявляющиеся в процессе материализации и дематериализации».

Из интернета:

ИНФОРМАЦИЯ – это «сумма всех пространственно-временных характеристик материи». То есть это все, чем можно охарактеризовать конкретный кусок материи: размеры, цвет и т.д., т.е. механические, оптические, электрические, химические и прочие параметры.

Информация (information) – это данные, сопровождающиеся смысловой нагрузкой. Тут еще есть про *знание*: «Знание – есть переживание, сравненное с другими переживаниями – это определение, которое дал Лосский Н. О. в своей работе «Обоснование интуитивизма», является наиболее точным» – это цитата. И даже формулы, правда не ясны единицы измерения:

информация = данные + смысл
знание = информация + сравнение.

А что такое данные?

Пример неудачного, на мой взгляд, определения этого термина: данные (data) – это такое же первоначальное понятие, как, скажем, в математике «точка»: попытка дать определение начальным понятиям приводит к необходимости дополнительно определять использованные термины. Итак, будем считать, что данные – это те или иные сведения (необязательно несущие смысловую нагрузку).

Информация (от лат. informatio – осведомление, разъяснение, изложение) – в широком смысле абстрактное понятие, имеющее множество значений, в зависимости от контекста (так что же такое информация в широком смысле?). В узком смысле этого слова – сведения (сообщения, данные) независимо от формы их представления (это в чистом виде из ФЗ «Об информации, информационных технологиях и о защите информации»).

Федеральный закон Российской Федерации от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации» (http://www.rg.ru/2006/07/29/informacia-dok.html): «информация – сведения (сообщения, данные) независимо от формы их представления».

Толковый словарь русского языка Ожегова приводит два определения слова «информация»:

- Сведения об окружающем мире и протекающих в нем процессах, воспринимаемые человеком или специальным устройством.
- Сообщения, осведомляющие о положении дел, *о состоянии* чего-нибудь.
- Информация – кодированный смысл.

Здесь введение в определение термина «смысл» порождает проблемы, связанные с личностью потребителя информации, его априорных знаний, целей, с которыми он воспринимает информацию. Получается уже не определение непосредственно информации, а определение процесса взаимодействия человека с информацией.

Бриллюэн Л. Наука и теория информации. М.: «Государственное издательство физико-математической литературы», 1960. – 391 с. http://edu.sernam.ru/book_scin.php?id=136

Стр. 15: «Методы этой теории (теории информации) могут с успехом применяться ко всем техническим проблемам, касающимся информации, как-то: кодирование, связь, вычислительные устройства и т.д. Во всех этих проблемах мы фактически перерабатываем информацию, или передаем ее из одного места в другое, и данная теория очень полезна для формулировки правил и установления точных пределов того, что может быть, а что не может быть сделано.

Но мы не в состоянии исследовать процесс мышления, и мы не можем в настоящее время ввести в нашу теорию какой-либо элемент, включающий человеческую оценку информации. Это исключение человеческого элемента является очень серьезным ограничением, но это есть та цена, которую мы должны были уплатить за возможность построения этой области научного знания. Введенные ограничения позволяют нам дать количественное определение информации и трактовать информацию как физически измеримую величину. Определение не может делать различия между очень важной информацией и новостью, не имеющей большой ценности для того, кто ее узнает.

Исключение человеческого элемента как раз дает возможность ответить на целый ряд вопросов. Инженера, который конструирует телефонную систему, не интересует, будет ли она использована для передачи сплетен, биржевых цен или дипломатических сообщений. Техническая задача всегда одна и та же: передать информацию, какова бы она ни была, правильно и точно. Конструктор вычислительной машины не знает, будет она применяться для составления астрономических таблиц или для коммерческих расчетов. Исключение человеческой оценки информации – это как раз путь к ее научному обсуждению, не подверженному влиянию предвзятых мнений и эмоций».

Еще:

Информация – совокупность данных, *зафиксированных на материальном носителе*, сохраненных и распространенных во времени и пространстве.

Можно продолжать перечисление очень долго, но уже достаточно, чтобы уловить тенденции определения.

Возможно также для каждой крупной решаемой проблемы давать свое определение «информации» в соответствии с конкретными решаемыми задачами; такое определение, которое «работало бы», т.е. не просто декларировало бы определение, но с его помощью можно было бы решать задачи.

Наиболее общим, философским определением термина «информации», и в то же время конструктивным, позволяющим строить производную систему терминов, применять ее в информационных теориях и количественно оценивать информацию, является следующее определение:

Информация – отраженное многообразие[23].

Здесь используются два основных свойства материи[24]:

Многообразие – множественность состояний материи.

---

[23] Урсул А. Д. Отражение и информация. М.: Мысль, 1973. 231 с.

[24] Объективная реальность, которая отображается нашими ощущениями, существуя независимо от них (объективно).

Отражение, проявляющееся в способности материальных объектов изменяться в процессе взаимодействия.

Причем, *процесс отражения* – это процесс такого воздействия (взаимодействия) одного объекта (отображаемого) на другой (отображающий), в результате которого содержание, структура и (или) особенности отображаемого объекта воспроизводятся в отображающем.

Информацию можно интерпретировать как результат отражения состояния объекта, процесса, явления, т.е. воспроизведение разнообразия и организованности одного объекта в другом объекте в результате их взаимодействия.

И еще определение[25]:

1. Информация – отраженное разнообразие, *т.е.* нарушение однообразия.

2. Информация – является одним из основных универсальных свойств материи.

Следует определиться и с термином «данные»:

Данные – информация, представленная в виде, пригодном для обработки автоматическими средствами при возможном участии человека (ГОСТ 15971–90[26]).

Или: *Данные* – факты, понятия или команды, представленные в формализованном[27] виде и позволяющие осуществлять их передачу или обработку как вручную, так и с помощью средств автоматизации (ГОСТ Р 50922-2006. Национальный стандарт Российской Федерации. Защита информации. Основные термины и определения (утв. и введен в дей-

[25] Информатика. Энциклопедический словарь для начинающих / под ред. Д. А. Поспелова. М.: Педагогика-Пресс, 1994.

[26] ГОСТ 15971–90 Системы обработки информации. Термины и определения (Утв. и введен в действие Постановлением Государственного комитета СССР по управлению качеством продукции и стандартам от 26.10.90 № 2698). М.: Издательство стандартов, 1991. 14 с.

[27] Формальный – выраженный на языке с ограниченным синтаксисом и определенной семантикой, основанной на установившихся математических понятиях (ГОСТ Р 50922–2006).

ствие приказом Федерального агентства по техническому регулированию и метрологии от 27 декабря 2006 г. № 373-ст).

И именно в этом смысле этот термин будет употребляться в текущем курсе.

В настоящее время существуют два противостоящих основных подхода (концепции) к определению информации:

Атрибутивный, утверждающий, что поскольку неорганической природе присуще универсальное свойство отражения, а отражение связано с передачей информации, то информация наличествует во всех материальных системах. «Атрибутисты» квалифицируют информацию как свойство всех материальных объектов, т.е. как атрибут материи.

Функциональный, утверждающий, что информация свойственна лишь обществу, живым существам и кибернетическим системам (т.е. там, где возможен обмен информацией), т.е. «функционалисты» связывают информацию лишь с функционированием сложных, самоорганизующихся систем.

Урсул А. Д.: (стр. 45) «Если ранее кибернетика и статистическая теория информации расширили понятие информации, отвергнув распространенное ее понимание лишь как духовного феномена, то применение теории информации в науках о неживой природе поставило вопрос о преодолении представления об информации как свойстве только кибернетических систем. В результате информация оказалась свойством не только общественных, живых и технических систем, но и вообще всех материальных систем (хотя в каждой из упомянутых областей информация имеет свою специфику). Эту позицию защищают видные советские ученые – кибернетики, в частности, А. И. Берг, В. М. Глушков, Б. Н. Петров, В. И. Сифоров и др. С этой точки зрения информация, как таковая, не возникает впервые на уровне жизни, а существует и существовала всегда».

Поэтому мы будем различать *внутреннюю* (*структурную*) и *внешнюю* (*содержательную*) информацию [16].

Свойства информации

Со свойствами информации дело обстоит примерно так же, как и с определением самого термина «информация».

Что такое «свойство»? Существует много определений термина свойства, отличающиеся строгостью и областью применения.

Толковый Словарь Ожегова: **СВОЙСТВО**: качество, признак, составляющий отличительную особенность кого/чего-нибудь.

Свойство (предмета) – атрибут (предмета);

Свойство продукции – объективная особенность продукции, которая может проявляться при ее создании, эксплуатации или потреблении (ГОСТ 15467–79).

В литературе встречаются следующие «свойства» информации:

– *достоверность* – информация достоверна, если она отражает истинное положение дел. Недостоверная информация может привести к неправильному пониманию или принятию неправильных решений (**появляется ЛПР, тогда это не столько свойство информации, сколько свойство взаимодействия информации и ЛПР**). Достоверная информация со временем может стать недостоверной, так как она обладает *свойством устаревать* (**в свойстве достоверность появляется еще свойство *старение информации***), *т.е.* перестает отражать истинное положение дел;

– *полнота* – информация полна, если ее достаточно для понимания и принятия решений (**опять ЛПР**). Как неполная, так и избыточная информация сдерживает принятие решений или может повлечь ошибки;

– *точность* – точность информации определяется степенью ее близости к реальному состоянию объекта, процесса, явления и т.п.;

– *ценность* – ценность информации зависит от того, насколько она важна для решения задачи (**ЛПР**), а также от того, насколько в дальнейшем она найдет применение в каких-либо видах деятельности человека;

– *своевременность* – только своевременно полученная информация может принести ожидаемую пользу (**ЛПР**). Одинаково нежелательны как преждевременная подача информации (когда она еще не может быть усвоена), так и ее задержка;

– *понятность* – информация становится понятной, если она выражена языком, на котором говорят те, кому предназначена эта информация (ЛПР);

– *доступность* – информация должна преподноситься в доступной (по уровню восприятия) форме (ЛПР). Поэтому одни и те же вопросы по-разному излагаются в школьных учебниках и научных изданиях;

– *краткость* – информацию по одному и тому же вопросу можно изложить кратко (сжато, без несущественных деталей) или пространно (подробно, многословно). Краткость информации необходима в справочниках, энциклопедиях, учебниках, всевозможных инструкциях.

В большинстве «свойств» характеризуются не свойства информации, как самостоятельной сущности, а свойства процесса взаимодействия информации и ее потребителя, чаще всего, – ЛПР.

Это не удивительно, поскольку *у информации два основных предназначения*: *быть источником новых знаний, быть исходными данными для принятия решений.*

Поэтому такое большое внимание уделяется процессу взаимодействия информации и ее потребителя.

Продолжим.

Национальный стандарт РФ «Информационная технология. Практические правила управления информационной безопасностью» (ГОСТ Р ИСО/МЭК 17799–2005):

Конфиденциальность: свойство информационных ресурсов, в том числе информации, связанное с тем, что они не станут доступными и не будут раскрыты для неуполномоченных лиц.

Целостность: неизменность информации в процессе ее передачи или хранения.

Доступность: свойство информационных ресурсов, в том числе информации, определяющее возможность их получения и использования по требованию уполномоченных лиц.

Еще:

1. Информация нематериальна, но обязательно связана с какой-либо материей, т.е. неотделима от материи.

Пример: размеры предмета не могут существовать сами по себе как отдельный материальный предмет.

2. Информацию можно переносить с одного вида материи на другой.

Пример: размеры предмета можно записать на какой-либо материальный носитель информации.

3. Главным переносчиком информации в пространстве является электромагнитная волна. Электромагнитное поле одновременно является информационным полем. Другие виды волн тоже способны передавать информацию.

4. В нашей Вселенной существует предел скорости передачи информации равный 300 000 км/с (округленно).

5. Информация не обладает энергией.

Но могут задать вопрос: «ведь на создание информации и на ее передачу расходуют энергию». Да, расходуют, но не на саму информацию, а на изменение поля или носителя или питание мозга.

6. С усложнением материи увеличивается и усложняется информация.

Но и материя и информация имеют предел развития (предел усложнения) и меняются циклически.

Конечно, приведенный перечень свойств информации тоже далеко не исчерпывающий.

Предлагается различать свойства структурной информации (вспоминаем атрибутивный подход) и содержательной информации (функциональный подход).

Свойства структурной информации:

1. Нематериальность и неотделимость от материи.
2. Запоминаемость (сохраняемость).
3. Передаваемость (переносимость с одного вида материи на другой) и копируемость (воспроизводимость).
4. Преобразуемость;
5. Стираемость (материя неуничтожима, она может лишь изменяться, но следы взаимодействия – отражения могут быть уничтожены).

Свойства содержательной информации

Свойства содержательной информации целесообразно определять как *качество информации*.

Качество продукции – совокупность свойств продукции, обусловливающих ее пригодность удовлетворять определенные потребности в соответствии с ее назначением (ГОСТ 15467–79[28]);

Качество (quality) – весь объем признаков и характеристик продукции или услуги, который относится к их способности удовлетворять установленным или предполагаемым потребностям (ИСО 8402). Примечание: в сфере контракта потребности определены, тогда как в других сферах *предполагаемые потребности должны быть установлены и определены* (ИСО 8402, примечание 1) (ГОСТ Р ИСО/МЭК 9126–93[29]);

Качество служебной информации – совокупность свойств служебной информации, обусловливающих ее пригодность удовлетворять

[28] ГОСТ 15467–79 Управление качеством продукции. Основные понятия. Термины и определения (утв. и введен в действие Постановлением Государственного комитета СССР по стандартам от 26 января 1979 г. № 244. Переиздание (июнь 1986 г.) с Изменением № 1, утвержденным в январе 1985 г. (ИУС 4-85).

[29] ГОСТ Р 51275–2006 Национальный стандарт Российской Федерации. Защита информации. Факторы, воздействующие на информацию. Общие положения. (утв. и введен в действие приказом Федерального агентства по техническому регулированию и метрологии от 27 декабря 2006 г. № 374-ст). М.: Стандартинформ. 2007. 7 с.

определенные потребности в соответствии с ее назначением (ГОСТ Р 51170–98, ст. 3.1.1[30]).

Для выделения основных свойств данных профессор Токийского университета Исикава Каору рекомендует вначале выделять свойства, характеризующие качество продукции на языке пользователя (так называемые «фактические показатели качества»), затем выявлять причины, влияющие на выделенные свойства, и определять составляющие качества и соответствующие им контролируемые (измеримые) показатели, сформулированные на языке персонала и разработчиков технологических систем.

С точки зрения пользователя качество данных определяется выполнением следующих условий (т.е. следующими «фактическими показателями качества»): наличие данных у пользователя в необходимый момент времени и совпадение (в пределах требуемой детализации и степени точности) информационной модели отображаемого явления с реальной действительностью.

Кроме того, часть пользователей, работающих с конфиденциальными сведениями, требует от системы сохранения соответствующих данных в тайне от других.

Кроме перечисленных, общих для всех пользователей показателей качества, могут быть выделены и другие, дополнительные показатели.

Таким образом, *качество информации*, или *свойства содержательной информации*:

Достоверность (целостность) – свойство информации не содержать скрытых ошибок. Состоит из показателей *точность* и *истинность,* как правило, находящихся в противоречии (чем точнее информация, тем менее она истинна и наоборот).

Конфиденциальность – свойство информации не быть доступной для НСД, НСИ и НСМ.

---

[30] ГОСТ Р 51170–98 Качество служебной информации. Термины и определения. Принят и введен в действие Постановлением Госстандарта России от 12 мая 1998 г. № 184. М.: ИПК Издательство стандартов. 1998. 7 с.

Кумулятивность – свойство содержать необходимые сведения в небольшом объеме информации.

Временные показатели качества:

Актуальность (*идентичность, динамичность*) – свойство информации, заключающееся в соответствии ее содержания текущему состоянию ОИ в динамике изменения этого состояния[31].

Доступность (*оперативность, своевременность*) – свойство информации, заключающееся в предоставлении ее пользователю, имеющему соответствующие полномочия, по требованию (в требуемое ему время).

Приведенный перечень технологических показателей качества содержательной информации не является исчерпывающим и в различных задачах переработки может быть изменен и дополнен в соответствии с назначением информации (например, специальная информация в виде файла, содержащего графическое изображение, имеет и такие показатели качества, как наличие или отсутствие цветности и количество цветов (оттенков, градаций серого), количество пикселей (точек) на единицу площади изображения и т.д.).

Перечень показателей качества информации определяет потребитель информации исходя из целей своего функционирования.

5.2. КОЛИЧЕСТВО ИНФОРМАЦИИ

Количество информации

Основной количественной характеристикой информации является ее количество – *количество информации*.

Для его оценки сначала рассмотрим вспомогательный термин – неопределенность (энтропия).

[31] Свойство данных соответствовать состоянию объекта; нарушение идентичности связано со старением данных по рассогласованию признаков (составляющих временных свойств) ГОСТ Р 51170–98.

Энтропия (Н) – есть информационная мера неорганизованности или разнообразия (сложности, неопределенности) ситуации или свойства объекта, процесса материального мира.

То самое многообразие из определения информации (Урсул А. Д.), множество возможных состояний.

Если пользователю информации известно о некотором объекте все, т.е. у него нет никакой неопределенности относительно состояния последнего, значит у него нет и необходимости в информации о нем.

Если же неопределенность существует, то получаемая информация призвана снизить эту неопределенность, в идеале до полной определенности.

Назовем исходную неопределенность *априорной* H_{apr}, а неопределенность после получения информации – *апостериорной* H_{aps}.

Тогда под *количеством информации*, содержащейся в сообщении, понимают разность между априорной и апостериорной неопределенностями:

$$I = H_{apr} - H_{aps}. \tag{5.1}$$

Из этого выражения следует, что

$$0 \le I \le H_{apr}\ , \tag{5.2}$$

причем:

$I = 0$ если $H_{apr} = H_{aps}$, т.е. после получения сообщения неопределенность не уменьшилась;

$I = H_{apr}$, если $H_{aps} = 0$, т.е. после получения сообщения неопределенности не осталось.

Обоснование меры неопределенности

Рассмотрим некоторое конечное множество объектов.

Простейший объект, обладающий неопределенностью, – индикатор, имеющий два состояния.

Светофор уже обладает способностью находиться в одном из трех состояний.

Дискретный канал связи может выдать символ из, например, *множества букв русского алфавита*.

Очевидно, что чем больше исходное множество возможных состояний источника информации, тем больше неопределенность о его состоянии.

Важную роль также играет равно- или разновероятность состояний.

Получение и создание информации

Установление состояния источника информации – это *получение информации*.

Но возможна и обратная операция – *синтез информации*.

При наборе текста на клавиатуре компьютера существует множество вариантов выбора, но пользователь, нажимая клавишу или сочетание клавиш, ограничивает это множество одним вариантом, что ОТРАЖАЕТСЯ на состоянии компьютера, в том числе, на экране монитора.

Таким образом – *выбор и реализация оператором некоторого варианта из множества возможных* – это создание информации.

Синтез, создание информации – это *ограничение разнообразия, организация, управление*.

Рассмотрим его еще раз.

Если возникает задача выбрать из алфавита какую-либо конкретную букву, то субъект, производящий выбор, сталкивается с определенным разнообразием возможностей. Иначе можно сказать, что само рассматриваемое множество обладает определенным разнообразием.

После того как выбор сделан (*например, выбрана буква* «*а*»), *разнообразие возможностей перестает существовать, так как множество, состоящее из одного элемента, не обладает разнообразием.*

В описанном процессе действия субъекта, производящего выбор, можно толковать как сообщение им определенной информации о выбранном элементе. Количественно эта информация по формуле (5.1) равна априорной неопределенности H_{apr}, поскольку апостериорная неопределенность H_{aps} равна нулю.

То есть в выражении $I = H_{apr} - H_{aps}$

$$H_{aps} = 0$$

и

$$I_{\text{созд}} = H_{apr}.$$

Заметим, что при получении информации $H_{aps} \geq 0$.

Аналогичные рассуждения, с соответствующим учетом особенностей, могут быть приведены для объектов, множество возможных состояний которых изменяется непрерывно, т.е. бесконечно.

Здесь уместно различать понятия – *множество состояний* объекта и множество его *различимых состояний* при данном способе информационного взаимодействия (в соответствии с определением информации Урсулом А. Д.).

Исходя из рассмотренного и логики здравого смысла, мера неопределенности должна обладать следующими свойствами или, иначе, удовлетворять следующим требованиям:

1. Быть неотрицательной

$$H \geq 0. \tag{5.3}$$

2. Быть равной нулю, если об объекте известно все, (неопределенность отсутствует), а в остальных случаях должна быть больше нуля

$$\begin{cases} H\,(N = 1) = 0; \\ H\,(N > 1) > 0, \end{cases} \tag{5.4}$$

где N – количество возможных состояний объекта;

3. Быть монотонно возрастающей функцией количества N состояний объекта

$$H_1 < N_2 \Rightarrow H(N_1) < H(N_2). \qquad (5.5)$$

4. Целесообразно, чтобы выполнялось свойство аддитивности, т.е. общая неопределенность нескольких статистически независимых систем равнялась сумме неопределенностей этих систем

$$H(N_1,\ N_2) = H(N_1) + H(N_2). \qquad (5.6)$$

Перечисленным выше требованиям могут удовлетворять различные меры энтропии.

Теперь для построения меры количества информации нам остается установить, какой именно функциональной зависимостью она связана с мощностью исходного множества.

Информационная мера неопределенности Р. Хартли (энтропия комбинаторная)

Говоря о неотрицательной монотонно возрастающей функции, мы задаем немалое разнообразие возможностей. Попробуем решить задачу, выбрав из этих возможностей самую простую.

Пусть количество информации H связано с мощностью исходного множества N линейной зависимостью $H = N$.

При $H = N$ никакой неопределенности нет (состояние одно, единственное), однако при предложенной линейной мере численно значение неопределенности равно единице, что противоречит второму требованию (5.4) к мере неопределенности.

Кроме того, введенная мера не удовлетворяет четвертому требованию (аддитивности). Для иллюстрации этого приведем пример.

Рассмотрим 32-разрядное слово в памяти вычислительной машины. Каждый разряд слова должен содержать 0 или 1. Пусть множество A состоит из всех возможных комбинаций нулей и единиц, которые могут располагаться в 32 разрядах слова. Нетрудно заметить, что таких комбинаций существует всего 2^{32}, т.е. мощность[32] N множества A равна 2^{32}:

$$N = card\ (A) = 2^{32}.$$

Следовательно, при принятой мере неопределенности выбор определенной комбинации a из множества A можно трактовать как сообщение количества информации $I(a) = 2^{32}$.

Аналогично для 64-разрядного слова $I(b) = 2^{64}$, т.е. при увеличении длины слова в 2 раза количество информации увеличилось в 2^{32} раза, а по логике следовало бы увеличение тоже в 2 раза.

Таким образом линейная зависимость между количеством элементов в множестве A и неопределенностью не удовлетворяет требованию аддитивности.

Всем перечисленным выше требованиям, в том числе и требованию аддитивности, удовлетворяет *комбинаторная мера* (*мера неопределенности Р. Хартли*).

Определение. Комбинаторная мера неопределенности есть логарифм количества состояний объекта

$$H(N) = \log_a N, \tag{5.7}$$

где $N = card\ (A)$; a – основание логарифма, выбираемое произвольно, с учетом особенностей задачи. В частности, если основание логарифма a равно 2, то единицей неопределенности $H(A)$ является бит, если десять – дит, а если логарифм натуральный ($a = e = 2{,}17\ldots$), то – нит (нат):

$H(A) = \log_2 N$, [бит];

$H(A) = \lg N$, [дит];

$H(A) = \ln N$, [нит].

[32] Выражением $Card\ (A)$ обозначается мощность, т.е. количество элементов множества A.

Ниже приведены графики логарифмической функции при различных основаниях, из которых видно что все графики качественно одинаковые: при значениях аргумента в интервале (0, 1) функции имеют отрицательное значение, при значении аргумента больше единицы – положительное, при значении аргумента, равно единице значение функций равно нулю. Основание влияет только на масштаб и не принципиально.

$L2(x) := \log(x, 2) \quad Ln(x) := \log(x, e) \quad L10(x) := \log(x, 10)$

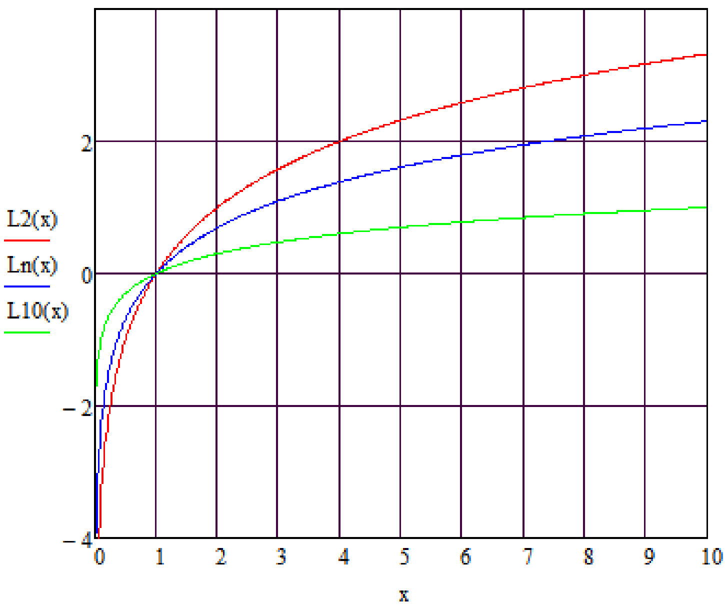


Рис. 5.1. Графики логарифмической функции

Мера неопределенности Р. Хартли удовлетворяет всем требованиям к неопределенности: неотрицательна; при одном возможном варианте, т.е. отсутствии неопределенности равна нулю ($\log_2 1 = 0$), в остальных случаях больше нуля; функция монотонно возрастающая; аддитивность также выполняется. Поясним последнее. Пусть имеется трехразрядная двоичная ячейка. Количество возможных вариантов равно $2^3 = 8$.

$H(8) = \log_2 8 = 3$ бит. Добавим в систему еще одну двоичную ячейку, количество возможных состояний которой равно двум и неопределенность равна 1 бит. Общее количество возможных состояний в системе стало 2^4 = 16. $H(16) = \log_2 16 = 4$ бит, т.е. 3+1 = 4 бит. Аддитивность выполняется.

Пример:

а) Неопределенность $H(A)$ двоичного N – разрядного слова равна N бит:

$$H(A) = log_2 2^N = N \text{ бит};$$

б) Неопределенность $H(A)$ десятичного N – разрядного числа равна N дит:

$$H(A) = lg\ 10^N = N \text{ дит}.$$

Следует помнить, что мера Р. Хартли применима только для тех случаев, множество А состояний объекта конечно

$$N = card\ (A) < \infty.$$

[Бриллюэн Л. Наука и теория информации. М.: «Государственное издательство физико-математической литературы», 1960. 391 с.] http://edu.sernam.ru/book_scin.php?id=136

«Мы рассматриваем задачу с некоторым числом возможных ответов, если мы не имеем специальной информации о действительном положении. Если окажется, что мы располагаем некоторой информацией о задаче, то число возможных ответов уменьшается, а полная информация может даже оставить нам лишь единственный возможный ответ. Информация есть функция отношения числа возможных ответов до и после (получения информации), и мы выбираем ЛОГАРИФМИЧЕСКИЙ закон для обеспечения АДДИТИВНОСТИ информации, содержащейся в независимых ситуациях».

Энтропия К. Шеннона

Пусть в качестве сообщения ожидается текст на русском языке.

Принята первая буква текста. Учитывая, что буквы алфавита имеют различную частоту встречаемости в текстах (например буква «а» и буква «ъ») количество информации в принятой букве «ъ» должно быть больше, чем в принятой букве «а». *Мера Хартли этого не учитывает*.

Статистические свойства элементов сообщения учитываются *селективной энтропией* или *статистической неопределенностью, неопределенностью К. Шеннона*.

Пусть дано множество $X = \{x_1, \ldots, x_i, \ldots, x_N\}$, состоящее из N элементов и вероятности $P(X) = \{P(x_1), \ldots, P(x_i), \ldots, P(x_N)\}$ появления этих элементов (табл. 5.1).

5.1. Распределение вероятностей *P*(*X*)
появления объектов множества *X*

X	x_1	x_2	…	x_i	…	x_N
$P(X)$	$P(x_1)$	$P(x_2)$	…	$P(x_i)$	…	$P(x_N)$

Поскольку все события появления элементов множества X составляют полную группу несовместных событий, справедливо равенство

$$\sum_{i=1}^{N} p(x_i) = 1.$$

Определение. *Частная энтропия* $H(x_i)$ *появления события* x_i (частная энтропия сообщения о том, что имеет место событие x_i)

$$H(x_i) = -\log_a(x_i). \tag{5.8}$$

Анализ выражения (6.8) показывает, что при $P(x_i)$ =1 частная энтропия равна нулю $H(x_i)$ = 0, никакой неопределенности нет, т.е. появление достоверного события x_i никакой информации не несет, в остальных случаях она больше нуля. Причем, чем «дальше» $P(x_i)$ от единицы, тем по существу неясность (неопределенность) больше и численно значение $H(x_i)$ также больше.

Определение. ***Статистическая энтропия Шеннона*** **(*средняя энтропия*)** *источника дискретных сообщений есть взвешенная сумма частных энтропий появления всех событий* (*математическое ожидание*)

$$H(X) = -\sum_{i=1}^{N} P(x_i)\log_a P(x_i)\,. \tag{5.9}$$

Пусть дано множество X, состоящее из двух элементов $X = \{x_1, x_2\}$ с соответствующими вероятностями появления элементов:

$$P(x_1) = P,\ P(x_2) = 1 - P.$$

Тогда зависимость $H(P)$ неопределенности $H(X)$ от вероятности P имеет вид, представленный на рис. 5.2.

Можно заметить, что энтропия Хартли есть частный случай энтропии Шеннона при $\forall i$: $P(x_i) = 1/N$, т.е. когда все события равновероятны, что соответствует наибольшей неопределенности

$$H(X) = -\sum_{i=1}^{N}\frac{1}{N}\log_a\frac{1}{N} = \sum_{i=1}^{N}\frac{1}{N}\log_a N = \log_a N. \tag{5.10}$$

С введенным понятием тесно связан ряд других производных мер.

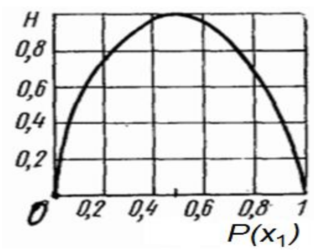


Рис. 5.2. Зависимость средней энтропии источника двух сообщений от вероятности появления первого сообщения

Определение. *Избыточность источника сообщений*

$$R(X) = \frac{\log N - H(X)}{\log N}.$$

Избыточность – единственный способ повышения достоверности информации.

В криптографической области избыточность облегчает криптоаналитику взлом шифра и поэтому является негативным качеством.

Пусть имеются два множества X и Y и известны: вероятность $P(x_i)$ появления элемента $x_i \in X$, вероятность $P(y_j)$ появления элемента $y_j \in Y$ и условная вероятность[33] $P(x_i/y_j)$ появления x_i при известном исходе y_j.

Пример: $X = Y$ – множество букв алфавита, y_j – буква, появившаяся на мониторе, x_i – буква, появляющаяся после буквы y_j.

Определение. *Энтропия условная* – энтропия, определяемая при известном исходе другой ситуации:

$$H(X/Y) = -\sum_{j=1}^{M} P(y_j) \sum_{i=1}^{N} P(x_i/y_j) \log_a P(x_i/y_j). \qquad (5.10)$$

$H(X/Y) \leq H(X)$ и при статистической независимости X и Y в Y нет информации о X и $H(X/Y) = H(X)$.

Если же X и Y взаимосвязаны функционально, т.е. по известному y_j точно определяется x_i, то $H(X/Y) = 0$.

Определение. *Энтропия совместная* двух случайных событий есть величина, определяемая выражением

$$H(X, Y) = H(Y) + H(X/Y) = H(X) + H(Y/X). \qquad (5.11)$$

---

[33] Справка. Из теории вероятности следует, что $P(x_i) = \sum_{j=1}^{M} P(x_i/y_j)$.

Анализ этого выражения показывает, что $H(X, Y) \leq H(X) + H(Y)$ и при статистической независимости X и Y

$$H(X/Y) = H(X) + H(Y). \tag{5.12}$$

Определение. *Взаимная информация – информация, содержащаяся в множестве Y относительно множества X*

$$I(Y/X) = H(X) - H(X/Y), \tag{5.13}$$

т.е. насколько уменьшится неопределенность множества X при снятии неопределенности с множества Y.

Из (5.11) и (5.13) следует равенство

$$I(Y/X) = H(X) + H(Y) - H(X, Y) = I(X/Y). \tag{5.14}$$

Из (5.12) и (5.14) следует, что если Y и X статистически независимы, то взаимная информация равна нулю $I(Y, X) = 0$.

Приведенная мера неопределенности, так же как и мера P. Хартли применима для объектов *с конечным количеством состояний*. Однако, заменой *вероятностей* на *плотности вероятностей* мера неопределенности по К. Шеннону может быть распространена на объекты, состояния которых являются *непрерывными случайными величинами*. В этом случае используется *дифференциальная энтропия*.

Дифференциальная энтропия источника непрерывной случайной величины:

$$H_{Д}(X) = -\int_{-\infty}^{+\infty} p(x) \log_a p(x)\, dx. \tag{5.15}$$

Энтропия принадлежности случайной величины интервалу (x_1, x_2)

$$H(x, x \in (x_1, x_2)) = H(x, x \in (-\infty, x_2)) - H(x, x \in (-\infty, x_1)) =$$

$$= -\int_{x_1}^{x_2} p(x) \log_a p(x) dx. \tag{5.15}$$

Мера неопределенности К. Шеннона лежит в основе множества математических выражений, позволяющих решать многие прикладные задачи в областях связанных с информацией: рассчитывать скорость передачи информации по каналу с помехами, осуществлять оптимальное кодирование, решать вопросы в области криптографии и т.д.

5.3. СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕОРИИ

Шеннон К. разрабатывал свою теорию для практических нужд связи, где важно дословно, *без искажений*, передать текст, и не важно, понял ли получатель смысл этого текста.

Проанализируем текст «Завтра будет буря». Осмысленность или информация текста «Завтра будет буря» очевидна. Достаточно, однако, сохранив все элементы (буквы) этого сообщения, переставить их случайным образом, например, «рдеа Звубуб траяи», как оно утратит всякий смысл. Но бессмысленной информации не бывает. Согласно формуле Шеннона оба предложения содержат одинаковое «количество информации». О какой же информации здесь идет речь? *Или, вообще, можно ли говорить об информации по отношению к разрозненным элементам сообщения*?..

Недостаточность Шенноновского подхода привела к появлению теорий, исследующих *смысловое содержание информации, ее ценности*.

Основы семиотической теории информации

Автором самой известной из них является Колин Черри.

В курсе лекций 1950-х годов и опубликованной монографии «Человек и информация» (перевод со второго издания в 1972 г.) К. Черри удается показать далеко не тривиальную сложность элементарного речевого общения, связанную с зависимостью передачи сообщения и его восприятия от огромного числа субъективных факторов и факторов среды.

Это позволяет автору сделать вывод о невозможности прямого переноса идей математической теории связи на человеческое общение. Она, по его мнению, «…может помочь в изучении обмена информацией между людьми, но не может дать объяснения сущности этого обмена…». Одним из оснований для такого вывода является то обстоятельство, *что математическая теория информации игнорирует семантическую и прагматическую сторону сообщений*.

Его теория называется «Семиотическая теория информации».

В дальнейшем множество ученых работало в области. Среди них – Юлий Анатольевич Шрейдер – математик (в 1950 году защитил кандидатскую диссертацию по функциональному анализу), специалист по информатике, семиотике, методологии науки и теории информации, публицист, впоследствии – философ (доктор философских наук – защитил докторскую диссертацию по философии в 1981 году) и богослов.

Юрий Сергеевич Степанов (род. 20 июля 1930, Москва) – российский ученый, филолог, академик АН СССР (с 15 декабря 1990 года; с 1991 – академик РАН), доктор филологических наук, профессор. Труды в области теоретической лингвистики, сравнительно-исторического индоевропейского языкознания, романской филологии и семиотики и др.

В настоящее время под *семиотикой* больше понимают связь знаков с содержащимся в них значением.

Семиотическая теория информации К. Черри исследует информацию на трех качественных уровнях: синтаксическом, семантическом и прагматическом и, соответственно, включает три раздела: *синтактику, семантику, прагматику*.

Определение. *Синтактика* – раздел семиотической теории информации, в котором исследуются символы, знаки, правила записи слов и предложений, связываются формальные свойства знаков и их комбинаций с количеством содержащейся в них информации. (Аналогичное значение имеет термин «синтаксис» в грамматике русского языка и в искусственных языках, например – языках программирования).

Определение. *Семантика* – раздел семиотической теории информации, в котором исследуется связь знаков с содержащимся в них смыслом, с объектами и их свойствами.

Пример: «10» – двоичное и «2» десятичное, «Дом» и «Home» – на синтаксическом уровне это различные объекты, а на семантическом – один и тот же.

К. Черри различает три подуровня семантики:

поверхностный – буквальный смысл слов и предложений;

глубокий – эмерджентность текста в целом;

диалоговый – уровень понимания в процессе общения с текстом.

Если *отражение многообразия* присуще всем информационным системам, то для кибернетической (в большей мере для живой) системы существует еще *разнообразие отражения* как функция ее *тезауруса*[34]. Это отражение – интерпретация «увиденного». Сам тезаурус тоже изменяется при получении информации:

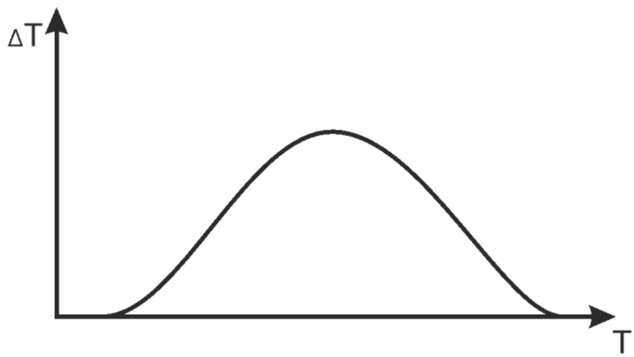


Рис. 5.3. Зависимость прироста тезауруса ΔT потребителя информации при прочтении текста от размера тезауруса Т

Интерпретация: если раньше потребитель информации не знал ничего, то новая информация не приведет к приросту тезауруса, так как потребитель не поймет смысл текста; если он что-то знает, то поймет содержание текста и его тезаурус увеличится, если же знает очень много, то полученное сообщение вряд ли добавит в тезаурус что-то новое.

[34] **Тезаурус** – совокупность сведений, которой обладает получатель информации, его понятийный аппарат.

Определение. *Прагматика* – раздел семиотической теории информации, в котором исследуются вопросы ценности информации, связь знаков с их потребительской полезностью для потребителя, вопросы достаточности информации.

Прилагательные «полезная», «бесполезная», «ценная» и т.д., применяемые к информации, предполагают наличие потребителя (полезная для кого?), в отличие от определения «точная», «надежная» и т.д. (надежная с вероятностью 0,98 безотносительно кого-либо). Это уровни прагматический и семантический соответственно.

Определение ценности информации. Апостериорный подход

Поскольку информация предназначена для удовлетворения определенных потребностей пользователей, важным вопросом является исследование ее на *прагматическом уровне*.

Может ли существовать единое определение ценности (и мера ценности) информации?

По-видимому, нет. Существует множество совершенно не сходных между собой ситуаций (и соответственно разных мер информации), в которых использование информации приносит (или может принести) выгоду. Этим объясняется, что уже сейчас существует немало определений понятия ценности информации, а появление новых продолжается.

1. Ценность = количество.
2. Апостериорный подход.
3. Ценность количества.

Апостериорный подход

Рассмотрим некоторую целесообразно действующую систему.

Вместо того чтобы количественно оценивать информационные сообщения, поступающие в систему, можно оценивать *действия* (их выгоду или невыгоду), осуществляемые системой на основании (как предполагается) поступившей информации, *не заботясь о том, что представляет собой и как измеряется поступившая информация*.

Естественно считать, что действия системы (и связанный с ними выигрыш) косвенно характеризуют и информацию, на основании которой эти действия совершаются.

Иными словами, рассматривается только «реализованная» системой информация. Реализация же представляет собой акт выбора системой одного из допустимых действий, вариантов поведения и т.д. Такой подход к оценке количества (или ценности) информации называется ***апостериорным***.

Отметим интересную особенность апостериорного подхода. Он позволяет – наряду с полезной информацией – рассматривать (и количественно оценивать) и «*дезинформацию*» и «*псевдоинформацию*», поступающие в систему. Поступление «дезинформации» проявляется в таких действиях, которые ухудшают положение системы, ведут к росту проигрыша и т.д., «псевдоинформация» не влияет на положение системы.

Следует обратить внимание, что статистическая теория информации (К. Шеннона) имеет дело только с неотрицательными величинами – энтропией и количеством информации, а потому и не может описать важный случай поступления дезинформации[35].

Апостериорный подход является, конечно, лишь общим принципом – слишком общим, чтобы применяться непосредственно. Но он дает возможность построить классификацию мер количества и ценности информации и прояснить сходство и различие между ними достаточно четко и с единой точки зрения.

Глубокое, по мнению автора, определение ценности информации приводится в [Ловцов Д. А. Проблема обеспечения информационной безопасности России. http://www.nasled.ru/pressa/obozrev/N02_99/2_10.HTM]:

«*Под ценностью информации в эргасистеме понимается значимость информации, определяемая способом динамического отображе-*

[35] Существуют модификации Шенноновской теории, позволяющие сделать это, основаны на применении апостериорного подхода.

ния множества ее качественных свойств и количественных характеристик на множество возможных управляющих решений, ведущих к достижению целей управления».

Основные меры ценности информации

Одной из мер, основанных в полной мере на апостериорном подходе, является мера А. А. Харкевича (Бонгарта[36]–Харкевича[37]).

Мера представляет интерес для изучения потому, что кроме того, что полностью основана на апостериорном подходе, совмещает элементы «чистой теории информации» и «теории абсолютной ценности информации», т.е. первого и второго из рассмотренных выше направлений, а также позволяет различать понятия «доброкачественная информация», «дезинформация», «пустая информация».

Аппарат теории применим в тех случаях, когда единственной задачей, стоящей перед системой, является достижение определенной цели (примеры: «пройти лабиринт», «достичь аэропорта назначения», «выиграть в игре»). При этом выигрыш или проигрыш не может быть описан в виде приращения или потерь одних лишь материальных ресурсов. Особенно отчетливо невозможность такого описания видна тогда, когда все средства, которыми располагает информационная система, направлены на избежание катастрофического проигрыша.

Итак, считается, что на основании *некоторой* информации, поступившей в систему (информации, природа которой нам безразлична), система принимает решение, изменяющее вероятность достижения цели. Физическая природа сигналов, логическая структура сообщений, их длина полностью игнорируются!

Естественно требовать, чтобы увеличению вероятности достижения цели отвечал случай положительного значения вводимой меры, умень-

---

[36] Бонгарт М. М. Проблемы узнавания. М.: Наука, 1967. 258 с.

[37] Александр Александрович Харкевич (21 января (3 февраля) 1904, Санкт-Петербург – 30 марта 1965, Москва) – советский ученый, член-корреспондент АН УССР, действительный член АН СССР, профессор, к 1951 г. – руководитель отдела технической физики киевского Института физики АН УССР.

шению вероятности – отрицательного значения, а сохранению прежнего значения вероятности – нулевого значения меры. Этим требованиям удовлетворяет *мера целесообразности управления* (одновременно являющаяся и мерой ценности информации, на основании которой система принимает решение), определяемая как

$$I = \log_a = \frac{P_1}{P_0},$$

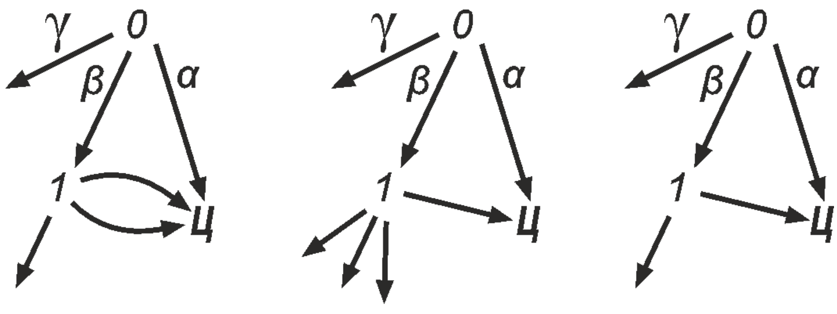
где P_1 и P_0 – вероятности достижения при управлении цели после и до получения сообщения соответственно.

Неопределенность здесь имеет вид «$-\log_a P$», где P – вероятность достижения объектом цели в данный момент и из данного положения.

Целесообразность управления выражается уменьшением неопределенности.

$$I = H_{apr} - H_{aps} = -\log_a P_0 - (-\log_a P_1) = \log_a P_1 - \log_a P_0 = \log_a P_1/P_0.$$

Рассмотрим пример (рис. 5.4).



а) *б*) *в*)

Рис. 5.4. К примеру целесообразности управления

Пусть объект может двигаться только в направлении дуг. При отсутствии информации выбор любого из возможных направлений производится с равными вероятностями. Точка *О* – исходное положение, точка *Ц* – цель. Висящие дуги соответствуют направлениям, выбрав которые объект полностью лишается возможности достижения цели.

Из точки *О* объект может попасть в *Ц* сразу (дуга α), никогда не попасть (дуга γ) или перейти в точку *1*, сохранив возможность достижения цели на следующем шаге (дуга β). Тогда целесообразность управления, переводящего объект из *О* в *1* (и, следовательно, ценность информации, на основании которой принято такое решение), будет равна (рис. 5.4, *а*)

$$I = \log_2 \frac{2/3}{1/3 + 1/3 \cdot 2/3} > 0,$$

так как P_1 = 2/3 (из точки *1* в точку Ц можно попасть двумя способами из трех возможных); P_0 = 1/3 + 1/3 · 2/3 (из точки *О* в точку *Ц* можно попасть непосредственно с вероятностью 1/3 и с вероятностью 1/3 сначала попасть в точку 1 и затем с вероятностью 2/3 в точку *Ц*).

Поскольку $I > 0$, то это случай «доброкачественной информации».

На рисунке 5.4, *б* представлен случай «дезинформации».

$$I = \log_2 \frac{1/4}{1/3 + 1/3 \cdot 1/4} < 0,$$

а на рис. 5.4, *в* случай «пустой информации», «псевдоинформации»:

$$I = \log_2 \frac{1/2}{1/3 + 1/3 \cdot 1/2} = 0.$$

Еще одной известной мерой ценности информации является выражение, предложенное В. И. Корогодиным[38]

$$I = \frac{P_1 - P_0}{1 - P_0}, \quad 0 \le I \le 1.$$

[38] Владимир Иванович Корогодин (4.01.1929 – 31.10.2005). Корогодин В. И., Корогодина В. Л. Информация как основа жизни. Дубна: Изд-во «Феникс», 2000. 208 с.

Основные положения теоретико-множественного представления информации

Эта теория является относительно новой, имеющей определенные перспективы.

Область ее применения – обеспечение качества информации в технологических процессах ее переработки, особенно при распределенных преобразованиях.

Учитывается семантический и прагматический уровни рассмотрения информации. Основоположником теории является д.т.н., профессор, академик А. В. Чечкин.

Исходя из определения информации, она всегда о чем-то, о каком-либо *объекте*, далее называемом **«*объект информации*» (*ОИ*)**.

Объект информации – первичное, строго не определяемое понятие, на природу которого не накладывается никаких требований: это могут быть реальные или идеальные объекты, в частности, математические. Очевидно, ОИ должен обладать некоторыми вариациями (состояниями), иначе бы не было смысла говорить о нем (требование многообразия состояний объекта). Множество возможных вариаций ОИ далее называется *опорным множеством состояний ОИ*.

Итак, пусть дано:

R – опорное множество, множество состояний объекта;

x – элемент множества R, $x \in R$, состояние объекта информации;

подмножество $\delta \subset R$, $x \in \delta$.

Если δ содержит те и только те точки R, которые обладают некоторым свойством, то возможно отождествление этого *подмножества* с понятием «*свойство*», причем истинное высказывание «точка x из R обладает свойством δ» является *элементарным сведением* о точке x (о состоянии объекта) и записывается в виде одноместного предиката $\delta(x)$.

Определение. *Любое подмножество*

$$\delta(x): x \in \delta(x),\ \delta(x) \subset R$$

есть *элементарное сведение о точке x.*

Пример.

1. Множество шаров в урне. Подмножества – свойства: белые, деревянные, большие и т.д.

2. Самолет: Опорное множество – все самолеты мира, сведения: 4-х моторный, 8 человек экипажа, 1913 г. выпуска.

На основе введенного термина «*элементарное сведение*» определен термин «*элементарная информация*».

Определение. *Элементарная информация о точке* x_0 – непустое семейство подмножеств $J(x_0) = \{\delta_i\ (x_0)\}$ – элементарных сведений о точке x_0, удовлетворяющее условию $\forall i : \delta\ (x_0) \subset \delta_i\ (x_0)$. То есть все надмножества подмножества $\delta\ (x_0)$. Интерпретация – все умозаключения, которые можно сделать на основе данного сведения.

Определение. *Неэлементарным сведением* о точке x_0 является предикат $<p, \delta\ (x_0)>$, где p – вероятность того, что $x_0 \in \delta$.

Мера неопределенности вводится следующим образом.

Пусть опорное множество $R' \subset R^1$ – компакт[39] на действительной числовой оси. Например, множество допустимых числовых значений какого-либо параметра технической системы. Тогда точка $x \in R'$ – значение этого параметра; подмножество – компакт δ: $x \in \delta$, $\delta \subset R'$ – элементарное сведение о значении числового параметра.

Если множество R' измеримо, то введем на нем *меру*[40] (а отрезок на числовой оси есть компакт и поэтому измерИм), которая может служить *мерой неопределенности* $H(\delta\ (x))$ *сведения* $\delta\,(x)$ *о точке* x.

---

[39] К математическим объектам – «компактам» – относят, в частности, все замкнутые ограниченные подмножества евклидова пространства любой конечной размерности.

[40] Математическое понятие «меры» является обобщением понятий: длины отрезка, площади плоской фигуры, объема пространственного тела, приращения неубывающей функции на отрезке или полуинтервале, интеграла от неотрицательной функции, взятого по некоторой области и т.д.

Определение. *Мера неопределенности* $H(\delta(x))$ *сведения* $\delta(x)$ *о точке* x есть мера множества $\delta(x)$

$$H(\delta(x)) = mes(\delta(x)).$$

Из введенных терминов следует, что сведение о действительном числе есть некоторый отрезок, содержащий это число, а неопределенность рассматриваемого сведения есть длина этого отрезка.

Введенная мера соответствует общим требованиям к мерам неопределенности:

1. $H(\delta(x)) \geq 0$ (по определению меры);
2. $H(\delta(x)) = 0 \Leftrightarrow \delta(x) \equiv x$ (по определению меры);
3. $\exists\ M = H_{\max} > 0$: $H(\delta(x)) \leq M$ (следует из определений меры и компакта).

Ограничения применения меры: для ОИ – действительных чисел – допускается вектор чисел. Целесообразно применение при расчетах по обеспечению конфиденциальности и точности при преобразованиях числовой информации.

Инвариантная модель информационного отношения субъектов с учетом качественных уровней информации

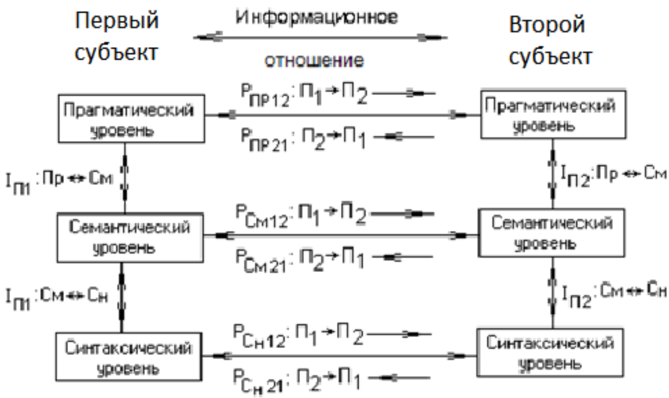


Рис. 5.5. Структура модели информационного отношения субъектов с учетом качественных уровней информации

Классификация мер количества информации

Изученные меры неопределенности, безусловно, не исчерпывают всего множества известных мер неопределенностей.

Проведем классификацию мер неопределенности:

- **комбинаторные меры:** к ним относится рассмотренная мера Р. Хартли, а также мера Ю. И. Шрейдера – комбинаторная (определяемая также через логарифм) мера разнообразия общесистемного тезауруса;
- **вероятностно-статистические меры:** К. Шеннона, А. Харкевича (вероятностная мера целесообразности управления), С. Кульбака-Лейблера (неопределенности распределения вероятностей).

Справка: Для дискретных вероятностных распределений P и Q с числом элементарных событий n расхождение Кульбака–Лейблера распределения Q относительно распределения P (или «расстояние от P до Q») определяется как:

$$D_{KL}(P \parallel Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}.$$

Данная мера расстояния в теории информации также интерпретируется как величина потерь информации при замене истинного распределения Р на распределение-модель Q.

- **меры сложности объектов:**

– алгоритмическая мера (А. Колмогорова) – мера сложности восстановления некоторого слова S из определенного алфавита по фиксированной оптимальной вычислимой функции $S = f(z)$, равная длине самого короткого слова z, в котором содержится полное описание слова S.

6. МОДЕЛИРОВАНИЕ В НАУЧНЫХ ИССЛЕДОВАНИЯХ

Одним из важнейших этапов научного исследования является моделирование. Модель, объективно обладая сходством с оригиналом, может с той или иной степенью полноты имитировать оригинал и позволяет проводить исследования объекта по его модели.

6.1. ОПРЕДЕЛЕНИЕ И КЛАССИФИКАЦИЯ МОДЕЛЕЙ

При необходимости исследования объектов, в случаях, когда исследование непосредственно самого объекта по каким-то причинам невозможно, затруднительно, дорого, требует недопустимо большого времени и т.п., применяют модели объектов.

Модель – это представление объекта или процесса в определенной форме, позволяющей проводить требуемые для исследования эксперименты.

Таким образом:

1. Модель – это в определенном смысле образ исследуемого объекта, предназначенный для того, чтобы проводить исследования не на объекте, а на его модели.
2. Модель является инструментом определенных исследований.

Поэтому в модели необходимо обеспечить подобие моделируемого объекта в сходстве характеристик, функций или форме описания поведения *в той области, в которой предполагается проводить исследования*. Это обеспечивает адекватность модели объекту. И в модели не должны быть сохранены те свойства моделируемого объекта, из-за которых невозможно или нецелесообразно проведение исследований на самом объекте.

Для различных исследований может быть построено несколько различных моделей объекта.

Определение.

Объект М есть модель объекта О, если:

1) М и О не идентичны друг другу,

2) М отвечает на вопросы относительно О (М позволяет делать верные умозаключения относительно О).

При моделировании различают 2 основных этапа:

1. Создание модели.

1.1. Проектирование.

1.2. Создание.

1.3. Проверка адекватности.

2. Проведение эксперимента.

2.1. Планирование эксперимента.

2.2. Проведение эксперимента.

2.3. Интерпретация результатов эксперимента.

Классификация моделей

В зависимости от средств реализации модели разделяют на:

1. *Материальные* (физические) модели – основаны на реализации *статических* и *динамических* свойств объекта моделирования.

Статические макеты представляют собой выполненные в натуральную величину модели оборудования или его отдельных частей (блоков), которые подвергаются проверке.

Функциональные макеты – это модели оборудования в натуральную величину, которые могут воспроизводить реальное функционирование аппаратуры.

2. *Абстрактные* модели – описание объекта моделирования (ОМ).

2.1. *Формальные* – описание с помощью формальных языков:

2.1.1. Математические – представление ОМ с использованием формализмов математики;

Все физические законы есть математические модели.

2.1.2. С использованием иных формальных языков.

2.2. *Неформальные*:

– *вербальные* – представляют ОМ в виде словесных описаний;

– *иконографические* – представляют ОМ в виде схем, графиков, изображений

Физическое моделирование в силу своей природы позволяет наиболее полно учесть многочисленные факторы, влияющие на исследуемый процесс. Поэтому, при физическом моделировании исследователю могут открыться какие-либо стороны объекта исследования, о которых он не знал.

Однако в некоторых случаях работа с физическими моделями является сложной и трудоемкой и, как правило, сопряжена с большими затратами времени и материальных средств. Кроме того, физическое моделирование принципиально возможно только при моделировании физических процессов.

Математические модели позволяют ликвидировать некоторые из указанных недостатков, хотя они и не обладают, как правило, полнотой физических моделей. Предпосылкой к использованию математических моделей при исследованиях вообще и моделировании в частности является *единство законов природы, объединяющих в некотором отношении далекие друг от друга явления, и одинаковость формы описывающих их уравнений*. Однако, в математическую модель исследователь, в отличие от физической модели, может ввести только то, что он уже знает.

Математические модели. Детерминированные и стохастические

Статические и динамические

Статические математические модели **–** модели, в математическом аппарате которых отсутствует параметр – время. Например – физические законы, связывающие температуру и давление, топология, геометрия, различные матричные операции и т.д.

В *динамической* системе обязательно присутствует параметр – время.

Динамическая система может быть представлена в виде следующей модели.

Дано:

1. Множество входных воздействий $X(t) = (x_1(t), x_2(t), ..., x_n(t))$.
2. Множество воздействий внешней среды $V(t) = (v_1(t), v_2(t), ..., v_m(t))$.
3. Множество выходных характеристик $Y(t) = (y_1(t), y_2(t), ..., y_k(t))$.
4. Множество состояний системы $Z(t) = (z_1(t), z_2(t), ..., z_l(t))$.

Состояние системы может изменяться как под действием внешних факторов, так и под влиянием внутренних причин.

5. Операторы выхода и перехода:

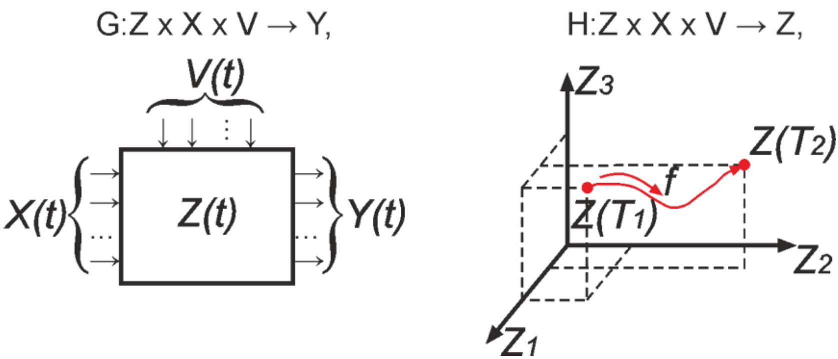


Рис. 6.1. Иллюстрация к вопросу о моделях динамических систем

На множестве состояний системы строится *пространство состояний* и текущее состояние является точкой данного пространства.

Совокупность точек, отражающих состояние системы за некоторый интервал времени, называется траекторией состояний.

Если в пространстве состояний ввести координату – *время*, то полученное пространство называется *фазовым пространством*.

Если значения входных воздействий и воздействий внешней среды детерминированы и детерминирована реакция системы на них, то такие системы называют *детерминированными динамическими системами* (*например – конечные автоматы*).

Иные системы называются *стохастические динамические системы* (СДС).

Это:

– системы с детерминированной реакцией и случайными входными воздействиями X, или случайными воздействиями V;

– системы со стохастическим поведением под действием детерминированных воздействий;

– системы со стохастическим поведением под действием стохастических воздействий.

Аналитические, алгоритмические и имитационные

Обычно в виде функций задаются общие законы природы или общие закономерности.

В самом общем виде математическая модель (ММ) представляет собой кортеж, содержащий набор сущностей и зависимость между этими сущностями, параметрами:

$$M = \langle X, Y, G, F \rangle,$$
$$F(x, y, g) = 0.$$

Из нее, как правило (но не всегда), можно построить различные варианты

$$y = F_y(x, g),\ g = F_g(x, y).$$

Моделирование реализуется в виде вычислительного эксперимента.

Примером такой модели может служить знаменитая формула К. Э. Циолковского:

$$\Delta v_{\text{ЛА}} = v \ln \frac{M_0}{M_{\text{к}}},$$

определяющая приращение скорости ракеты $\Delta v_{ЛА}$(*v-ню*) при импульсном сжигании топлива через скорость истечения рабочего тела v и отношение начальной M_0 и конечной $M_к$ масс ракеты.

Аналитические модели могут быть реализованы в виде аналитического выражения (формулы) или алгоритмически.

Например, решение квадратного уравнения в виде аналитического выражения

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

или алгоритмически.

Стохастические процессы также могут быть смоделированы аналитическими моделями с использованием формализма теории вероятностей.

Имитационные модели

При моделировании систем и процессов желательно получать модель в виде аналитических выражений, однако это не всегда возможно.

Имитационные модели реализуются алгоритмически.

Например, при исследовании функционирования сложной системы, состоящей из множества простых элементов, вероятностные характеристики (законы распределения) которых известны, невозможно построить аналитические выражения для закона распределения ее состояния (выходов), а исследование ее в реальных условиях связано с неприемлемо большими затратами ресурсов.

В этом случае надо моделировать сущность процесса функционирования системы. Такое моделирование называется *имитационным*, а конкретную имитацию реализации процесса (функционирования системы) – *имитационных экспериментом* или *реализацией* процесса.

Пример: расчет вероятности поражения двумя выстрелами. Возможно аналитическое и имитационное моделирование.

Выбор и обоснование математической модели представляет собой задачу, трудность которой определяется наличием и степенью изученности процессов, подобных моделируемому, наличием неопределенностей, сопровождающих данный конкретный процесс.

6.2. СВОЙСТВА МОДЕЛЕЙ

Модели должны обладать определенными свойствами, т.е. удовлетворять определенным требованиям.

Общие требования к моделям

Адекватность – степень соответствия модели реальным характеристикам объекта.

Проверяется следующими способами[41]:

– проверкой непротиворечивости характеристик модели физическим законам и известным ограничениям (например, вероятность не должна быть больше единицы);

– сравнением с такими же характеристиками других, уже проверенных моделей или реальных моделируемых объектов;

– сравнением с моделями сходных объектов и экстраполяция их характеристик в сторону исследуемого объекта;

– и т.д.

Конечность – модель отображает оригинал лишь в конечном числе его отношений. Любая модель имеет ограничения. Эти ограничения устанавливает разработчик в зависимости от целей исследования.

Универсальность – модель должна описывать не отдельный объект, а определенный класс объектов, характеризующийся вполне определенными признаками.

Экономичность – характеристика затрат ресурсов на реализацию модели.

[41] Чугунов А. В. Теория электронных вычислительных машин и вычислительных систем. М.: МО СССР. 1990. 487 с. С.138–139.

Модель – это инструмент исследования и к ней могут предъявляться и другие требования в зависимости от предназначения модели.

Требования к математическим моделям

Адекватность ММ – это способность ММ описывать выходные параметры с относительной погрешностью не более некоторого заданного значения δ.

Множество входных данных, на котором обеспечивается условие требуемой точности, называется *областью адекватности* данной ММ.

В более общем смысле под адекватностью ММ понимают правильное качественное и достаточно (для целей моделирования) точное количественное описание именно тех характеристик моделируемого объекта, которые важны в данном конкретном случае. Модель, адекватная при выборе одних характеристик, может быть неадекватной при выборе других характеристик того же моделируемого объекта.

Точность – одна из составляющих адекватности, оценивается степенью совпадения результатов моделирования с экспериментальными данными.

Относительная погрешность для i-го параметра можно определить по формуле

$$\varepsilon_i = \left| \frac{y_{i\text{м}} - y_{i\text{э}}}{y_{i\text{э}}} \right|,$$

где $y_{i\text{м}}$, $y_{i\text{э}}$ – значения параметра y_i, рассчитанное с помощью модели и экспериментальное соответственно.

Для n параметров в качестве оценки точности можно определить

$$\varepsilon = \max_{i=\overline{1,n}} \varepsilon_i \quad \text{или} \quad \varepsilon = \sqrt{\sum_{i=1}^{n} \varepsilon_i^2}.$$

Конечность. Формула не может описывать всех свойств моделируемого объекта. Формула (модель) должна быть адекватна только исследуемым свойствам объекта.

Универсальность – способность математической модели описывать не отдельный объект, а определенный класс объектов, характеризующийся определенными признаками.

Экономичность ММ оценивают затратами на вычислительные ресурсы (машинное время и память), необходимые для реализации ММ на ЭВМ.

Эти затраты зависят от числа арифметических операций при использовании модели, от размерности пространства фазовых переменных, от особенностей применяемой ЭВМ и других факторов.

Очевидно, что требования экономичности высокой точности и достаточно широкой области адекватности ММ противоречивы и на практике могут быть удовлетворены лишь на основе разумного компромисса.

Робастность (*устойчивость*) (от английского слова robust – крепкий, устойчивый) характеризует ее устойчивость по отношению к погрешностям исходных данных, способность нивелировать эти погрешности и не допускать их чрезмерного влияния на результат вычислительного эксперимента.

Причинами низкой робастности ММ могут быть:

– необходимость при ее количественном анализе вычитания близких друг к другу приближенных значений величин;

– деление на малую по модулю величину;

– использование в ММ функций, быстро изменяющихся в промежутке, где значение аргумента известно с невысокой точностью;

– большие числа обусловленности матриц.

Наглядность ММ является ее желательным, но необязательным свойством. Тем не менее использование ММ и ее модификация упрощаются, если ее составляющие (например, отдельные члены уравнений) имеют ясный содержательный смысл. Это обычно позволяет ориентировочно предвидеть результаты вычислительного эксперимента и облегчает контроль их правильности.

6.3. МОДЕЛИ, ОПИСЫВАЕМЫЕ АЛГЕБРАИЧЕСКИМИ И ТРАНСЦЕНДЕНТНЫМИ УРАВНЕНИЯМИ

Рассмотрим вначале наиболее простые математические модели, в которых моделируемая величина y явно выражается через другие параметры модели:

$$y = f(\bar{a},\ \bar{x}) + \varphi(\bar{b},\ \bar{x})\eta, \tag{6.1}$$

где $f(\bar{a},\ \bar{x})$ и $\varphi(\bar{b},\ \bar{x})$ – некоторые детерминированные функции; $\bar{x} = (x_1,\ x_1,\ ...,\ x_m)$ – некоторые известные параметры, одним из которых является время t, в качестве остальных параметров могут выступать различные известные величины, влияющие на развитие величины y; $\bar{a} = (a_1,\ a_1,\ ...,\ a_n)$ и $\bar{b} = (b_1,\ b_1,\ ...,\ b_p)$ – неизвестные, подлежащие определению *во время построения модели* параметры (коэффициенты); η – случайный процесс с нулевым математическим ожиданием.

Функция $f(\bar{a},\ \bar{x})$ является детерминированной основой моделируемого процесса y. Она характеризует значения, которые имела бы величина y, если бы она не подвергалась воздействию случайной помехи η.

Функция $\varphi(\bar{b},\ \bar{x})$ характеризует воздействие случайной помехи η. В различных практических задачах о η (кроме равенства нулю его математического ожидания) могут быть сделаны другие предположения:

– η почти всегда имеет постоянную дисперсию;
– η обычно имеет нормальное распределение;
– η – очень часто некоррелированный случайный процесс.

Наиболее распространенной в настоящее время моделью является модель с аддитивной (суммарно входящей) помехой вида

$$y = f(\bar{a},\ \bar{x}) + \eta,\ \ \varphi(\bar{b},\ \bar{x}) \equiv 1. \tag{6.2}$$

Случай $\varphi(\bar{b},\ \bar{x}) = 0$ соответствует детерминированному процессу, который редко встречается в практике.

На практике встречаются также модели вида

$$y = f(\bar{a}, \bar{x}) \cdot \varphi(\bar{b}, \bar{x})^{\eta}, \qquad (6.3)$$

которые в некоторых случаях (когда $f(\bar{a}, \bar{x}) > 0$ и $\varphi(\bar{b}, \bar{x}) > 0$ могут быть сведены к модели (6.3) логарифмированием:

$$\ln y = \ln f(\bar{a}, \bar{x}) + \eta \ln \varphi(\bar{b}, \bar{x}). \qquad (6.4)$$

Модели (6.3) и (6.4) часто используются для моделирования экономических процессов.

Остановимся на наиболее часто встречающихся на практике видах детерминированных основ процессов $f(\bar{a}, \bar{x})$. Начнем с рассмотрения моделей, у которых $\bar{x}$ имеет размерность, равную единице, т.е. когда моделируемая величина является функцией только одного известного параметра – времени *t*:

$$\bar{x} = t.$$

При этом модель (6.2) можно записать в виде

$$y = f(\bar{a}, t) + \eta. \qquad (6.5)$$

Наиболее простой моделью этого типа является модель с постоянной детерминированной основой

$$y = a + \eta, \qquad (6.6)$$

т.е. модель, детерминированная основа которой является константой.

$$f(\bar{a}, \bar{x}) = a. \qquad (6.7)$$

Модель с постоянной детерминированной основой (6.6) является весьма распространенной, так как она описывает стационарные случайные процессы, широко представленные на практике. Наблюдения за процессом с этой моделью в различные моменты времени представляют собой случайную выборку из некоторого распределения с математическим ожиданием, не зависящим от времени.

Детерминированная основа *линейной* модели

$$y = a_0 + a_1 t + \eta \qquad (6.8)$$

может быть представлена в виде

$$f(\bar{a}, t) = a_0 + a_1 t. \qquad (6.9)$$

Этой моделью описываются процессы, развивающиеся с постоянной скоростью a_1.

Пример. Наиболее простым способом расчета упреждения при стрельбе по воздушной цели является расчет, при котором вычисление будущего положения цели ведется исходя из предположения о ее равномерном и прямолинейном движении за время полета снаряда в упрежденную точку. При этом координаты цели рассчитываются в соответствии с моделью:

$$x = x_0 + v_{0_x} t; \quad y = y_0 + v_{0_y} t; \quad z = z_0 + v_{0_z} t.$$

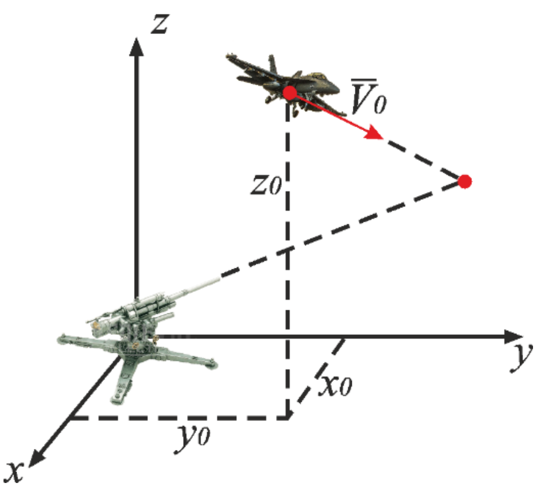


Рис. 6.2. Линейная модель

Квадратичная модель

$$y = a_0 + a_1 t + a_2 t^2 + \eta \qquad (6.10)$$

имеет детерминированную основу вида

$$f(\overline{a}, t) = a_0 + a_1 t + a_2 t^2. \qquad (6.11)$$

Эта модель учитывает кроме скорости изменения моделируемой величины еще и ее постоянное ускорение a_2. Такая модель могла бы быть применена в условиях предыдущего примера, если бы цель от момента выстрела до момента встречи со снарядом имела бы постоянное ускорение. При этом вычисление упреждения проводилось бы при гипотезе о том, что цель будет иметь постоянное ускорение за время полета снаряда в упрежденную точку.

Рассмотренные три вида детерминированных основ процессов (постоянная, линейная, квадратичная) являются распространенными частными случаями более общего вида детерминированной основы – *полиномиальной*:

$$f(\overline{a}, t) = \sum_{i=0}^{n} a_i t^i, \qquad (6.12)$$

которая в общем случае представляет собой полином n-й степени. Следовательно, модель с постоянной детерминированной основой есть полином нулевой степени, с линейной – полином первой степени и квадратичной – полином второй степени.

Степень полинома в некоторых случаях может быть выбрана на основании физического анализа процесса. Так, перемещение цели в рассмотренном примере с постоянным линейным ускорением может быть представлено полиномом второй степени, а изменение ее скорости – полиномом первой степени.

В тех случаях, когда моделируемый процесс изучен недостаточно, к выбору степени полинома следует подходить с большой осторожностью. Действительно, увеличивая степень полинома, можно с необходимой точностью описать изучаемый процесс на участке наблюдения.

В частности, при n наблюдениях можно построить полином Лагранжа $n - 1$ степени, который пройдет через все наблюдаемые точки. Однако реальные процессы искажены помехой. В связи с этим попытка построить моделирующий полином, проходящий через все наблюдаемые точки, приведет, как правило, к неточной модели.

Экспоненциальные модели

$$y = a_1 e^{-a_2 t} + \eta. \tag{6.13}$$

Эти модели описывают процессы, в которых *скорость изменения моделируемого параметра пропорциональна величине этого параметра*.

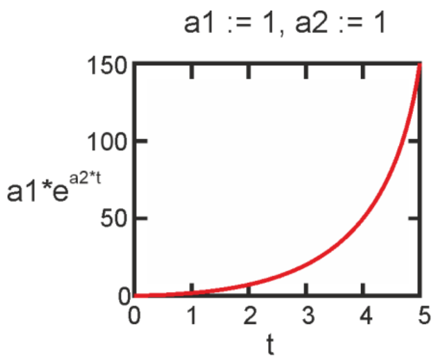


Рис. 6.3. График экспоненциального процесса

Экспоненциальные модели являются частным случаем более общего класса логистических моделей, детерминированная основа которых имеет вид

$$f(\bar{a}, t) = \frac{a_1}{1 + a_2 e^{-a_3 t}} \tag{6.14}$$

и представляют собой S-образную кривую с асимптотой a_1 при $t \longrightarrow \infty$.

При выполнении условия $a_2 e^{-a_3 t} >> 1$, которое может иметь место при малых t и $a_2 > 1$ получаем, как частный случай, экспоненциальную модель.

При построении модели $f(\bar{a}, t)$ необходимо найти вектор $\bar{a}$. Компоненты вектора $\bar{a}$ находят, исходя из требования минимума разницы между экспериментальными и расчетными (по модели) значениями.

$$a1 := 1, a2 := 1, a3 := 1$$

$$f(t) := \frac{a1}{1 + a1 * e^{-a3 * t}}$$

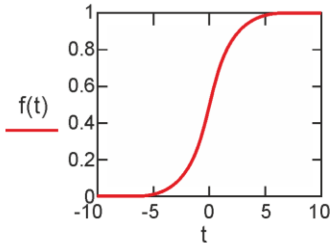


Рис. 6.4. Логистическая модель процесса

Поиск компонент вектора $\bar{a}$ называется регрессией.

На следующих рисунках представлены примеры одномерной регрессии.

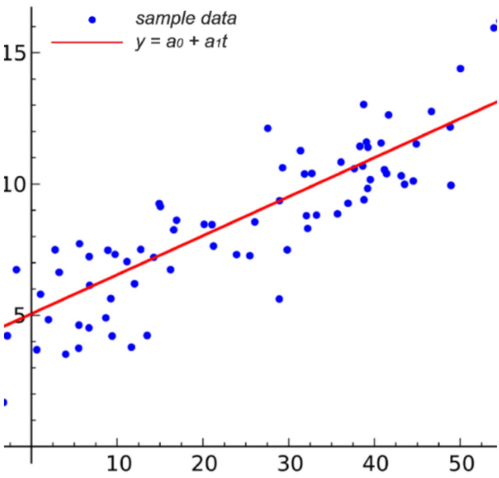


Рис. 6.5. Линейная регрессия

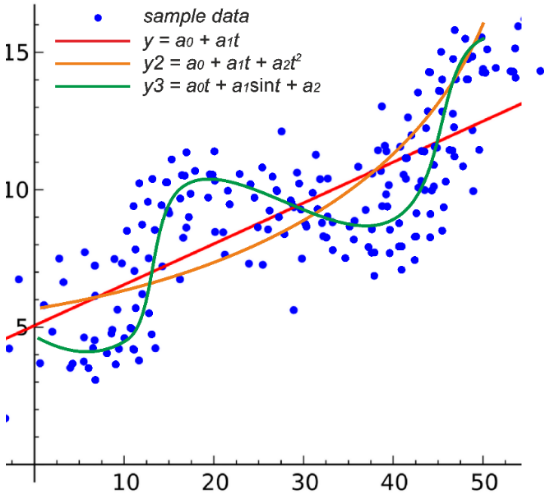


Рис. 6.6. Полиномиальная зависимость первой и второй степени и сложная зависимость

На первом рисунке в качестве аналитической зависимости принята полиномиальная зависимость первой степени $y = a_0 + a_1t$.

На втором рисунке – полиномиальная зависимость второй степени $y_2 = a_0 + a_1t + a_2t^2$ и сложная зависимость $y_3 = a_0t + a_1\sin t + a_2$.

Следует отметить, что все три зависимости являются зависимостями как от t, так и от вектора некоторых переменных $\bar{a}$ – весовых коэффициентов, причем от вектора весовых коэффициентов – *линейно*! Поэтому все эти зависимости являются *линейными регрессиями*.

Весовые коэффициенты $\bar{a}$ отражают влияние соответствующих факторов или комбинаций факторов на значение функции y.

В общем виде задача ставится так:

$$\text{\textit{Дано}:}\ \begin{pmatrix} t_1 & y_1 \\ t_2 & y_2 \\ \dots & \dots \\ t_M & y_M \end{pmatrix}.$$

Для $y_{\text{расч}} = \sum_{i=0}^{n} a_i t^i$ найти $\overline{a}^* = (a_0^*,\, a_1^* \ldots a_n^*)$ такой, что

$$y_{\text{расч\_}j} = \sum_{i=0}^{n} a_i^* \cdot t_j^{i-1} = a_0^* + a_1^* t_j + a_2^* t_j^2 + \cdots + a_n^* t_j^n,$$

$$\Delta_j = y_{\text{расч\_}j} - y_j,$$

$$\sum_{j=1}^{M} \Delta_j^2 \to \min.$$

Решение.

Запишем матрицу: $A = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^n \\ 1 & t_2 & t_2^2 & \cdots & t_2^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & t_M & t_M^2 & \cdots & t_M^n \end{pmatrix}.$

Тогда

$$\overline{a}^* = (A^T A)^{-1} \cdot (A^T Y).$$

Предполагается, что $(A^TA)^{-1}$ невырожденая и не плохообусловленая.

Если измерений M больше или равно количеству факторов n, то det $(A^TA) \neq 0$ и решение единственно.

Если $M < n$, то $\det(A^TA) = 0$ и решений бесконечное количество.

6.4. МОДЕЛИ, ОПИСЫВАЕМЫЕ ДИФФЕРЕНЦИАЛЬНЫМИ УРАВНЕНИЯМИ

В ряде случаев не представляется возможным выразить моделируемую величину через другие величины.

Модель исследуемого процесса может представлять собой дифференциальное уравнение или систему дифференциальных уравнений (линейных, нелинейных, в частных производных и т.д.). Зачастую задача моделирования характеристик процессов, описываемых такими моделями, не может быть решена аналитически.

Модели процессов, основанные на дифференциальных уравнениях, естественно, не могут быть составлены только по результатам наблюдений за моделируемым процессом в прошлом. Для составления подобных моделей необходим глубокий физический анализ явлений, приводящих к изменению моделируемого процесса, и проведение специальных научных исследований.

Дифференциальные уравнения представляют собой модели несравненно более богатые по содержанию, чем, например, модели, описываемые алгебраическими и трансцендентными уравнениями. Достоинством последних является простота и сравнительное удобство в работе. Однако дифференциальные уравнения позволяют более глубоко изучить динамический процесс формирования моделируемой величины, рассмотреть влияние известных параметров модели на вопросы устойчивости моделируемого процесса, а также решать задачу выбора величин параметров процесса (если мы в состоянии это сделать), чтобы моделируемый процесс в будущем был смоделирован адекватно.

Примеры.

Выше рассматривалась алгебраическая модель равноускоренного движения

$$f(\bar{a}, t) = a_0 + a_1 t + a_2 t^2.$$

Здесь переменные a_0, a_1 и a_2 рассматривались как независимые, а по существу процесса

$$a_1 t = \frac{da(t)}{dt}$$

или

$$a_2 = \frac{da_1(t)}{dt} = \frac{d^2 a(t)}{dt^2}.$$

Пример. Пусть исходное число боевых единиц у одной стороны n_0, текущее n и каждая из них с вероятностью P при каждом выстреле поражает боевую единицу противника, имея скорострельность λ.

Все единицы одной стороны одинаковы. У противника соответственно: $n_{0\text{пр}}$, $n_{\text{пр}}$, $P_{\text{пр}}$, $\lambda_{\text{пр}}$. Пусть бой полностью упорядочен, т.е. все боевые единицы разведаны и огонь ведется только по непораженным единицам. Тогда можно предположить, что скорость потерь одной стороны пропорциональна произведению числа боевых единиц другой на скорострельность и вероятность поражения:

$$\frac{dn(t)}{dt} = -n_{\text{пр}}(t) P_{\text{пр}} \lambda_{\text{пр}},$$

$$\frac{dn_{\text{пр}}(t)}{dt} = -n(t) P \lambda.$$

Решение этих дифференциальных уравнений позволяет определить количество уцелевших единиц каждой стороны в любой момент времени. Следует заметить, что в уравнениях динамики боя можно учитывать различные факторы, относящиеся к организации боевых действий, таких, как

– ввод резервов;
– упреждающий удар одной из сторон;
– истощение боезапаса;
– присутствие разнотипных боевых единиц и т.д.

Можно показать, что параметрами, определяющими победу той или иной стороны при использовании приведенных уравнений, являются величины:

$$\Phi = P\lambda n_0^2 \quad \text{и} \quad \Phi_{\text{пр}} = P_{\text{пр}} \lambda_{\text{пр}} n_{0\_\text{пр}}^2,$$

т.е. побеждает та сторона, у которой больше Ф. Другими словами, больше произведение вероятности поражения боевой единицы на скорострельность и квадрат исходного числа боевых единиц. Этот результат не противоречит и здравому смыслу. Однако, как следует из выражений для Ф, наибольшее влияние оказывает на победу начальное число боевых единиц, так как оно входит в квадрате.

Подчеркнем еще раз, что полученные результаты являются верными «в среднем», т.е. результат $\Phi > \Phi_{\text{пр}}$ еще не означает, что в каком-то конкретном бою при тех же величинах невозможна победа второй стороны, потому, что здесь участвует *вероятность*.

Пример. Уравнения движения центра масс артиллерийского снаряда в воздухе, из курса «Внешняя баллистика», могут быть записаны в следующем виде

$$\frac{d^2x}{dt^2} = -cH(y)G(\mathrm{v})\frac{dx}{dt}; \qquad \frac{d^2y}{dt^2} = -cH(y)G(\mathrm{v})\frac{dy}{dt} - g,$$

где x, y – координаты центра масс снаряда в прямоугольной системе координат, связанной с плоской невращающейся Землей; c – баллистический коэффициент; $H(y)$ – функция плотности воздуха при нормальных метеорологических условиях, которая зависит от высоты; $G(\mathrm{v})$ – функция сопротивления воздуха; g – ускорение силы тяжести, принимаемое постоянным по величине и направлению.

Интегрируя СДУ, получим в качестве решения (при заданных начальных условиях) две функции $x(t)$ и $y(t)$.

Хорошо зарекомендовал себя рассматриваемый математический аппарат в форме системы дифференциальных уравнений (СДУ) Колмогорова для исследования динамики протекания эпидемий.

Из области эпидемиологии

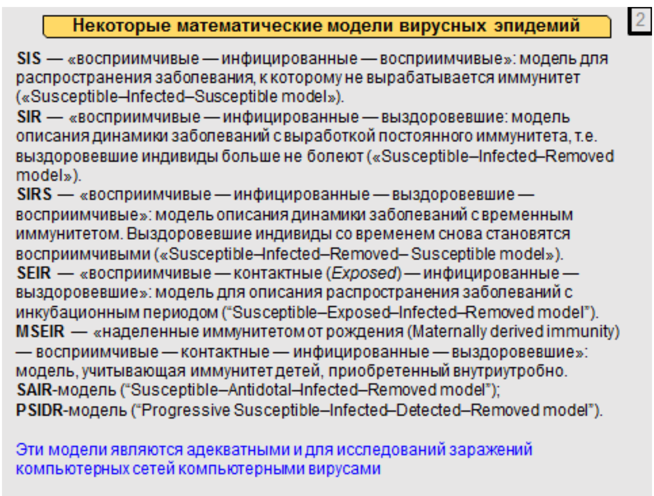


Рис. 6.7. Вирусные эпидемии

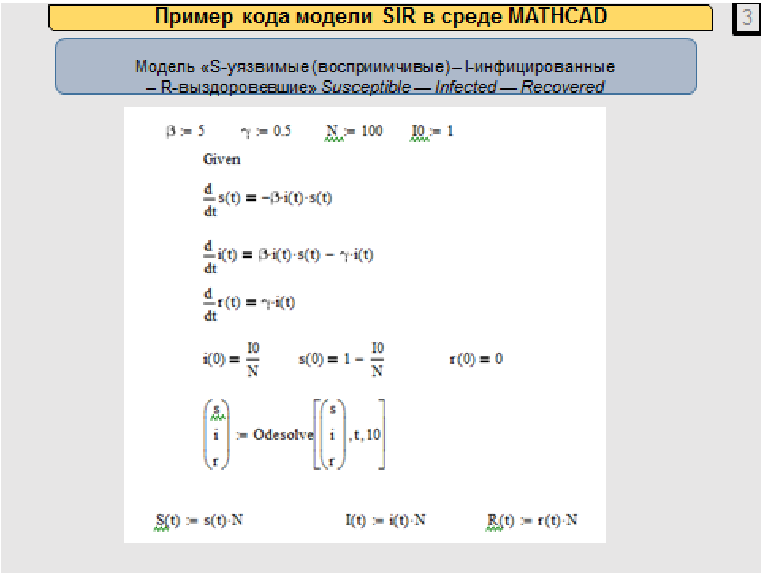


Рис. 6.8. Пример кода модели SIR в среде MATHCAD

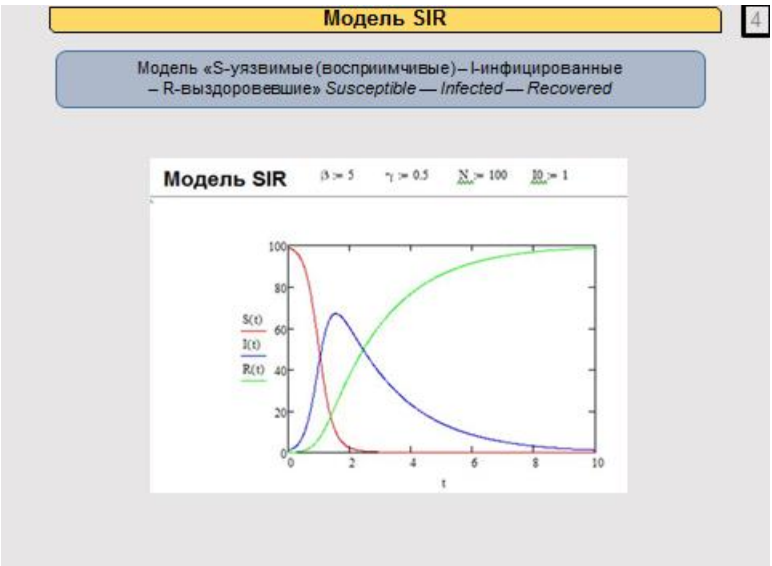


Рис. 6.9. Модель SIR

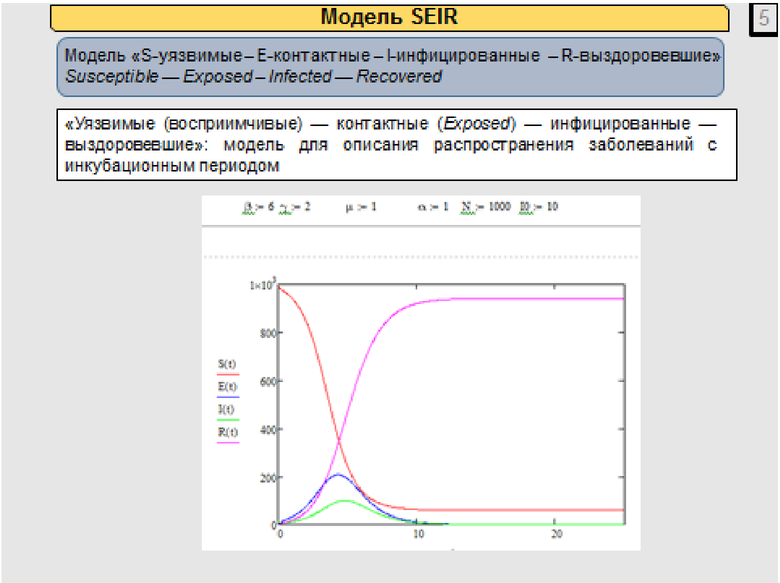


Рис. 6.10. Модель SEIR

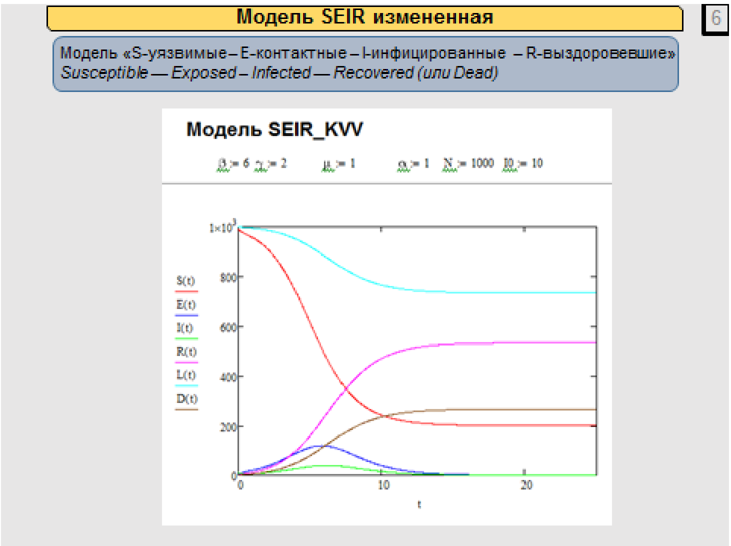


Рис. 6.11. Модель SEIR с изменениями

Отдельно следует упомянуть теорию фильтрации и моделирования нестационарных случайных процессов, разработанную Р. Калманом, которая легла в основу *моделирующего фильтра,* обеспечивающего моделирование многомерного случайного процесса $x(t)$:

$$\frac{dx(t)}{dt} = F(t)x(t),$$

$$z(t) = H(t)x(t) + v(t).$$

Здесь $F(t)$ – матрица, характеризующая динамические свойства процесса; $H(t)$ – матрица, отражающая наблюдение динамического процесса $x(t)$; $z(t)$ – данные наблюдений за процессом $x(t)$; $v(t)$ – помеха при наблюдении.

Оптимальная оценка моделируемого процесса $\hat{x}(t)$ в текущий момент времени, определяемая из выражения

$$\frac{d\hat{x}(t)}{dt} = F(t)\hat{x}(t) + K(t)\left[z(t) - H(t)\hat{x}(t)\right],$$

где $K(t)$ – матричный коэффициент усиления, определяемый в зависимости от динамических свойств системы, ограничений, накладываемых на наблюдения за моделируемым процессом, и характеристик помех.

6.5. ВЕРОЯТНОСТНЫЕ МОДЕЛИ

В предыдущих вопросах были рассмотрены некоторые модели, описываемые алгебраическими, трансцендентными и дифференциальными уравнениями, которые строятся на основании физического анализа существа моделируемого процесса и представляют собой аналитические выражения, связывающие моделируемую величину с рядом других величин.

Однако, часто можно судить только о вероятностях того или иного события, или делать умозаключения на основе знания каких-либо вероятностей.

Рассмотрим часто встречающийся математический аппарат СДС – стохастические динамические системы.

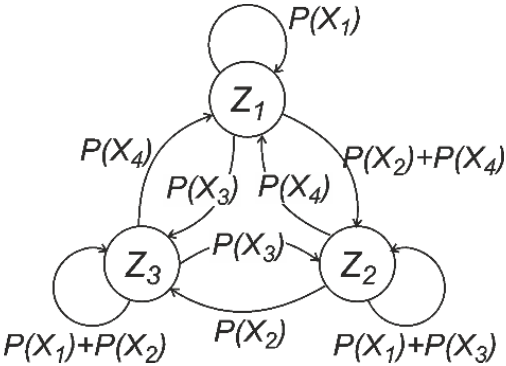


Рис. 6.12. Модель СДС с тремя возможными состояниями

Характеристиками СДС являются:

– вектор вероятностей состояний

$$P_z(t) = (P_{z1}(t), P_{z2}(t), ..., P_{zn}(t)), \sum_i P_{zi}(t) = 1;$$

– матрица переходных вероятностей $||P_{ij}||$, где $p_{ij} = p_{ij}(t, \tau, \Theta)$ – условная вероятность того, что в момент времени $t + \tau$ окажется в состоянии Z_j, если в момент времени t она находилась в состоянии Z_i.

В зависимости от вида $p_{ij}(t, \tau, \Theta)$ существуют различные модели СДС.

Марковские процессы

Процесс называется *марковским*, если переходная вероятность зависит только от i, j, t и τ, т.е. процесс без последействия (не зависит от того в каких состояниях система была раньше)

$$p_{ij} = p_{ij}(t, \tau).$$

Марковские процессы различают:

- по времени:

– непрерывные $t \in R$;

– дискретные $t \in \{t_1, t_2, ...\}$, $\tau = \text{const}$;

- по изменению переходной вероятности во времени:

– однородные

$$p_{ij}(t) = p_{ij} = \text{const};$$

– неоднородные

$$p_{ij}(t) \neq \text{const}.$$

Дискретные МП называют Марковскими цепями.

Марковские процессы (МП) задаются:

1. Множеством состояний $Z = \{z_1, z_2, ..., z_l\}$;
2. Определением начального состояния Z_0;
3. Заданием матрицы $\Lambda(t)$ интенсивности переходов из одного состояния в другое, элементами которой являются величины λ_{ij}:

$$\lambda_{ij}(t) = \lim_{\tau \to 0} \frac{P_{ij}(t, \tau)}{\tau}.$$

Для однородных МП $\lambda_{ij}(t) = \text{const}$ и $\Lambda(t) = \Lambda$.

С использованием интенсивностей переходов строится основное для МП математическое соотношение – система дифференциальных уравнений Колмогорова:

$$\frac{dP_j(t)}{dt} = -P_j(t)\sum_{i \neq j} \lambda_{ji} + \sum_{i \neq j} P_i(t)\lambda_{ij}, \quad i, j = 0, 1, 2, ..., L.$$

Или, в развернутом виде

$$\begin{cases} \dfrac{dP_1(t)}{dt} = -P_1(t)\sum\limits_{i \neq 1} \lambda_{1i} + \sum\limits_{i \neq 1} P_1(t)\lambda_{i1}; \\ \dfrac{dP_2(t)}{dt} = -P_2(t)\sum\limits_{i \neq 2} \lambda_{2i} + \sum\limits_{i \neq 2} P_i(t)\lambda_{i2}; \\ \cdots \\ \dfrac{dP_j(t)}{dt} = -P_j(t)\sum\limits_{i \neq j} \lambda_{ji} + \sum\limits_{i \neq j} P_i(t)\lambda_{ij}; \\ \sum_j P_j(t) = 1. \end{cases} \tag{6.16}$$

Решением СДУ при начальных условиях $P_i(0)=1$, $\forall\ j \neq i$: $P_j(0)=0$ является совокупность зависимостей $P_j(t)$, j = 0, 1, 2, ..., J.

При этом, надо помнить что $\sum_j P_j(t)=1$.

СДУ Колмогорова легко строится, если МП задан графом.

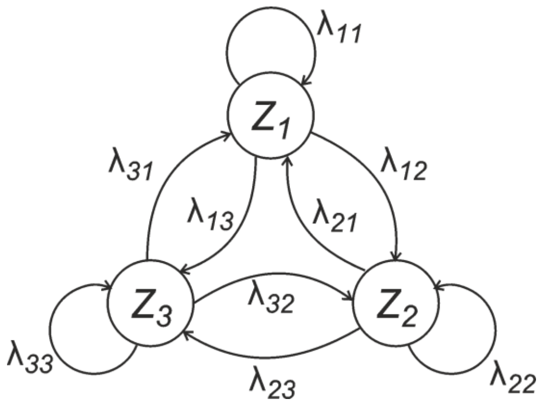


Рис. 6.13. Иллюстрация примера марковского процесса, заданного графом

Для каждого состояния строится ДУ по правилу:

– слева пишется производная j-го состояния;

– справа столько слагаемых, сколько стрелок связано с данным (j-м) состоянием, причем:

– если стрелка направлена из состояния, то слагаемое берется с минусом, а если в состояние, то с плюсом;

– каждое слагаемое представляет собой произведение интенсивности перехода по стрелке и вероятности того состояния, из которого выходит стрелка.

Для *устоявшихся режимов* левые части уравнений (6.16), кроме последнего, заменяют нулями и *получают систему линейных алгебраических уравнений*

$$
\begin{cases}
-P_1(t)\sum\limits_{i\neq 1}\lambda_{1i}+\sum\limits_{i\neq 1}P_1(t)\lambda_{i1}=0;\\
-P_2(t)\sum\limits_{i\neq 2}\lambda_{2i}+\sum\limits_{i\neq 2}P_i(t)\lambda_{i2}=0;\\
\cdots\\
-P_j(t)\sum\limits_{i\neq j}\lambda_{ji}+\sum\limits_{i\neq j}P_i(t)\lambda_{ij}=0;\\
\sum_j P_j(t)=1.
\end{cases} \tag{6.17}
$$

Марковские цепи

Дискретные марковские процессы называются *марковские цепи*. Они задаются:

1. Множеством состояний $Z = \{z_1, z_2, \ldots, z_I\}$.
2. Определением начального состояния $Z(0)$.
3. Заданием матрицы переходных вероятностей

$$
P=\begin{pmatrix}
P_{11} & P_{12} & \cdots & P_{1j} & P_{1n}\\
P_{21} & P_{22} & \cdots & P_{2j} & P_{2n}\\
\cdots & \cdots & \cdots & \cdots & \cdots\\
P_{i1} & P_{i1} & \cdots & P_{ij} & P_{in}\\
P_{n1} & P_{n2} & \cdots & P_{nj} & P_{nn}
\end{pmatrix}.
$$

Тогда состояние системы через n тактов рассчитывается по формуле

$$
P(n)=P(n-1)\cdot P=P(0)\cdot P^n. \tag{6.18}
$$

Эргодические цепи

Если существует некоторое j-е состояние, такое, что $P_{jj} = 1$ (или $\forall i\ P_{ji} = 0$), то это состояние называется *поглощающим*.

Цепи, содержащие поглощающие состояния, называются *цепями с поглощением*.

Если цепь не содержит ни одного поглощающего состояния, она называется *эргодической цепью*.

Для эргодических цепей существуют *асимптотические значения вероятностей состояний*, т.е. это значения при $n \to \infty$, которые уже не зависят от n. Это стационарные вероятности. Обозначим их как $\pi = \{\pi_1, \pi_2, ..., \pi_n\}$.

Поскольку они стационарные, значит они равны и при n и при $n + 1$, т.е.

$$\pi \cdot P = \pi, \ \sum \pi_i = 1.$$

Следовательно

$$\begin{cases} \pi_1 = \sum_{j=1}^{n} \pi_j \cdot P_{j1}; \\ \pi_2 = \sum_{j=1}^{n} \pi_j \cdot P_{j2}; \\ \dots \\ \pi_i = \sum_{j=1}^{n} \pi_j \cdot P_{ji}; \\ \dots \\ \pi_{n-1} = \sum_{j=1}^{n} \pi_j \cdot P_{jn-1}; \\ \sum \pi_i = 1. \end{cases} \qquad (6.19)$$

Пример.

Дано: $P = \begin{pmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{pmatrix}.$

Решение: $\begin{cases} \pi_1 = 0{,}7 \cdot \pi_1 + 0{,}4 \cdot \pi_2 \\ \pi_1 + \pi_2 = 1 \end{cases} \Rightarrow \pi_1 = 0{,}571, \quad \pi_2 = 0{,}429.$

Справка. Решение в матричной форме.

Итак,

$\pi \cdot P = \pi$

$\pi = \{\pi_1, \pi_2, ..., \pi_n\}$ – вектор-строка.

$\pi \cdot P - \pi = (0,\ 0,\ ...,\ 0)$,

$\pi \cdot (P - E) = (0,\ 0,\ ...,\ 0)$, E = единичная матрица.

Транспонируем равенство, используя

$$(ab)^T = b^T a^T,$$

$$(P - E)^T \pi^T = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}.$$

Но это однородное линейное уравнение относительно π

$$A'X = B',$$

где $A' = (P - E)^T$, $X = \pi^T$, $B' = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}$.

Поскольку уравнение однородное, оно имеет бесконечное множество решений. В нем не хватает коэффициента, задающего масштаб. Воспользуемся тем, что $\Sigma\pi_i = 1$. Заменим последнее уравнение на $\Sigma\pi_i = 1$. Это равносильно замене матрицы A' на A, отличающуюся от A' тем, что в ней вся последняя строка состоит из единиц, и вектора B' на $B = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}$. $AX = B$, откуда $A^{-1}AX = A^{-1}B$, $X = \pi^T = A^{-1}B$.

Пример. Рассчитаем асимптотические вероятности сразу

$$ORIGIN := 1$$

$$P := \begin{pmatrix} 0.2 & 0.1 & 0.3 & 0.4 \\ 0.6 & 0 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0.4 & 0 \\ 0.1 & 0.3 & 0.2 & 0.4 \end{pmatrix} \quad E := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad A := (P - E)^T = \begin{pmatrix} -0.8 & 0.6 & 0.2 & 0.1 \\ 0.1 & -1 & 0.4 & 0.3 \\ 0.3 & 0.3 & -0.6 & 0.2 \\ 0.4 & 0.1 & 0 & -0.6 \end{pmatrix}$$

$$A_{4,1} := 1 \quad A_{4,2} := 1 \quad A_{4,3} := 1 \quad A_{4,4} := 1 \quad B := \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$A = \begin{pmatrix} -0.8 & 0.6 & 0.2 & 0.1 \\ 0.1 & -1 & 0.4 & 0.3 \\ 0.3 & 0.3 & -0.6 & 0.2 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad \pi := A^{-1} \cdot B \quad \pi = \begin{pmatrix} 0.264 \\ 0.214 \\ 0.31 \\ 0.212 \end{pmatrix}$$

или, поэтапно, задав вектор вероятностей состояний в начальный момент

$$S := \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$S0 := S^T \cdot P^0 = (0 \quad 1 \quad 0 \quad 0)$$

$$S1 := S^T \cdot P^1 = (0.6 \quad 0 \quad 0.3 \quad 0.1)$$

$$S2 := S^T \cdot P^2 = (0.19 \quad 0.21 \quad 0.32 \quad 0.28)$$

$$S3 := S^T \cdot P^3 = (0.256 \quad 0.231 \quad 0.304 \quad 0.209)$$

$$S4 := S^T \cdot P^4 = (0.272 \quad 0.21 \quad 0.31 \quad 0.209)$$

$$S5 := S^T \cdot P^5 = (0.263 \quad 0.214 \quad 0.31 \quad 0.213)$$

$$S6 := S^T \cdot P^6 = (0.264 \quad 0.214 \quad 0.31 \quad 0.212) .$$

Процессы «гибели и размножения»

Процессы гибели и размножения (ПГиР) есть частный вид МП, характеризующийся тем, что состояния соединены «в цепь», т.е. каждое состояние связано интенсивностями только с двумя соседними состояниями.

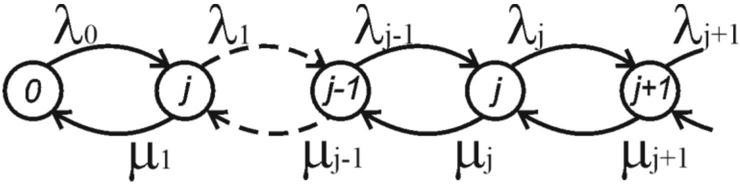


Рис. 6.14. Граф процесса «гибели и размножения»

ПГиР вводится следующими 6-ю допущениями:

1. $P_{j,\ j+1}(\Delta t) = \lambda_j \Delta t + O(\Delta t),\ j \geq 0$;
2. $P_{j,\ j-1}(\Delta t) = \mu_j \Delta t + O(\Delta t),\ j \geq 1$;
3. $P_{j,\ j}(\Delta t) = 1 - (\lambda_j + \mu_j)\,\Delta t + O(\Delta t)$;
4. $\mu_0 = 0$; $\lambda_0 > 0$; λ_j, $\mu_j > 0$, $j > 0$; при $n < \infty$ $\lambda_n = 0$;
5. $P_{ji}(\Delta t) = 0$, $|\,j - i\,| > 1$;
6. λ_j, μ_j не зависят от времени.

Общая СДУ, описывающая динамику ПГиР:

$$\begin{cases} \dfrac{dP_0(t)}{dt} = -\lambda_0 P_0(t) + \mu_1 P_1(t); \\ \dfrac{dP_j(t)}{dt} = -(\lambda_j + \mu_j)P_j(t) + \mu_{j+1} P_{i+1}(t) + \lambda_{j-1} P_{j-1}(t), \ \ j \geq 1. \end{cases}$$

Системы массового обслуживания

СМО – это математические модели систем, предназначенных для обслуживания потоков заявок (требований), поступающих через случайные промежутки времени, причем время обслуживания одной заявки также случайно.

Примеры: ТЛФ-станция, возникновение неисправности в системе, попытки НСД и т.д.

Классификация СМО:

– *по характеру источников требований*: с конечным числом заявок; с бесконечным числом заявок;

– *по возможности ожидания*: с ожиданием (очередь конечная или бесконечная); с отказами;

– *по числу обслуживающих приборов* (канал обслуживания – устройство, способное в течение некоторого времени обслужить одно (и только одно) требование): одноканальные, многоканальные;

– *по числу фаз обслуживания*: однофазные, многофазные;

– *по правилу формирования очереди*: с общей очередью, с несколькими очередями;

– *по дисциплине обслуживания*: бесприоритетные, приоритетные.

Приоритеты:

– абсолютный – при его поступлении прерывается выполнение требования низшего приоритета;

– относительный – приоритетное требование занимает место в очереди и не происходит прерывания процесса обслуживания.

Виды абсолютного приоритета:

– с потерями;

– с дообслуживанием;

– с обслуживанием заново.

Динамика функционирования N-канальной СМО с отказами может быть описана системой ДУ.

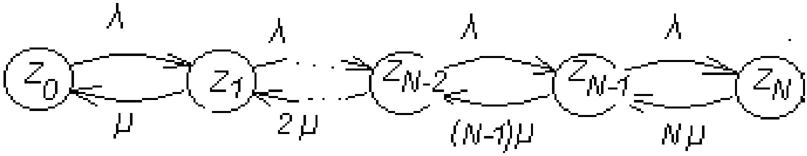


Рис. 6.15. Динамика функционирования *N*-канальной СМО с отказами

$$\begin{cases} \dfrac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t); \\ \dots \\ \dfrac{dP_i(t)}{dt} = -(\lambda + i\mu)P_i(t) + (i+1)\mu P_{i+1}(t) + \lambda P_{i-1}(t),\ i = 1, 2, ..., N-1; \\ \dots \\ \dfrac{dP_N(t)}{dt} = -N\mu P_N(t) + \lambda P_{N-1}(t), \end{cases}$$

где $P_i(t)$ – вероятность того, что в произвольный момент времени $t \geq 0$ i каналов из N будут заняты; λ – плотность (интенсивность) потока заявок; μ – величина, обратная среднему времени обслуживания, интенсивность обслуживания, среднее число заявок, обслуженных одним каналом за единицу времени.

Систему интегрируют при начальных условиях $P_0(0) = 1$, $P_{i \neq 0}(0) = 0$.

6.6. ОСНОВЫ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

При моделировании стохастических систем желательно получать модель в виде аналитических выражений.

Тогда моделирование будет заключаться в подставлении исходных данных, расчета и получения исчерпывающей информации в виде законов распределения или числовых характеристик.

Однако, такое возможно далеко не всегда.

В этом случае *надо моделировать сущность процесса* функционирования системы. Такое моделирование называется *имитационным*, а конкретную имитацию реализации процесса (функционирования системы) – *имитационным экспериментом*, или *реализацией* процесса.

В ИМ операторы G, H задаются алгоритмически.

Основными элементами имитационной модели являются *активности*, *события* и *процессы*.

Активность – это элемент имитационной модели, отражающий процесс выполнения какой-либо работы, действия, это динамический объект. Активность характеризуется *временем реализации* и *затраченными ресурсами* (или соответствующими распределениями вероятностей).

Событие – элемент имитационной модели, отражающий мгновенное изменение какого-либо объекта модели, т.е. факт совершения чего-либо, например, окончания какой-либо активности.

Процесс – логически связанная последовательность активностей.

Процесс может быть активностью в модели более высокого иерархического ранга.

Такая организация элементов позволяет строить *сложные иерархические структуры модели*. Причем, возможны независимые отдельные работы по созданию и отладке элементов модели, испытание их на всем множестве возможных значений влияющих факторов. Затем объединение в блоки и их отладку.

Все это в целом позволяет более наглядно представлять модель, более тщательно ее отлаживать, более точно соответствует реальному объекту, системе.

***Важно*!** В ИМ, как правило, приходится моделировать процессы, идущие в реальной системе параллельно, т.е. действия, выполняемые по какому-либо алгоритму, зависят от действий и событий других алгоритмов. Поэтому в ИМ *активности и события должны быть скоординированы по времени*, чтобы не нарушить причинно-следственные связи.

Существуют следующие механизмы реализации системного времени модели (модельного времени):

– непрерывное время;

– задание времени с помощью постоянных интервалов;

– задание времени с помощью переменных интервалов (по изменению состояний системы).

В последнем случае различают:

– событийно-ориентированные модели;

– процессо-ориентированные модели;

– объектно-ориентированные модели и их частный случай – агентно-ориентированные.

Объектно-ориентированное моделирование

Объектно-ориентированная имитационная модель обладает следующими характеристиками:

– модель представляет собой совокупность объектов;

– объекты включают данные и операции над ними;

– объекты представляют собой модель «актора», который может выполнить работу, изменить свое состояние и взаимодействовать с другими объектами;

– объекты являются основой современных языков программирования;

– объекты являются естественным способом описания системы.

Агентно-ориентированное моделирование

Выделяют еще одну разновидность систем моделирования, еще одну парадигму – ***агентно-ориентированную***, для которой характерны следующие свойства:

– агентный подход – это особый случай объектно-ориентированного подхода;

– поведение всей системы можно рассматривать как результат поведения большого числа объектов, называемых агентами;

– агенты являются автономными. Могут взаимодействовать друг с другом и преследуют свои цели.

Известные программы агентного ИМ Netlogo и AnyLogic.

Пример, модель в среде Netlogo.

Моделируется взаимоотношение патрулирующих объектов (зеленые) и объектов-нарушителей (красных). Изначально, патрулирующие объекты равномерно распределяются по району патрулирования. При появлении нарушителя ближайший патрулирующий объект направляется к нему, после их контакта нарушитель исчезает, патрулирующие объекты вновь распределяются равномерно по району патрулирования.

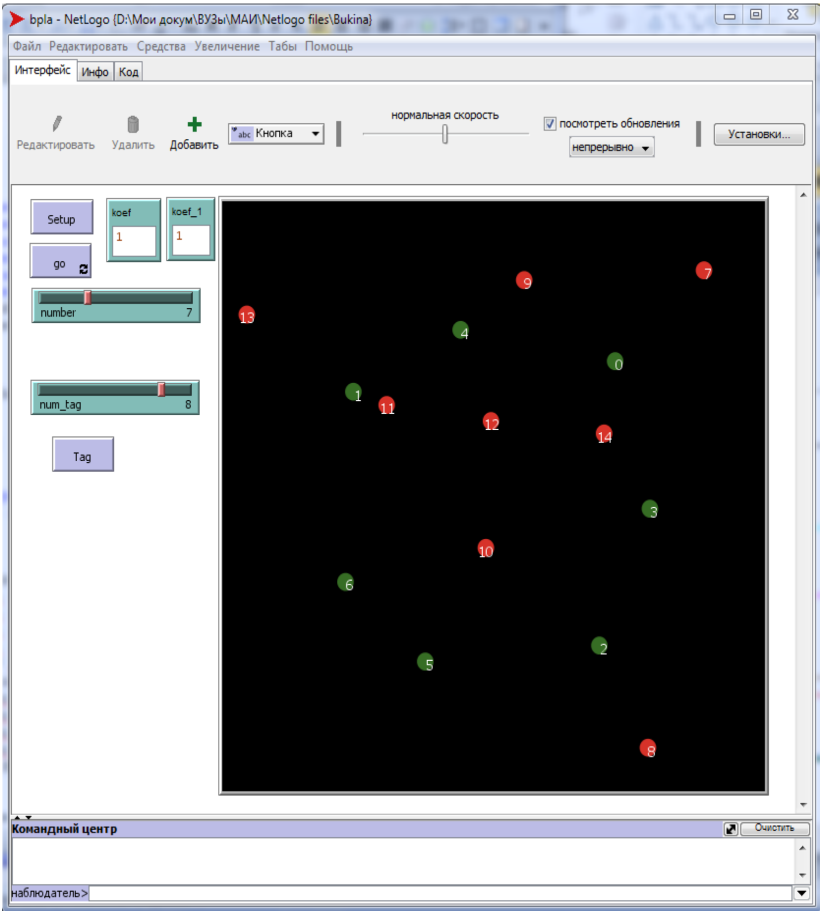


Рис. 6.16. Модель в среде Netlogo

Листинг

```
breed [bplas bpla] ; создание класса объектов – БПЛА
breed [targets target] ; создание класса объектов – цели
bplas-own [zan] ; задание классу БПЛА свойства «занят»
links-own [npr] ; задание классу «линии» свойства» «направление»
to setup ; начальная процедура
 ca ; очистка пространства
 create-bplas number ; задание количества БПЛА
 ask bplas ; для всех БПЛА:
 [setxy random-xcor random-ycor ; установить случайное положение
 set shape «circle» set color 53 ; установить красный цвет
 set size 1 ; установить размер
 set label who ; показать свой номер
 set zan false ; установить свойство «занят» – «ложь»
 create-links-from other turtles ; установить связи с каждым БПЛА
 ]
 ask links [hide-link ] ; скрыть с экрана эти связи
end

to go ; основная процедура программы
 ask-concurrent bplas ; заставить все БПЛА отодвигаться от краев полигона и друг от друга
 [if zan = false
 [
 let len_list [link-length] of my-in-links
 set len_list lput ((max-pxcor – xcor) / number) len_list
 set len_list lput ((xcor – min-pxcor) / number) len_list
 set len_list lput ((max-pycor – ycor) / number) len_list
 set len_list lput ((ycor – min-pycor) / number) len_list
 let v_list map PF len_list
```

```
 let napr_list [link-heading] of my-in-links
 set napr_list lput 270 napr_list
 set napr_list lput 90 napr_list
 set napr_list lput 180 napr_list
 set napr_list lput 0 napr_list
 let dv_list (map SENT napr_list v_list )
 foreach dv_list
 [x ->
 set heading item 0 x
 fd item 1 x
 ]
 ]
 ]
 tg
end
```

to-report ugol[ug]; переводит угол в интервале от -360 до 360 в интервал от 0 до 360

```
let u(0)
ifelse (ug >= 360)
 [set u (ug – 360)]
 [ifelse (ug < 0)
 [set u (ug + 360)]
 [set u (ug)]
 ]
 report u
end
```

to-report PF[r] ; процедура «потенциальная функция», устанавливает зависимость силы влияния БПЛА друг на друга в зависимости от расстояния

```
ifelse (r > 0)
[report 1 / (1 + r / koef)]
[report 1]
end

to-report SENT [a b]
report sentence a b
end
```

to tag ; процедура создания и размещения случайным образом целей

```
create-targets num_tag
ask targets
[setxy random-xcor random-ycor
set shape «circle» set color red
set size 1
set label who]
end
```

to Tg ; процедура, которая заставляет ближайший БПЛА «обслужить» цель. Инициируется целью.

```
ask targets
[
let Nm_tg who
ask min-one-of bplas with [zan = false] [distance myself]
[
let lb who
set zan true
let len distance myself ;
create-link-to myself
ask link lb Nm_tg
```

```
[
;ask links [hide-link ]
set npr link-heading
let np npr
ask bpla lb [set heading np]
]
fd len
ask target Nm_tg [die]
set zan false
]
]
end
```

Пожар леса

https://github.com/bedla/forest-wildfire-simulation

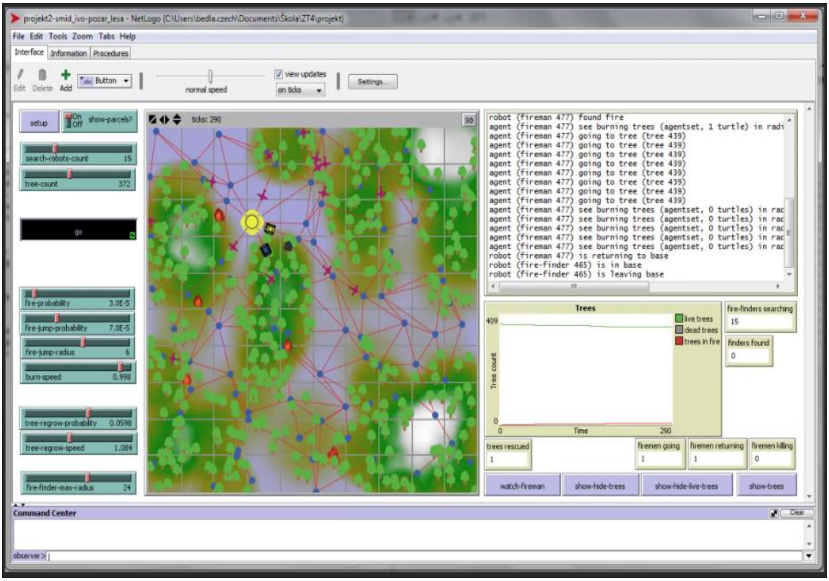


Рис. 6.17. Имитационная модель пожара леса в среде Netlogo

Атака вируса на сеть:
https://commons.wikimedia.org/wiki/File:NetLogo_604-VirusonaNetwork.jpg

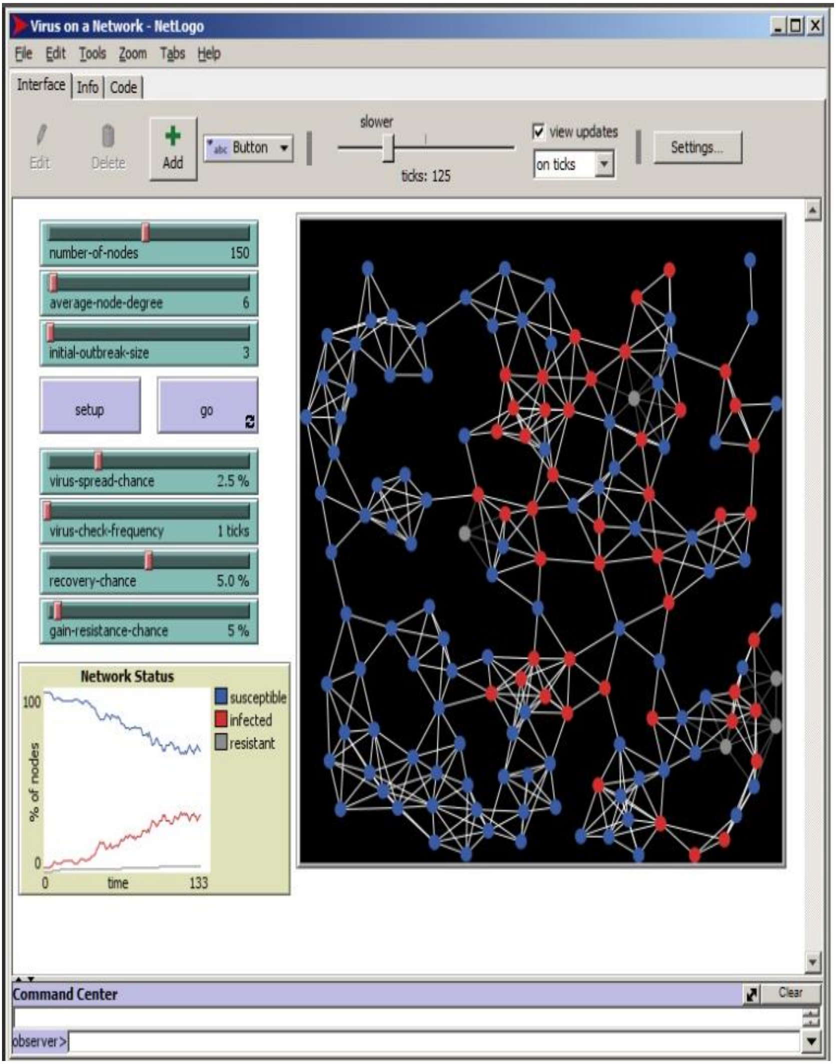


Рис. 6.18. Имитационная модель атаки компьютерного вируса на сеть в среде Netlogo

Более широкими возможностями обладает AnyLogic

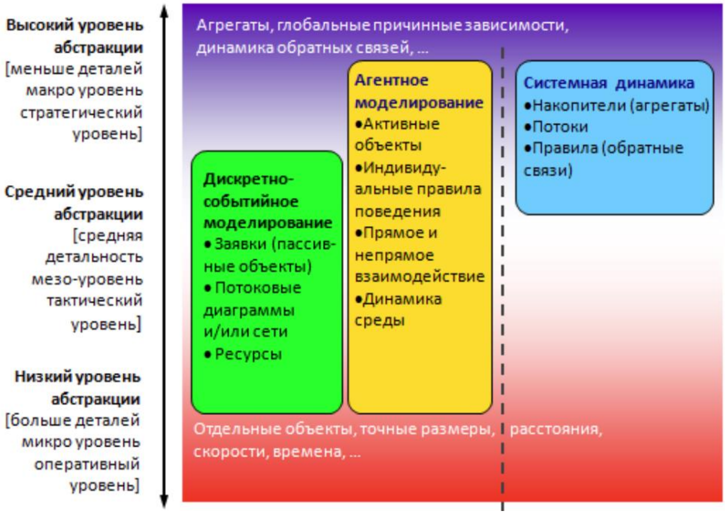


Рис. 6.19. Агрегаты AnyLogic

Примеры:

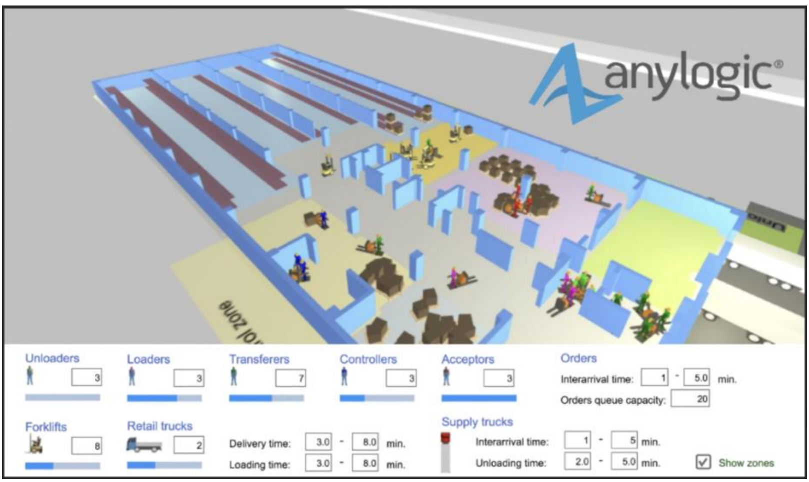


Рис. 6.20. Пример имитационной модели в среде AnyLogic

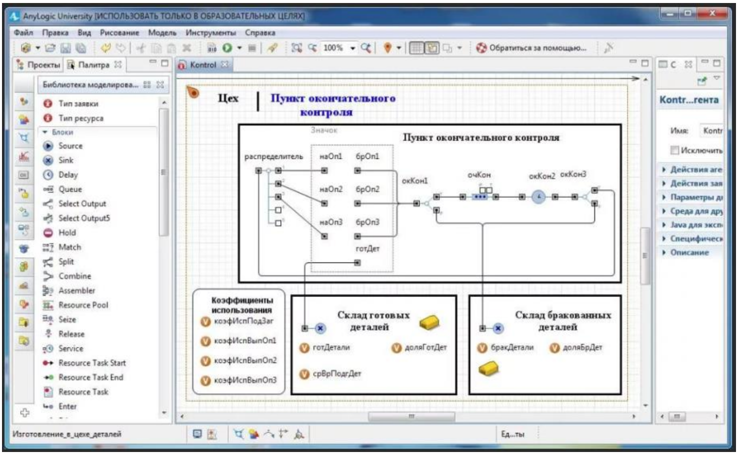


Рис. 6.21. Скрин AnyLogic

Еще некоторые программы имитационного моделирования.

Программное средство «Сетевой симулятор»

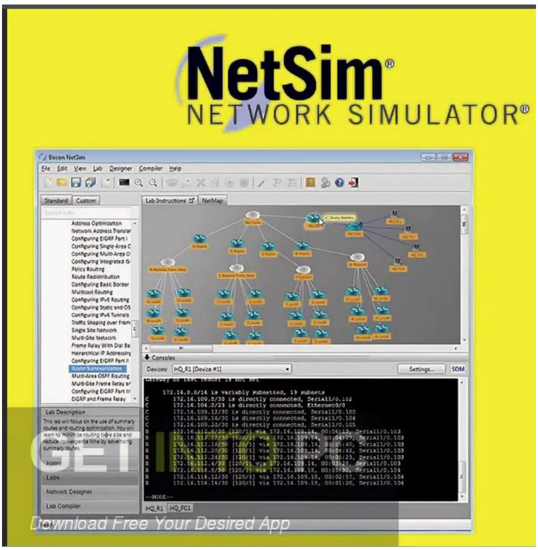


Рис. 6.22. Среда имитационного моделирования «Сетевой симулятор»

SimulationX (коммерческая): для разработки, моделирования, симулирования, анализа и виртуального тестирования сложных мехатронных систем. На единой платформе программа моделирует поведение и взаимодействие различных физических объектов механики (1D и 3D), приводной техники, электрических, гидравлических, пневматических и термодинамических систем, а также магнетизма и аналоговых и цифровых систем управления. Одними из основных приложений SimulationX являются исследования в области автомобильных приводов.

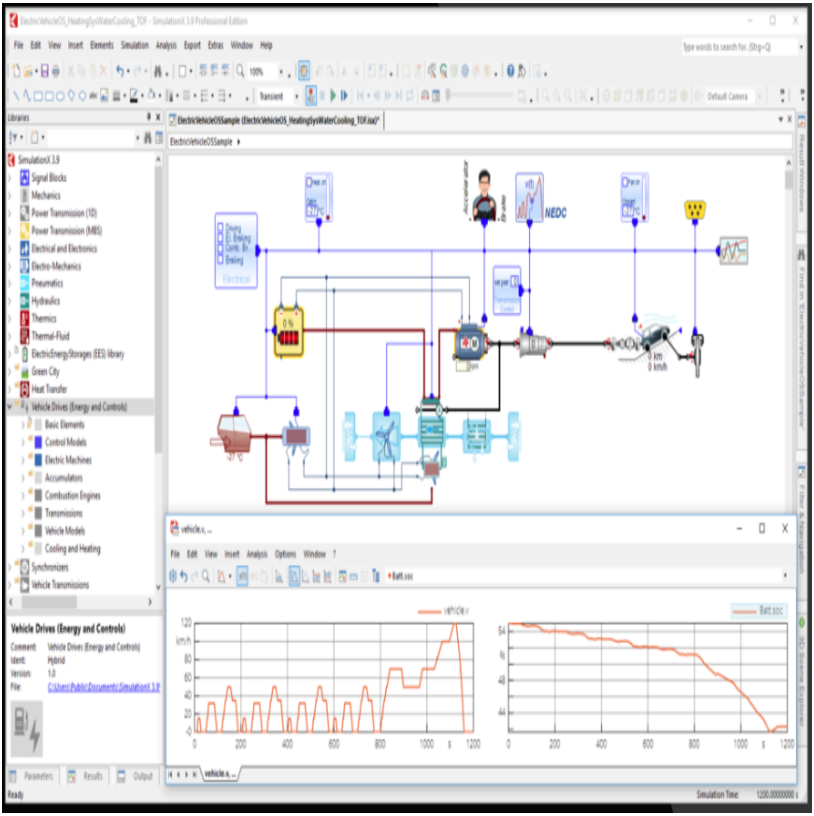


Рис. 6.23. Имитационная модель автомобильных приводов в среде SimulationX

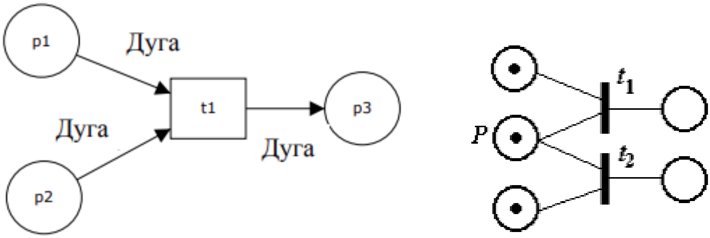


Рис. 6.24. Сети Петри

CPN Tools предлагает мощный класс сетей Петри для описания моделей. Согласно стандартной классификации такие сети называют *иерархическими временными раскрашенными сетями Петри*.

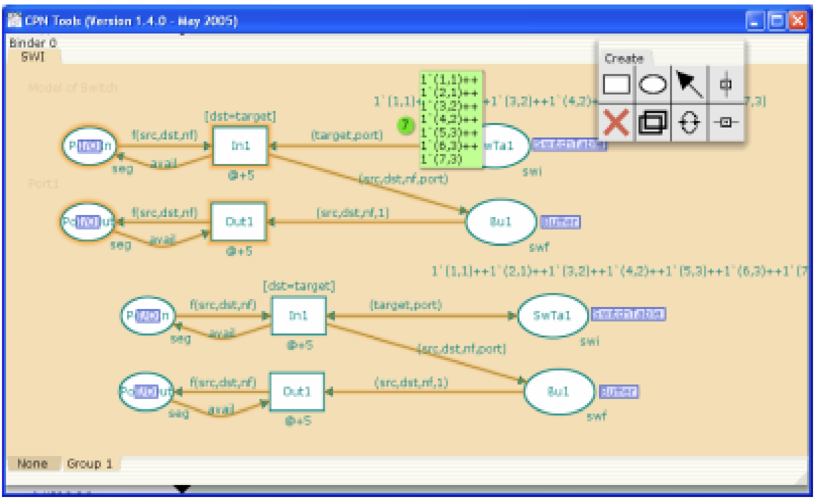


Рис. 6.25. Иерархические временные раскрашенные сети Петри в среде имитационного моделирования CPN Tools

В зависимости от способа диалога исследователя с моделью различают *автоматические* и *диалоговые* модели. В последних исследователь имеет возможность активно управлять ходом имитируемого процесса.

После создания ИМ проводятся имитационные эксперименты (ИЭ).

Перед проведением ИЭ следует организовать *планирование экспериментов*, т.е. организовать выбор и упорядочение множества исходных данных для ИЭ. Этот вопрос решается *теорией планирования эксперимента*.

При имитации СДС результат ИЭ есть случайное событие, характеризующееся набором случайных величин (векторов $Y(t)$, $Z(t)$). Для получения исчерпывающей информации о СДС этого недостаточно. Нужны знания о законах распределения или числовые характеристики этих законов.

Для этого нужно проводить серию экспериментов при одних и тех же исходных данных (начальных условиях).

Последним этапом является статистическая обработка результатов ИЭ методами математической статистики.

Достоверность полученных результатов зависит от объема (количества) экспериментов.

Так, при определении *оценки математического ожидания* количество реализаций должно быть

$$N \approx \left(\Phi^{-1}_{\beta/2}\right)^2 \cdot \frac{\sigma^2}{\Delta_m^2},$$

где $\Phi^{-1}_{\beta/2}$ – функция, обратная функции Лапласа при аргументе $\beta/2$; β – доверительная вероятность; σ – СКО; Δ_m – допустимая разность между истинным значением СВ $x_{\text{ист}}$ и его оценкой СВ $x_{\text{изм}}$.

Аргумент в обратной функции Лапласа равен половине доверительной вероятности исходя из следующего. Для нормированной случайной величины $m_x = 0$, $\sigma = 1$. Тогда

$$\beta(|x_{\text{ист}} - x_{\text{изм}}| \le \Delta_m) = P(m_x - \Delta_m \le x_{\text{ист}} \le m_x + \Delta_m) =$$
$$= P(-\Delta_m \le x_{\text{ист}} \le \Delta_m) = \Phi(\Delta_m) - \Phi(-\Delta_m) = \Phi(\Delta_m) + \Phi(\Delta_m) = 2\Phi(\Delta_m).$$

Отсюда $$\Phi(\Delta_m) = \frac{\beta(|x_{\text{ист}} - x_{\text{изм}}| \le \Delta_m)}{2}.$$

$\Phi^{-1}_{\beta/2}$ в маткаде рассчитывается по формуле

$$\Phi_{\text{обр Лапл}}(\beta) := \text{qnorm}\left(\frac{\beta}{2} + 0{,}5,\ 0,\ 1\right)$$

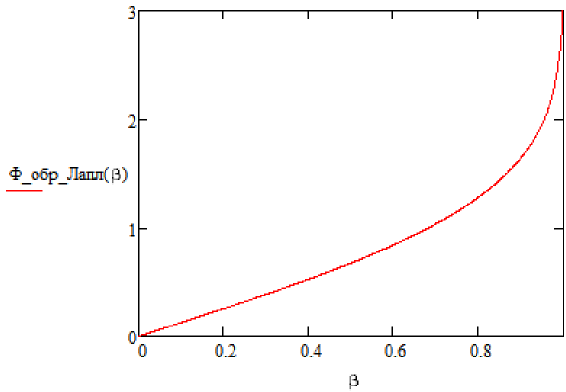


Рис. 6.26. Обратная функция Лапласа

Пример:

$\sigma := 1,\ \ \beta := 0{,}8,\ \ \Delta m := 0{,}15,$

$$N := \Phi\_\text{обр\_Лапл}(\beta)^2 \cdot \frac{\sigma^2}{\Delta m^2} = 72{,}994.$$

При неизвестной дисперсии сначала проводится серия из n предварительных опытов и рассчитывается оценка дисперсии

$$\left.\begin{aligned} &\tilde{m} = \frac{\sum\limits_{i=1}^{n} x_i}{n}; \\ &\tilde{D} = \frac{\sum\limits_{i=1}^{n} (x_i - \tilde{m})^2}{n-1}; \\ &\text{или}\ \ \tilde{D} = \left(\frac{\sum\limits_{i=1}^{n} x_i^2}{n} - \tilde{m}^2\right)\frac{n}{n-1}; \end{aligned}\right\}.$$

Затем в выражении

$$N_M \approx \left(\frac{\Phi_{\beta}^{-1} \sigma}{\Delta_m} \right)^2, \quad \text{оно же} \quad N \approx \left(\Phi_{\beta}^{-1} \right)^2 \frac{\sigma^2}{\Delta^2},$$

делается замена

$$N \approx t(n, \beta)^2 \frac{\hat{\sigma}^2}{\Delta^2},$$

где $\hat{\sigma}^2 = \tilde{D}$ – оценка дисперсии; $t(n, \beta)$ – определяется из табл. 6.1.

Таблица 6.1

β	n					
	5	10	20	40	60	∞
0,8	1,53	1,38	1,33	1,30	1,30	1,28
0,9	2,13	1,83	1,73	1,69	1,67	1,64
0,95	2,77	2,26	2,09	2,01	2,00	1,96
0,99	4,60	3,25	2,86	2,70	2,66	2,58

Для определения оценки вероятности P события требуемое количество экспериментов N вычисляется по формуле

$$N_P = \frac{t(n, \beta)^2 \cdot P(1-P)}{\Delta_P^2},$$

где Δ_P – допустимое (абсолютное) отклонение значения вероятности от ее оценки.

Таким образом, можно сформулировать основные принципы имитационного статистического моделирования:

– моделируется сущность исследуемого процесса;

– сущность исследуемого процесса моделируется путем имитации данного процесса;

– целесообразно строить модульную иерархическую структуру имитационной модели;

– события, процессы и активности необходимо координировать по времени;

– перед проведением экспериментов организуется планирование экспериментов, включающее планирование упорядоченного множества наборов исходных данных и количество реализаций процесса;

– после осуществления экспериментов проводится статистическая обработка результатов методами математической статистики.

Выводы.

1. Моделирование является мощным инструментом научных исследований, существенно повышающим их эффективность.

2. Среди различных видов моделирования ведущее место в научных исследованиях занимает математическое моделирование.

3. Исследование сложных процессов и систем целесообразно с помощью имитационных моделей, реализованных с использованием специального программного обеспечения.

4. При моделировании важную роль играет планирование экспериментов и корректная обработка их результатов.

7. ПРАКТИЧЕСКИЕ ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ МОДЕЛИРОВАНИЯ

7.1. ОБЩИЕ СВЕДЕНИЯ О СИСТЕМАХ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ (САПР)

7.1.1. ПРИНЦИПЫ СОЗДАНИЯ САПР

Проектирование – процесс разработки документов (описаний) по созданию нового объекта в заданных условиях. При этом если новизна объекта определяет творческую составляющую процесса проектирования, то первая составляющая, связанная с разработкой комплекта необходимых документов, связана в большинстве случаев с выполнением трудоемких рутинных операций. В условиях быстрого морального старения технических средств необходимо ускорение процессов их проектирования. Основным приемом реализации такого ускорения является автоматизация процесса проектирования.

Модель процесса проектирования приведена на рис. 7.1

Из анализа этой модели следует, что процесс проектирования – управляемый итеративный процесс, минимизирующий в процессе итераций величину отклонения модели от желаемого результата.

Автоматизация проектирования – процесс внедрения в проектирование компьютера и других технических средств – помощников конструктора – при рациональном распределении функций между ними. Обычно подразумевается, что человек выполняет задачи творческого характера, а компьютер – жесткие алгоритмические процедуры. По мере развития искусственных «интеллектуальных» возможностей ЭВМ перераспределение обязанностей между человеком и машиной все более сдвигается в сторону усиления роли компьютера: в него закладываются пополняющиеся справочники, типовые варианты, экспертные системы и т.д., которые «подсказывают» (или автономно реализуют) тот или иной оптимальный вариант. При такой тенденции вполне закономерно наступление ситуации, когда машины будут выполнять почти все процедуры этапа проектирования, т.е. машины будут «думать», принимать решения в условиях неопределенности и т.д. Возможно наступление времени, когда машины будут создавать машины. Однако функция постановки задачи на проектирование, определение условий, введение ограничений, т.е. стратегическое управление проектированием при любом варианте должно остаться за человеком.

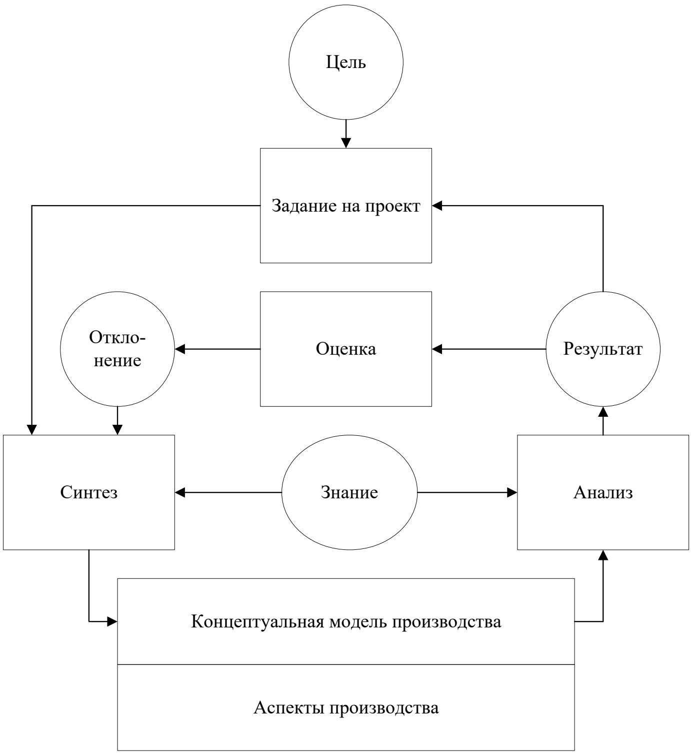


Рис. 7.1. Модель процесса проектирования

Таким образом, основной элемент современной системы проектирования – «вооруженный» техническими средствами (основным из которых является компьютер с соответствующим математическим и программным обеспечением, базами данных и знаний) человек-конструктор. Поэтому под системой автоматизированного проектирования (САПР) следует понимать систему конструктор-компьютер (с соответствующей периферией). Напомним, что любая система – это целостная совокупность элементов и связей между ними, т.е. система – это целостная структура. Следствием целостности является эмерджентность – появление в системе надсистемных, сверхаддитивных свойств, которыми элементы системы по отдельности не обладают. Основным надсистемным свойством САПР является существенное сокращение времени на проектирование, уменьшение размеров материального носителя информации (вследствие инвариантности сигнала к своему носителю), появление возможностей его неоднократного использования и дальнейшего наращивания и преобразования.

Основные принципы создания САПР:

- принцип системной модульной иерархии, согласно которому САПР представляет собой совокупность иерархических подсистем (уровней) как на техническом, так и на программном уровнях. На высшем уровне используется наименее детализированное представление, отражающее только общие черты и особенности проектируемой системы. С понижением уровня возрастает степень детализации. При этом каждый уровень рассматривается как отдельный модуль (отдельная система) со своими входами и выходами. Другими словами, элемент верхнего уровня предстает системой на нижнем уровне, каждый элемент которого есть система на более низком уровне и т.д. При этом под модулем целесообразно понимать функциональный узел [1], предназначенный для многократного применения и имеющий упорядоченные значения параметров. Проектирование представляет последовательно-параллельное выполнение проектных процедур по созданию описаний на каждом иерархическом уровне;

- принцип унификации (как проявление принципа минимальности [2]), направленный на использование типовых и унифицированных элементов и узлов. При этом существует базовый комплект САПР, который периодически подвергается модернизации и на основе которого создаются модифицированные варианты САПР;
- принцип информационной согласованности (при переходе от одной программы к другой не требуется видоизменение исходных данных);
- принцип развития (как проявление более общих принципов: виртуальности – возможности несуществующего и его пространственно-временной ограниченности и предвидения [2]). Этот принцип обеспечивает пополнение, совершенствование и обновление составных частей САПР как открытой системы.

Стадии создания САПР:

- технико-экономическое обоснование создания САПР;
- проведение научно-исследовательских работ и составление тактико-технического задания на создание САПР;
- эскизное проектирование САПР;
- разработка технического проекта (окончательные решения по структуре, процедурам, процессу функционирования, видам обеспечения, пректно-сметной документации);
- разработка рабочей документации по видам обеспечения;
- изготовление несерийных компонентов;
- ввод САПР в действие (строительно-монтажные работы, пуско-наладочные работы, опытная эксплуатация САПР, приемочные испытания, приемка САПР в эксплуатацию).

7.1.2. КЛАССИФИКАЦИЯ САПР

В соответствии с ГОСТ 23501.108–85 в качестве классификационных признаков САПР используются: тип, разновидность, сложность объекта проектирования, уровень автоматизации проектирования, характер и число выпускаемых проектных документов, число уровней в структуре технического обеспечения САПР.

По типу объектов проектирования выделяют группы: САПР изделий машиностроения; САПР изделий приборостроения; САПР технологических процессов в машино- и приборостроении; САПР объектов строительства; САПР технологических процессов в строительстве; САПР программных изделий; САПР организационных систем.

По сложности объектов проектирования в зависимости от числа элементов выделяют: простые объекты с числом элементов до 10^2, объекты средней сложности (10^2 – 10^3), сложные объекты (10^3 – 10^4), очень сложные объекты (10^4 – 10^5), объекты очень высокой сложности (число элементов более 10^6).

По уровню автоматизации проектирования выделяют САПР низкоавтоматизированные (до 25% проектных процедур), среднеавтоматизированные (25…50%), высокоавтоматизированные (более 50%). Чтобы отнести САПР к третьей группе в ней должны быть использованы методы многовариантного оптимального проектирования.

По числу выпускаемых проектных документов различают САПР малой, средней и высокой производительности. При этом число проектных документов колеблется от 10^3 до 10^6.

По числу уровней в структуре технического обеспечения: одноуровневые САПР (типа автоматизированного рабочего места), двухуровневые САПР (радиальной или кольцевой структуры с объединением автоматизированных рабочих мест в кольцевую вычислительную сеть), трехуровневые САПР (добавляется программно-управляемое оборудование).

7.1.3. ВИДЫ ОБЕСПЕЧЕНИЯ САПР

К основным видам обеспечения САПР относят: математическое обеспечение, программное обеспечение, информационное обеспечение, техническое обеспечение, лингвистическое обеспечение, организационное обеспечение.

Математическое обеспечение САПР. Основа математического обеспечения – математические модели объектов проектирования, методы и алгоритмы выполнения проектных процедур и разработки программного обеспечения САПР. Часть математического обеспечения – универсальная. К ней относятся принципы построения математических моделей, численные методы, методы поиска экстремумов и др. Возможность проведения экспериментов на математических моделях проектируемых объектов, полностью или значительно сократив дорогостоящее физическое моделирование, – одно из основных преимуществ машинных способов проектирования. При этом математические модели должны удовлетворять требованиям универсальности, адекватности, точности и экономичности.

Математическое обеспечение обычно разделяют на две части:

- математические методы и построенные на их основе модели объектов проектирования;
- описание технологии автоматизированного проектирования.

Первая часть МО, связанная с построением математической модели объекта проектирования, хотя и специфична для каждого объекта, но в рамках существующих математических методов в большинстве случаев реализуема. Получение второй составляющей в виде совокупности процедур, инвариантных к объекту проектирования, – трудная задача. Очевидно, наиболее целесообразным можно считать подход, связанный с разработкой математического аппарата моделирования типового процесса проектирования и созданием базовых программно-методических комплексов.

Программное обеспечение (ПО) САПР. Представляет собой совокупность всех программ и эксплуатационной документации к ним, необходимых для выполнения автоматизированного проектирования. Основные требования к ПО: адаптируемость, развитие, гибкость (возможность оперативного изменения), надежность, компактность. Оно делится на общесистемное и специальное (прикладное).

Общесистемное ПО предназначено для организации функционирования технических средств, т.е. для планирования и управления вычислительным процессом, распределения имеющихся ресурсов. Оно создается для многих приложений и потому не отражает специфику САПР, т.е. является инвариантным к объектам проектирования и поэтому должно быть защищено от пользователей. Пример – операционные системы ЭВМ. Основные функции общесистемного ПО САПР: ввод, вывод и обработка инструкций пользователей; диалоговая взаимосвязь с пользователем в процессе проектирования; хранение, поиск, анализ, модификация данных, защита их целостности; решение общесистемных задач; контроль и диагностика в процессе решения задач проектирования.

Состав общесистемного САПР: мониторная диалоговая система, система управления базами данных, информационно-поисковые системы, геометрические и графические процессоры, средства формирования графической и текстовой информации, средства для выполнения общетехнических расчетов. В настоящее время наиболее распространены проблемно-ориентированные операционные системы.

Специальное (*прикладное*) *ПО* создается для непосредственного выполнения проектных процедур, т.е. его основная задача – получение проектных решений. Оно имеет обычно форму пакетов прикладных программ, каждый из которых обслуживает определенный этап процесса проектирования или группу однотипных задач внутри различных этапов.

Иногда отдельно выделяют *базовое программное обеспечение,* к которому относят программы, обеспечивающие правильное функционирование прикладных программ.

Информационное обеспечение САПР. Основу информационного обеспечения (ИО) САПР составляют всевозможные данные, которыми пользуются проектировщики в процессе проектирования непосредственно для выработки проектных решений. Эти данные могут быть представлены в виде тех или иных документов на различных носителях, содержащих сведения справочного характера о материалах, комплектующих изделиях, типовых проектных решениях, параметрах элементов, сведения о состоянии текущих разработок в виде проектных решений, параметров, структур и т.п. При этом данные одного процесса преобразования могут быть исходными для другого процесса. Все данные составляют информационный фонд САПР, к которому имеется оперативный доступ. Данные информационного фонда можно разделить на следующие типы:

- программные модули, мало изменяющиеся в течение жизненного цикла САПР;
- исходные и результирующие данные, необходимые при выполнении программных модулей в процессе преобразования;
- нормативно-справочную проектную документацию, включающую в себя: справочные данные о материалах, элементах, унифицированных узлах, государственные и отраслевые стандарты, типовые проектные решения, регламентирующие документы;
- текущую проектную документацию, отражающую состояние и ход выполнения проекта.

Тенденцией современного этапа развития САПР является переход от использования только данных (банка данных в виде базы данных и системы управления ею) к использованию как данных, так и знаний (в виде базы знаний). Вопросы продуцирования и использования знаний исследуются в разделе кибернетики, получившем название «Искусственный интеллект».

Техническое обеспечение САПР. Представляет собой совокупность взаимосвязанных и взаимодействующих технических средств в виде автоматизированных рабочих мест на базе персональных ЭВМ, обеспечивающих реализацию процедур автоматизированного проектирования.

Лингвистическое обеспечение САПР. Его основу составляют специальные проблемно-ориентированные языки проектирования, предназначенные для описания процедур автоматизированного проектирования и проектных решений. Основное требование к этим языкам – универсальность применения, простота их освоения проектировщиком и эффективность описания.

Методическое обеспечение САПР. Включает в себя: положения и инструкции по эксплуатации САПР, приказы, штатные расписания, квалификационные требования и другие документы, регламентирующие организационную структуру подразделений проектной организации и взаимодействие с комплексом средств автоматизированного проектирования.

7.1.4. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В САПР

Математические модели, используемые в САПР, служат для описания свойств объектов в процедурах автоматизированного проектирования. Если проектная процедура включает создание математической модели и оперирование ею с целью получения полезной информации об объекте, то говорят, что процедура выполняется на основе математического моделирования. Основные признаки классификации и типы математических моделей в соответствии с этими признаками приведены в табл. 7.1.

7.1. Основные типы математических моделей

Признак классификации	Математические модели
Характер отображаемых свойств объекта	Структурные; функциональные
Принадлежность к иерархическому уровню	Микроуровня; макроуровня; метауровня
Степень детализации описания внутри одного уровня	Полные; макромодели
Способ представления свойств объекта	Аналитические; алгоритмические; имитационные
Способ получения модели	Теоретические; эмпирические

Структурные модели предназначены для отображения структурных свойств объекта. Выделяют топологические и геометрические математические модели. В топологических моделях с пространственно-временных позиций отображаются состав и взаимосвязи элементов объекта. С помощью моделей этого типа решают задачи привязки конструктивных элементов к определенным пространственным позициям или определенным моментам времени. Топологические модели имеют форму таблиц, графов, списков.

Геометрические модели кроме взаимного расположения деталей содержат информацию о форме каждой детали. Наиболее очевидны аналитические математические модели, которые посредством уравнений линий и поверхностей описывают форму объекта. Однако для сложного объекта его полное аналитическое описание будет слишком громоздким. Поэтому в этих случаях применяют кусочную аппроксимацию формы объекта, т.е. модель представляют в виде некоторого множества хорошо согласующихся друг с другом «кусочков» (каркасная модель). Кроме того, в случае сложных объектов часто используют кинематическую модель, в которой поверхность представляется в параметрическом $R(u, v)$, где $R=(x, y, z)$, а u и v – параметры. Тогда поверхность объекта можно представить как результат перемещения в трехмерном пространстве кривой $R(u)$.

Канонические модели позволяют внести простой геометрический смысл в параметры модели, имеющие простую связь с его формой. Например, для плоского многоугольника такими параметрами являются координаты вершин, для цилиндра – направляющие косинусы и координаты некоторой точки оси и радиус цилиндра.

Функциональные модели предназначены для отображения физических или информационных процессов, протекающих в объектах при его функционировании или изготовлении. Обычно функциональные модели представляют собой системы уравнений, связывающих фазовые переменные, внутренние, внешние и выходные параметры.

Особенностью математических моделей на микроуровне является отражение непрерывных физических процессов в пространстве и времени. Типичный представитель – дифференциальные уравнения в частных производных, с помощью которых рассчитываются поля механических напряжений и деформаций, электрических потенциалов, давлений, температур и т.д. На макроуровне используют укрупненную дискретизацию пространства на функциональном уровне, что приводит к использованию для описания модели систем обыкновенных дифференциальных уравнений. В этих уравнениях независимой переменной является время, а вектор зависимых переменных составляют фазовые переменные, характеризующие состояние укрупненных элементов дискретизованного пространства. Такими переменными являются силы и скорости механических систем, напряжения и силы тока электрических систем, давления и расходы гидравлических систем. Системы обыкновенных дифференциальных уравнений являются универсальными моделями для анализа как динамических, так и установившихся состояний объектов. Кроме того, для установившихся режимов модели могут быть представлены в виде системы алгебраических уравнений. На метауровне в качестве элементов принимают достаточно сложные совокупности деталей, причем для многих объектов можно пользоваться также системой обыкновенных дифференциальных уравнений.

Полные математические модели содержат фазовые переменные, характеризующие состояние всех элементов объекта и межэлементных связей. В макромодели отдельные элементы и связи объединены.

Аналитические модели представляют собой явные выражения для вектора выходных параметров Y системы как функций от вектора входных X и вектора внутренних Q параметров $Y = F(X, Q)$. Такие модели наиболее экономичны, однако не всегда их возможно получить с достаточной точностью для сложных объектов. В таком случае используют алгоритмические модели, выражающие связи выходных параметров с входными и внутренними в виде алгоритмов как функционалов. Имитационные модели отражают поведение исследуемого объекта во времени при задании внешних воздействий на него. Пример – модели динамических объектов в виде системы обыкновенных дифференциальных уравнений.

Теоретические модели создаются в результате исследования процессов и их закономерностей, присущих рассматриваемому классу объектов и явлений. Эмпирические – в результате изучения внешних проявлений свойств объекта с помощью измерений фазовых переменных на внешних входах и выходах и обработки результатов измерений.

Поскольку современные технические системы в большинстве своем можно отнести к сложным техническим системам, то целесообразно чуть более подробно остановиться на наиболее приемлемом для их исследования и создания методе имитационного моделирования, позволяющем получить достоверные сведения о процессах, протекающих в сложных системах, и выбрать наиболее целесообразное поведение системы в различных условиях функционирования. Исходными данными для такого моделирования являются известные параметры технической системы и значения начальных условий.

Наиболее часто имитационное моделирование используют в случаях:

- необходимости получения дополнительного знания о неизвестных сторонах объекта моделирования в случае отсутствия законченной постановки задачи исследования;
- повышенной сложности и трудоемкости аналитических методов исследования. При этом имитационное моделирование предлагает робастный (более простой и достоверный) способ решения задачи;
- изучения отдельных частных процессов в исследуемой системе;
- невозможности наблюдения функционирования объекта исследования в реальных условиях (например, применение ракетно-ядерного оружия);
- необходимости растягивания или сжатия по оси времени исследуемого явления или процесса, прогнозирования его поведения в неясных ситуациях;
- решения задачи подготовки специалистов (например, операторов) по эксплуатации сложной дорогостоящей техники.

При построении имитационной модели (ИМ) исследователя интересует прежде всего возможность вычисления некоторого функционала, заданного на множестве реализаций процесса функционирования изучаемой сложной системы и характеризующего поведение объекта имитации. Наиболее важным для исследователей функционалом является показатель эффективности системы. Имитируя различные реальные ситуации на ИМ, исследователь получает возможность решения таких задач, как оценка эффективности различных принципов управления системой, сравнение вариантов структуры системы, определение степени влияния изменений параметров системы и начальных условий имитации ее поведения на показатель эффективности системы. В целом можно отметить, что, несмотря на невозможность построения адекватной во всех деталях имитационной модели, а также априорной оценки степени близости модели к реальности, имитационное моделирование получает все большее распространение. Основные достоинства имитационного моделирования: возможность описания поведения компонент сложной системы на высоком уровне детализации, отсутствие ограничений на вид зависимостей между параметрами модели и состоянием внешней среды, возможность исследования динамики взаимодействия компонент системы во времени и пространстве ее параметров.

7.2. ОБЗОР ВОЗМОЖНОСТЕЙ НАИБОЛЕЕ ИЗВЕСТНЫХ И РАСПРОСТРАНЕННЫХ ПРОГРАММНЫХ СРЕДСТВ АВТОМАТИЗАЦИИ ПРОЕКТИРОВАНИЯ

Основная тенденция развития как общесистемного, так и специализированного современного программного обеспечения – ориентация на «непрограммирующего пользователя». В этом случае пользователь такого пакета получает возможность сосредоточиться на сущности самой задачи, а не на способах ее программной реализации. В свою очередь пользователь должен ясно представлять возможности используемого пакета и заложенных в нем методов, а также уметь выбрать необходимый пакет, соответствующий решаемой задаче. Поэтому в настоящее время для научно-технических расчетов на компьютерах все чаще и чаще используются не традиционные языки программирования и не электронные таблицы, а специальные математические программы типа Mathematica, MatLab, Maple, Mathcad, Gauss, Reduce, Eureka и др.

7.2.1. ПАКЕТ Mathcad

Все этапы создания и использования математической модели легко проследить при работе с пакетом Mathcad фирмы «MathSoft Inc.» (США).

Математические пакеты, в особенности Mathcad, позволяют специалистам в конкретной научно-технической области очень быстро освоить работу на компьютере и реализовать на них математические модели, не вдаваясь в тонкости программирования на традиционных языках (fortran, C, Pascal, BASIC и др.).

Mathcad – универсальный математический пакет, предназначенный для выполнения инженерных и научных расчетов. Математическое обеспечение пакета позволяет решать многие задачи в объеме инженерного вуза.

Пакет объединяет в себе: редактор математических формул, интерпретатор для вычислений, библиотеку математических функций, процессор символьных преобразований, текстовый редактор, графические средства представления результатов. Пакет Mathcad относится к интегрированным пакетам, т.е. позволяет не только произвести вычисления, но и получить документ – итоговый отчет с комментариями, формулами, таблицами и графиками.

К положительным качествам MATHCAD следует отнести открытость – все приведенное в документе может быть воспроизведено, а интеграция в одном документе исходных данных, метода решения и результатов позволяет сохранить настройки для решения подобных задач.

Основные возможности и преимущества работы в среде Mathcad:

- математические выражения в среде Mathcad записываются в их общепринятой нотации: числитель находится сверху, а знаменатель – внизу; в интеграле пределы интегрирования также расположены на своих привычных местах, что очень важно при анализе математических моделей, форма и содержание которых едины;

- процесс создания «программы» идет параллельно с ее отладкой;
- в пакет Mathcad интегрирован довольно мощный математический аппарат, позволяющий решать возникающие проблемы без вызова внешних процедур;
- пакет Mathcad дополнен справочником по основным математическим и физико-химическим формулам и константам, которые можно автоматически переносить в документ без опасения внести в них искажения;
- к пакету Mathcad разработаны дополнительные электронные учебники по различным дисциплинам: решение обыкновенных дифференциальных уравнений, статистика, термодинамика, теория управления, сопротивление материалов и т.д. Прежде чем решать возникшую проблему, пользователь может изучить электронный учебник и перенести из него в свой документ нужные фрагменты, отдельные формулы и константы;
- решая поставленную задачу, пользователь может вводить не только числовые значения переменных, но и дополнить их размерностями. При этом можно выбирать и систему единиц (СИ, кг-м-с, г-см-с, британская), и конкретные размерности (мм, дюймы, футы и т.д.);
- система Mathcad оборудована средствами анимации, что позволяет реализовать созданные модели не только в статике (числа, таблицы, графики), но и в динамике (анимационные клипы);
- в систему Mathcad интегрированы средства символьной математики, что позволяет решать поставленные задачи (этап задачи) не только численно, но и аналитически;
- кроме того, не следует забывать, что пакет Mathcad – это полноценное Windows-приложение. Решая поставленную задачу, можно

в статике (через буфер обмена Windows) или в динамике (OLE-технологии) передать данные в среду другой программы (в среду языка fortran, например) и там решить часть задачи.

- Mathcad – приложение Stencil for Visio дает возможность пользователям легко включать вычисления и стандартную математическую нотацию Mathcad в рисунки Visio, что облегчает обмен данными между Mathcad и Visio;
- мастер сбора данных в реальном времени позволяет пользователям читать и посылать данные через National Instruments аналоговые интерфейсные платы;
- модуль ODBC (Open Database Connectivity) позволяет запрашивать информацию из SQL-совместимых баз данных, включая Microsoft Access, FoxPro и другие SQL- совместимые базы данных;
- модуль MATLAB дает возможность переносить данные и функции из MATLAB 5 в рабочие документы Mathcad;
- приложение AutoCAD Add-In поддерживает более 10 входов в Mathcad объект и позволяет брать выходные данные из ячеек DataTable и копировать их в буфер обмена;
- возможно чтение и сохранение рабочих документов Mathcad в HTML-формате с поддержкой MathML;
- набор разработчика (Component Software Developers Kit – SDK) позволяет разрабатывать пользовательские Mathcad модули в C++;
- функции преобразования координатных систем позволяют получать различные типы графиков. Поддерживаются следующие преобразования систем координат:

– от сферической к Евклидовой;

– от цилиндрической к Евклидовой;

– от Евклидовой к сферической;

– от Евклидовой к цилиндрической;

– от полярной к Евклидовой;

– от Евклидовой к полярной.

Основное преимущество пакета перед типичными языками программирования и другими программными средствами – естественный математический язык, на котором формулируется решаемая задача и более широкие возможности.

7.2.2. ПАКЕТ MATLAB

Модульный подход к моделированию прослеживается и в современном пакете MATLAB фирмы «The MathWorks Inc» (США), который по существу переместился с «больших» машин на персональные компьютеры.

Система MATLAB предназначена для выполнения инженерных и научных расчетов и высококачественной визуализации получаемых результатов. Эта система применяется в математике, вычислительном эксперименте, имитационном моделировании.

В пакет входит множество хорошо проверенных численных методов, операторы графического представления результатов, средства создания диалогов. Отличительной особенностью MATLAB по сравнению с обычными языками программирования является матричное представление данных и большие возможности матричных операций над данными. Используя пакет MATLAB, можно построить довольно сложную математическую модель или написать свою программу (весьма похожую на Фортран-программу). А можно, используя SIMULINK и технологию визуального моделирования, составить имитационную модель и исследовать ее в реальном масштабе времени.

В настоящее время MATLAB используется во множестве областей, среди которых обработка сигналов и изображений, проектирование систем управления, финансовые расчеты и медицинские исследования. Его открытая архитектура делает возможным использование MATLAB и сопутствующих продуктов для исследования данных и создания собственных инструментов, использующих функциональные возможности MATLAB.

Для проектирования систем управления, цифровой обработки сигналов, коммуникационных систем широко используется Simulink, позволяющий моделировать динамические системы, оценивать их работу, модифицировать проект с помощью графических блок-диаграмм. Simulink – это интерактивная среда для моделирования и анализа широкого класса динамических систем.

Благодаря тесной интеграции с MATLAB, Simulink имеет непосредственный доступ к широкому диапазону средств проектирования и анализа. Традиционный подход к проектированию систем обычно заключается в создании прототипа, за которым следует всестороннее тестирование и внесение соответствующих изменений. Этот подход требует больших временных и финансовых затрат. Эффективной и общепринятой альтернативой является имитационное моделирование. Simulink – мощный инструмент для моделирования, обеспечивающий быстрое построение и тестирование виртуальных прототипов и дающий доступ к любому уровню детализации проекта с минимальными усилиями. Используя Simulink для итеративного исправления проекта до построения прототипа, инженер может разработать проект быстро и эффективно.

MATLAB содержит инструменты для сбора данных, анализа и обработки данных, визуализации и цифровой обработки сигналов и изображений, создания алгоритмов и проектирования, моделирования и имитации, программирования и разработки приложений.

Краткий обзор возможностей

MATLAB выполняет множество компьютерных задач для поддержки научных и инженерных работ, начиная от сбора и анализа данных до разработки приложений. Среда MATLAB объединяет математические вычисления, визуализацию и мощный технический язык. Встроенные интерфейсы позволяют получить быстрый доступ и извлекать данные из внешних устройств, файлов, внешних баз данных и программ. Кроме того, MATLAB позволяет интегрировать внешние процедуры, написанные на языках Си, Си++, Фортран и Java с MATLAB-приложениями.

MATLAB имеет широкий спектр применения, включая цифровую обработку сигналов и изображений, проектирование систем управления, естественные науки, финансы и экономику, приборостроение. Открытая архитектура позволяет легко использовать MATLAB и сопутствующие продукты для исследования данных и быстрого создания пользовательских инструментов.

Основные функции:

- быстрые и точные численные алгоритмы;
- графика для анализа и отображения данных;
- интерактивный язык и среда программирования;
- инструменты для настройки пользовательских интерфейсов;
- интерфейсы с внешними языками, такими как Си, Си++, Фортран и Java;
- поддержка импорта данных из файлов и внешних устройств плюс доступ к базам данных и вспомогательному оборудованию при помощи приложений;
- преобразование MATLAB приложений в Си и Си++ при помощи набора Compiler Suite.

Этот широкий набор возможностей делает MATLAB идеальной базой для решения многих технических проблем.

Работа в среде MATLAB

Среда MATLAB спроектирована для интерактивных или автоматических вычислений. Используя встроенные математические и графические функции и простые в использовании инструменты, пользователь получает возможность анализировать и отображать данные «на лету». Структурированный язык и программные инструменты позволяют сохранить результат интерактивных исследований, разрабатывать собственные алгоритмы и приложения.

Пользователи, работающие в широком спектре применений с различным уровнем сложности, используя интерфейс MATLAB, могут настроить его для удовлетворения своего стиля работы и избирательного использования функций, необходимых для каждой фазы проекта.

Моделирование явлений и процессов в среде MATLAB с помощью программы Simulink

Фирмой MathWorks вместе с MATLAB поставляется сопутствующая программа Simulink – интерактивная система для имитации динамических систем. Эта программа позволяет представить исследуемую динамическую систему при помощи соединенных между собой функциональных блоков (блок-диаграммы), а затем исследовать ее поведение в динамике.

Сам процесс построения диаграмм очень прост: необходимые блоки перетаскиваются из библиотеки блоков в рабочее окно пакета при помощи мыши, затем нужным образом соединяются также при помощи мыши, затем настраиваются их параметры при помощи диалоговых окон, вызываемых двойным щелчком мыши по блоку, и, наконец, оформляется внешний вид блок диаграммы – размер и цвет блоков, их наименование, выбирается размер и тип шрифта для надписей и т.д.

Имеется разница между моделированием систем блочными диаграммами пакета Simulink и моделированием командами MATLAB. При Simulink-моделировании команды всех блоков выполняются одновременно в течение каждого временного шага (так называемая имитация временного потока). Команды MATLAB выполняются последовательно (имитация потока данных). Для специальных приложений могут понадобиться как те, так и другие модели систем.

Применение сочетания программ MATLAB и Simulink привело к разработке широкого класса профессиональных инструментальных приложений (**toolboxes** – тулбоксы, наборы инструментов) для генерации, анализа и оптимизации систем.

Наборы инструментов – это больше, чем набор полезных функций. Без преувеличения можно сказать, что они представляют собой последнее слово в разработке (исследованиях) в таких областях, как управление, обработка сигналов, идентификация систем и многих других. Поэтому, освоив и применяя наборы инструментов MATLAB, можно достичь уровня разработчиков (исследователей) мирового класса.

Для многих наборов инструментов (таких как коммуникация, обработка сигналов, энергетические системы и других) поставляются наборы блоков (**blocksets**) для создания динамических моделей при помощи блок-диаграмм программы Simulink.

Далее приведен список профессиональных наборов инструментов, распространяемых в настоящее время компанией MathWorks. Этот список быстро расширяется, каждый год разрабатываются новые пакеты.

Набор инструментов **«Communications»** содержит средства для разработки современных коммуникационных систем, включая моделирование в реальном масштабе времени. Он охватывает такие области применения, как электронные телекоммуникации, телефония, авиация и космонавтика, а также компьютерная периферия.

Набор инструментов **«Системы управления»** – основной пакет MATLAB для моделирования, анализа и проектирования автоматических систем управления. Широко применяется в разработке высокотехнологических систем, как, например, автомобили и аэрокосмическая техника, компьютерная периферия и управление процессами, а также в менее очевидных приложениях, как, например, стиральные машины и фотокамеры.

Финансовый набор инструментов программы MATLAB содержит необходимые функции для ввода, обработки и вывода финансовых данных, финансового анализа и прогноза. Применения включают ценовую политику, расчет процентов и доходов, анализ производных и оптимизацию портфелей. Он оперирует во взаимодействии с наборами инструментов: статистическим и оптимизации. Рекомендуется также графический интерфейс пакета Simulink для моделирования финансовых систем как нестохастическими методами, так и методами Монте-Карло.

Набор инструментов **идентификации систем методом спектрального анализа (Frequency-Domain System Identification – FDI)** включает набор *m*-файлов для моделирования линейных систем на основе измерений их частотных окликов.

Набор инструментов **«Нечеткая логика» («The Fuzzy Logic»)** содержит средства, предназначенные для проектирования, моделирования и анализа систем с нечетким откликом. Он имеет легко осваиваемые и в то же время мощные средства для преобразования входных данных в выходные данные системой правил и связей произвольной сложности, выраженных обычным языком. Системы могут быть имитированы в MATLAB или включены в блочные диаграммы Simulink с возможностью генерации кода для независимого выполнения.

Набор инструментов **Higher-Order Spectral Analysis** содержит инструменты для обработки сигналов, являющихся результатом нелинейных процессов или искаженных негауссовским шумом, с использованием спектрального разложения высокого порядка.

Набор инструментов **обработки изображений (Image Processing)** содержит функции для анализа, статистической обработки, усиления, восстановления и двумерного преобразования изображений (фильтры, цвет, геометрия, морфология).

Набор инструментов **управление линейными матричными неравенствами (LMI Control)** позволяет эффективно решать системы линейных матричных неравенств (Linear Matrix Inequalities), которые возникают при решении многих задач в таких областях, как управление, распознавание, фильтрация, проектирование структур, теория графов и линейная алгебра. **LMI Control** также содержит функции для проектирования и анализа таких характеристик систем управления, как помехоустойчивость, производительность и др.

Набор инструментов **Model Predictive Control** особенно важен для управления системами с большим количеством входных и выходных переменных, имеющих много связей. Широко применяется в химической инженерии для управления процессами.

Набор инструментов **Мю-анализа и синтеза (Mu-Analysis and Synthesis)** содержит набор функций для использования в анализе и проектировании устойчивых линейных систем со многими переменными. Его цель – сделать доступными некоторые последние достижения в теории систем управления в среде MATLAB, а именно H-бесконечное оптимальное управление, *m*-анализ и синтез.

Набор инструментов **Foundation** включает более 200 подпрограмм численного расчета из хорошо известных библиотек NAG Fortran, применяемых для задач пограничного слоя, оптимизации, адаптивной квадратуры, подгонки при помощи кривой или поверхности и многих других.

Пакет Neural Network (нейронные сети) – это набор MATLAB-функций для проектирования и имитации нейронных сетей. Нейронные сети – это компьютерные архитектуры, инспирированные биологическими нервными системами. Они применяются в областях, в которых формальный анализ чрезвычайно труден или невозможен, как, например, распознавание образов, идентификация и управление нелинейными системами.

Набор инструментов **Optimization** включает методы нахождения экстремумов линейных и нелинейных функционалов при наличии связей и ограничений.

Набор инструментов **Partial Differential Equation** предназначен для решения уравнений в частных производных в пространстве двух измерений и времени методом конечных элементов. Он включает набор функций и интуитивный графический пользовательский интерфейс для предварительной обработки, решения и последующей визуализации.

В модуле имеются восемь готовых модулей для таких инженерных и физических применений, как перенос тепла, строительная механика, электростатика, магнитостатика и диффузия.

Набор инструментов **QFT Control Design** – применение спектрального приближения в проектировании контроллеров для неопределенных систем. Он находит оптимальное решение на основе компромисса между сложностью контроллеров (а следовательно, и возможностью внедрения) и их техническими характеристиками.

Набор инструментов **Robust Control** содержит специализированный набор инструментов для анализа и синтеза систем управления, устойчивых по отношению к случайным возмущениям, которые могут возникнуть в реальном мире.

Набор инструментов **Signal Processing** содержит средства обработки сигналов. Применения включают аудиосистемы (аналоговая и цифровая звукозапись), видеосистемы (цифровое телевидение, обработка сигналов и сжатие данных), телекоммуникации (факс и голосовой телефон), медицину (CAT-сканирование, магниторезонансное изображение), геофизику и эконометрику.

Набор инструментов **Spline** содержит набор *m*-файлов для конструирования сплайнов, которые используются для кусочно-полиномиальных аппроксимаций других функций. Аппроксимации сплайнами имеют некоторые преимущества по сравнению с другими видами аппроксимаций.

Набор инструментов **Statistics** содержит набор *m*-файлов и графических средств для исследования фундаментальных законов статистики и теории вероятностей, статистического анализа данных и моделирования методом Монте-Карло.

Набор инструментов **Symbolic Math** позволяет в среде MATLAB выполнять аналитические вычисления, используя команды и некоторые дополнительные специализированные функции основной символьной библиотеки программы Maple V.

Набор инструментов **System Identification** включает набор средств, предназначенных для оценки и идентификации систем. Он позволяет построить математическую модель физической системы (такой, как электрический мотор или даже финансовый рынок) на основе только входных и выходных характеристик.

Набор инструментов **Wavelet** содержит обширный набор программ, позволяющих повысить производительность методов, использующих технику Фурье-анализа для изучения многомасштабных или нестационарных явлений, а также улучшить их понимание. Он применяется во многих приложениях обработки сигналов и изображений, включая аудио-обработку, коммуникации, геофизику, финансы и медицину.

Мастерская реального времени Simulink (Real-Time Workshop). Этот пакет автоматически генерирует Си-код прямо из блочных диаграмм пакета Simulink, что обеспечивает создание непрерывных, дискретных и гибридных систем на широком спектре компьютерных платформ, включая аппаратуру реального времени.

Мастерская реального времени может быть использована для быстрого создания программного обеспечения для встроенных контроллеров реального времени. Коды для систем обработки цифровых сигналов могут быть сгенерированы, кросс-компилированы и загружены в выбранный процессор. Мастерская реального времени поддерживает

DSP-платы и широкий диапазон изготовляемой и доступной для пользователей аппаратуры.

Современные программы численного моделирования систем и процессов становятся все более автоматизированными, облегчая пользователю процесс постановки и решения широкого класса сложных задач. Еще больший эффект дают современные возможности качественного визуального представления результатов.

7.2.3. ПРОГРАММНЫЙ КОМПЛЕКС ELECTRONICS WORKBENCH

Разработка радиоэлектронных устройств требует высокой точности и глубокого анализа. При разработке современного радиоэлектронного оборудования невозможно обойтись без компьютерных методов разработки, ввиду сложности и объемности выполняемых работ.

Electronics Workbench представляет собой программный продукт, позволяющий производить моделирование, тестирование, разработку и отладку электрических цепей.

Для работы данного программного комплекса необходим IBM-совместимый компьютер с процессором I486 и выше.

Electronics Workbench имеет достаточно простой интерфейс пользователя и прост в обращении.

Electronics Workbench содержит в себе достаточно большое количество моделей радиоэлектронных устройств, а также позволяет создавать пользователю свои модели.

В программном комплексе предусмотрена работа не только с «идеальными» элементами, но и с «реальными». Есть возможность имитации различного вида шумов и помех, что позволяет разработчику максимально приблизить модель к реальной.

Также Electronics Workbench позволяет проводить анализы электрических цепей, выполнение которых при стандартном подходе является достаточно трудоемким процессом.

Интерфейс программного комплекса Electronics Workbench

Приложение Electronics Workbench представляет собой средство программной разработки и имитации электрических цепей.

Интерфейс пользователя состоит из полоски меню, панели инструментов и рабочей области.

Полоса меню состоит из следующих компонент: меню работы с файлами (File), меню редактирования (Edit), меню работы с цепями (Circut), меню анализа схем (Analysis), меню работы с окнами (Window), меню работы с файлами справок (Help).

Панель инструментов состоит из «быстрых кнопок», имеющих аналоги в меню, кнопок запуска и приостановки схем, набора радиоэлектронных аналоговых и цифровых деталей, индикаторов, элементов управления и инструментов.

Технология подготовки схем

Прежде чем создавать чертеж принципиальной схемы средствами программы EWB, необходимо на листе бумаги подготовить ее эскиз с примерным расположением компонентов и с учетом возможности оформления отдельных фрагментов в виде подсхем. Целесообразно также ознакомиться с библиотекой готовых схем программы для выбора аналога (прототипа) или использования имеющихся решений в качестве подсхем.

В общем случае процесс создания схемы начинается с размещения на рабочем поле EWB компонентов из библиотек программы в соответствии с подготовленным эскизом. Разделы библиотеки программы EBW 5.4 поочередно могут быть вызваны с помощью меню Windows или с помощью иконок, расположенных под линейкой контрольно-измерительных приборов. Каталог выбранной библиотеки располагается в вертикальном окне справа или слева от рабочего поля (устанавливается в любое место перетаскиванием стандартным способом – за шапку заголовка).

Для открытия каталога нужной библиотеки необходимо подвести курсор мыши к соответствующей иконке и нажать один раз ее левую кнопку. Необходимый для создания схемы значок (символ) компонента переносится из каталога на рабочее поле программы движением мыши при нажатой левой кнопке, после чего кнопка отпускается (для фиксирования символа) и производится двойной щелчок по значку компонента. В раскрывающемся диалоговом окне устанавливаются требуемые параметры (сопротивление резистора, тип транзистора и т.д.) и выбор подтверждается нажатием кнопки Accept или клавиши Enter. На этом этапе необходимо предусмотреть место для размещения контрольных точек и иконок контрольно-измерительных приборов.

Если в схеме используются компоненты одинакового номинала (например, резисторы с одинаковым сопротивлением), то номинал такого компонента рекомендуется задать непосредственно в каталоге библиотеки и затем переносить компоненты в нужном количестве на рабочее поле. Для изменения номинала компонента необходимо два раза щелкнуть мышью по символу его графического изображения и в раскрывающемся после этого окне внести изменения.

После размещения компонентов производится соединение их выводов проводниками. При этом необходимо учитывать, что к выводу компонента можно подключить только один проводник. Для выполнения подключения курсор мыши подводится к выводу компонента и после появления прямоугольной площадки синего цвета нажимается левая кнопка и появляющийся при этом проводник протягивается к выводу другого компонента до появления на нем такой же прямоугольной площадки, после чего кнопка мыши отпускается, и соединение готово. При необходимости подключения к этим выводам других проводников в библиотеке Passive выбирается точка (символ соединения) и переносится на ранее установленный проводник. Чтобы точка почернела (первоначально она имеет красный цвет), необходимо щелкнуть мышью по свободному месту рабочего поля. Если эта точка действительно имеет электрическое соединение с проводником, то она полностью окрашивается черным цветом. Если на ней виден след от пересекающего проводника, то электрического соединения нет и точку необходимо установить заново. После удачной установки к точке соединения можно подключить еще два проводника.

Если соединение нужно разорвать, курсор подводится к одному из выводов компонентов или точке соединения и при появлении площадки нажимается левая кнопка, проводник отводится на свободное место рабочего поля, после чего кнопка отпускается. Если необходимо подключить вывод к имеющемуся на схеме проводнику, то проводник от вывода компонента курсором подводится к указанному проводнику и после появления точки соединения кнопка мыши отпускается.

Следует отметить, что прокладка соединительных проводников производится автоматически, причем препятствия – компоненты и другие проводники – огибаются по ортогональным направлениям (по горизонтали или вертикали).

Точка соединения может быть использована не только для подключения проводников, но и для введения надписей (например, указания величины тока в проводнике, его функционального назначения и т.п.). Для этого необходимо дважды щелкнуть по точке и в раскрывшемся окне ввести необходимую запись (не более 14 символов), причем запись можно смещать вправо путем введения слева нужного количества пробелов. Это свойство может быть использовано и в том случае, когда позиционное обозначение компонента (например Cl, R10) накладывается на рядом проходящий проводник или другие элементы схемы.

Если необходимо переместить отдельный сегмент проводника, к нему подводится курсор, нажимается левая кнопка и после появления в вертикальной или горизонтальной плоскости двойного курсора производятся нужные перемещения. Подключение к схеме контрольно-измерительных приборов производится аналогично. Причем для таких приборов, как осциллограф или логический анализатор, соединения целесообразно проводить цветными проводниками, поскольку их цвет определяет цвет соответствующей осциллограммы. Цветные проводники целесообразны не только для обозначения проводников одинакового функционального назначения, но и для проводников, находящихся в разных частях схемы (например, проводники шины данных до и после буферного элемента). Примеры такого оформления можно найти в каталогах готовых схем.

При обозначении компонентов необходимо придерживаться рекомендаций и правил, предусмотренных ЕСКД (единой системой конструкторской документации). Что касается пассивных компонентов, то при выборе их обозначений особых трудностей не возникает. Трудности возникают при выборе активных элементов – микросхем, транзисторов и т.п., особенно при необходимости использования компонентов отечественного производства, когда требуется установить точное соответствие функциональных обозначений выводов и параметров зарубежных и отечественных компонентов. Для облегчения этой задачи можно воспользоваться таблицами соответствия зарубежных и отечественных компонентов.

При импортировании в создаваемую схему другой схемы или ее фрагментов целесообразно действовать следующим образом:

- командой **File>Save As** записать в файл создаваемую схему, указав его имя в диалоговом окне (расширение имени файла указывать не обязательно, программа сделает это автоматически);
- командой **File>0pen** загрузить на рабочее поле импортируемую схему стандартным для Windows образом (некоторые особенности описаны в конце главы);

- командой **Edit> Select All** выделить схему, если импортируется вся схема, или выделить ее нужную часть и командой **Edit>Copy** скопировать выделенную схему в буфер обмена;
- командой **File>0pen** загрузить создаваемую схему;
- командой **Edit>Paste** вставить содержимое буфера обмена на рабочее поле; после вставки импортируемая схема будет выделена (и отмечена красным цветом) и может оказаться наложенной на создаваемую схему;
- клавишами управления курсором или мышью отбуксируйте импортированную часть в нужное место, после чего можно отменить выделение;
- после подключения импортированной схемы необходимо щелчками мыши пройтись по всем ее компонентам, чтобы исключить их смещения, возникающие при буксировке и приводящие к ступенчатым искажениям проводников.

Перемещения отдельных фрагментов схемы при ее компоновке выполняются вышеописанным образом после выделения фрагмента. После подготовки схемы рекомендуется составить ее описание (окно-ярлык вызывается из меню **Window>Description)**, в котором указывается ее назначение; после проведения моделирования указываются его результаты. К сожалению, программа EWB позволяет вводить описание только на английском языке. Кроме того, в EWB не предусмотрены средства для редактирования графических изображений компонентов, а также введения новых шрифтов.

Перейдем теперь к краткому обзору библиотечных компонентов программы EWB. При описании библиотек после названия компонента в скобках указываются назначаемые пользователем параметры. Например, для конденсатора это емкость, значение которой может быть установлено с помощью диалогового окна, а также температурные коэффициенты и разбросы (для EWB 5.4), для операционного усилителя – тип.

Группы компонентов программного комплекса electronics workbench

Группа **Custom** – вспомогательные компоненты. В разделе Custom программы размещаются подсхемы, если они имеются в данной схеме (в исходном состоянии раздел пуст), а также все библиотечные компоненты предыдущей версии EWB 3.0 в случае импорта из этой версии схемных файлов. В EWB 5.0 этот раздел называется **Favorites**. Заполнение раздела моделями компонентов или подсхем осуществляется программой автоматически одновременно с загрузкой схемного файла и очищается после окончания работы с ним.

Группа **Passive** – пассивные компоненты **–** содержит пассивные компоненты. Точка соединения проводников (●) используется также для введения на схему надписей длиной не более 14 символов (других способов введения текста в EWB не существует). Например, если на схеме требуется указать значение тока в какой-либо ветви, то на проводнике этой ветви ставится точка, затем двойным щелчком по точке вызывается диалоговое окно, в котором и выполняется соответствующая надпись.

Группа **Active** – активные компоненты. Она содержит полупроводниковые диоды, биполярные транзисторы, операционные усилители, аналоговые делительное и множительное устройства, а также линии связи.

Группа **FET** – полевые транзисторы.

Группа **Control** – коммутационные устройства и управляемые источники.

Группа **Hybrid** – гибридные компоненты. Для компонентов библиотеки Hybrid, за исключением таймера, допускается редактирование в диалоговом окне следующих параметров:

– верхний уровень входного напряжения (High-Level Input Voltage Vih, В);

– нижний уровень входного напряжения (Low-Level Input Voltage Vil, В);

– время установления при переходе от нижнего уровня к верхнему и наоборот (Propagation Delay Time, Low-to-High Level Output Tplh; Propagation Delay Time, High-to-Low Level Output Tphi, c);

– пороговое напряжение (Threshold Voltage Vth, В).

Например, для АЦП (ADC) первые два параметра отражают диапазон преобразуемых напряжений, третий – время преобразования, четвертый – цену младшего разряда.

Группа **Indie** – индикаторные приборы – содержит амперметр и вольтметр с цифровым отсчетом, одиночные и многосегментные светоиндикаторы, 8-разрядное устройство записи данных и звуковой сигнализатор.

Группа **Gates** – логические элементы – состоит из моделей базовых логических элементов и моделей цифровых ИМС ТТЛ- и КМОП-серий.

Группа **Comb'I** – комбинированные цифровые компоненты.

Группа **Seg'I** – триггеры.

Группа **IC** – цифровые микросхемы. В группе IC собраны модели цифровых ИМС серий SN74 и CD4000 (отечественные ИМС серий 155 и 176 соответственно). Для конкретных ИМС вместо символов хх ставятся соответствующие номера, например, SN7407 – 6 буферных элементов с открытым коллектором, CD4081 – 4 элемента 2И и т.д.

Контрольно-измерительные приборы. Панель контрольно-измерительных приборов находится под полем меню рабочего окна программы EWB и содержит цифровой мультиметр, функциональный генератор, двухканальный осциллограф, измеритель амплитудно-частотных и фазо-частотных характеристик, генератор слов (кодовый генератор), 8-канальный логический анализатор и логический преобразователь. Общий порядок работы с приборами такой: иконка прибора курсором переносится на рабочее поле и подключается проводниками к исследуемой схеме. Для приведения прибора в рабочее (развернутое) состояние необходимо дважды щелкнуть курсором по его иконке.

7.3. ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ ПРОЕКТИРОВАНИЯ УЗЛОВ, УСТРОЙСТВ И СИСТЕМ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

В настоящей главе рассмотрены примеры синтеза и анализа узлов и систем автоматического управления с помощью рассмотренных ранее программных средств автоматизации проектирования.

7.3.1. ПАРАМЕТРИЧЕСКИЙ СИНТЕЗ УЗЛОВ СИСТЕМ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ НА ПРИМЕРЕ РАСЧЕТА УСИЛИТЕЛЯ ЗВУКОВОЙ ЧАСТОТЫ

Исходные данные для выполнения расчета:

– принципиальная схема однокаскадного усилителя звуковой частоты представлена на рис. 7.2;

– полоса рабочих частот УЗЧ $f_{\text{н}} = 200\,\text{Гц}$, $f_{\text{в}} = 12\,\text{кГц}$;

– допустимые коэффициенты частотных искажений $M_{\text{в}} \leq 1{,}08$; $M_{\text{н}} \leq 1{,}2$;

– максимальная температура окружающей среды: +40 °C;

– напряжение источника питания $E_{\text{к}}$ = 10 В;

– амплитуда переменных тока и напряжения на нагрузке $i_{\text{н}\sim} = 12$ мА; $u_{\text{н}\sim} = 0{,}3$ В;

– сопротивление нагрузки $R_{\text{н}} = 300$ Ом, емкость нагрузки $C_{\text{н}} = 50\,n\Phi$;

– режим усиления класса А.

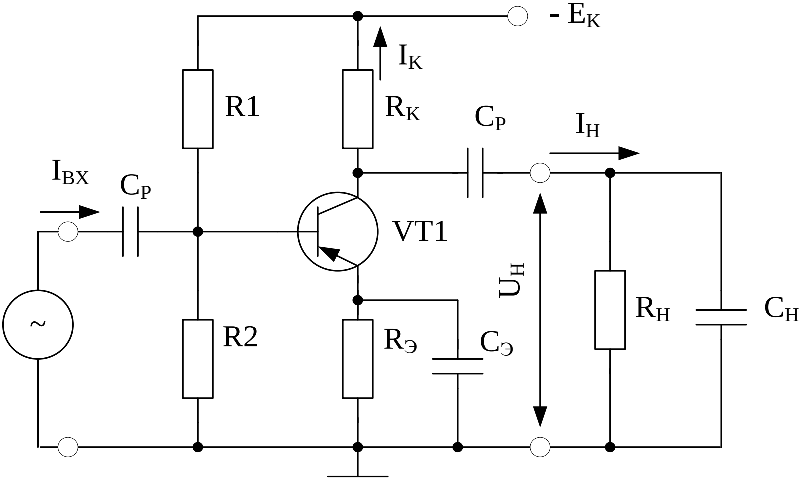


Рис. 7.2. Усилитель звуковой частоты

Рис. 2.2. Схема усилителя звуковой частоты

Необходимо:

– выбрать транзистор;

– рассчитать величины сопротивлений резисторов $R_{\text{к}}$ и $R_{\text{э}}$;

– рассчитать величины сопротивлений резисторов R_1 и R_2 делителя напряжения смещения;

– рассчитать емкости конденсаторов $C_{\text{р}}$ и $C_{\text{э}}$.

Расчет параметров схемы усилителя.

Выбор транзистора.

Транзистор выбирается по следующим показателям:

– по электрической прочности $U_{\text{кэ max}} > E_{\text{к}}$;

– по предельной частоте усиления $f_{\alpha} > 3 f_{\beta} \cdot h_{21\text{э ср}}$;

– по рабочей температуре $t^0_{TMMAK} > t^0_{CMMAK}$,

где t^0_{TMMAK} – максимальная рабочая температура транзистора; $t^0_{C\ MAKC}$ – максимальная температура окружающей среды.

Используя справочные данные [4, 7], выбирается транзистор, параметры которого записываются в табл. 7.1.

7.1. Параметры транзистора

Тип транзистора	$U_{\text{кэ max}}$, В	$I_{\text{к max}}$, мА	$I_{\text{кбо}}$, мкА	$h_{21\text{э}}$	$C_{\text{к}}$, пФ	$f\alpha$, МГц
КТ363А	12	30	30	80…300	≤2	≥1000

Определяются величины сопротивлений резисторов $R_{\text{к}}$ и $R_{\text{э}}$. Для этого:

– находится переменная составляющая коллекторного тока. Из рисунка 7.2 видно, что

$$i_{R_{K\sim}} = \frac{u_{\text{н}\sim}}{R_{Kop}},$$

а ориентировочное значение сопротивления будет рассчитываться по выражению

$$R_{Kop} = \frac{(0{,}3...0{,}6)E_K}{1{,}5 i_{\text{н}\sim}} = \frac{0{,}45 \cdot 10}{1{,}5 \cdot 12 \cdot 10^{-3}} = 250 \text{ Ом}.$$

Тогда

$$i_{R_{K\sim}} = \frac{0{,}3}{250} = 1{,}2 \text{ мА}; \; i_{K\sim} = 1{,}2 + 12 = 13{,}2 \text{ мА};$$

– определяется минимальный расчетный коллекторный ток покоя $I_{\text{кп}}$:

$$I_{\text{кп}} = (1{,}05...1{,}2) \cdot i_{K\sim} = (1{,}05...1{,}2) \cdot 13{,}2 = 1{,}1 \cdot 13{,}2 = 14{,}5 \text{ мА};$$

– определяется величина сопротивления R_K:

$$R_K = \frac{(0{,}3...0{,}6) \cdot E_K}{I_{\text{кп}}} = \frac{(0{,}3...0{,}6) \cdot 10}{14{,}5 \cdot 10^{-3}} = 310 \text{ Ом}.$$

Номинальное значение R_K выбирается из стандартного ряда сопротивлений $R_K = 330$ Ом;

– определяется величина сопротивления $R_э$:

$$R_э = \frac{(0{,}1\ldots0{,}3)E_K}{I_{кп}} = \frac{(0{,}1...0{,}3)\cdot 10}{14{,}5\cdot 10^{-3}} = 138 \text{ Ом}$$

и выбирается номинальное значение $R_э$ из стандартного ряда: $R_э = 130$ Ом.

Определение параметров режима покоя усилителя:

– расчет напряжения между коллектором и эмиттером $U_{кэп}$ для начальной рабочей точки (покоя):

$$U_{кэп} = E_к – I_{кп}\,(R_к + R_э) = 10 – 14{,}5\cdot 10 – 3\cdot(330 + 130) = 3{,}33 \text{ В}$$

(незначительным падением напряжения на $R_э$ от тока базы покоя пренебрегают);

– по входным и выходным характеристикам выбранного транзистора КТ363А [4, 7] для определенных значений $U_{кэп} = 3{,}33$ В и $I_{кп} = 14$ мА определяются значения $U_{бэп}$ и $I_{бп}$: $I_{бп} = 0{,}37$ мА, $U_{бэп} = 0{,}88$ В.

Определение величины сопротивлений делителя напряжения смещения R_1 и R_2:

– сопротивление $R_2 = (5...15)\cdot h_{11э}$, поэтому для найденной точки покоя на входных статических характеристиках определяется значение $h_{11э}$

$$h_{11э} = \left.\frac{\Delta\, U_{бэ}}{\Delta\; I_б}\right|_{U_{кэ}=\text{const}} = \frac{0{,}9 - 0{,}86}{(0{,}41 - 0{,}3)\cdot 10^{-3}} = 360 \text{ Ом},$$

а затем определяется сопротивление R_2

$$R_2 = (5...15)\cdot 360 = 10\cdot 360 = 3600 \text{ Ом}.$$

Номинальное значение R_2 выбирается из стандартного ряда (см. [4]): $R_2 = 3{,}6$ кОм;

– для определения R_1 вначале находится падение напряжения на R_2:

$$U_{R_2} = U_{\text{бэп}} + U_{R_э} = U_{\text{бэп}} + R_э(I_{\text{кп}} + I_{\text{бп}}) = 2{,}8 \text{ В},$$

откуда

$$I_{\partial} = \frac{U_{R_2}}{R_2} = \frac{2{,}8}{3{,}6\cdot 10^3} = 0{,}78 \text{ мА},$$

$$R_1 = \frac{E_K - U_{R_2}}{I_{\partial}} = \frac{10 - 2{,}8}{0{,}78\cdot 10^{-3}} = 9{,}17 \text{ кОм}.$$

Из стандартного ряда выбирается ближайший номинал $R_1 = 9{,}1$ кОм.

Определение емкости конденсаторов $C_{\text{р}}$ и $C_э$:

– вначале определяется эквивалентное сопротивление R_0

$$\frac{1}{R_0} = \frac{1}{h_{11э} + R_э} + \frac{1}{R_2},$$

откуда

$$R_0 = \frac{R_2(h_{11э} + R_э)}{R_э + h_{11э} + R_э} = \frac{3600\cdot(360+130)}{3600+360+130} = 430 \text{ Ом},$$

$$C_{\text{р}} = \frac{0{,}159}{100\cdot(330+430)\sqrt{1{,}06^2 - 1}} = 5{,}98 \text{ мкФ}.$$

Из стандартного ряда выбирается $C_{\text{р}}$:

$$C_{\text{р}} = 5{,}6 \text{ мкФ};$$

– для определения $C_{\text{э}}$ вначале определяется коэффициент $\alpha\,(h_{21\text{б}})$ и $h_{11\text{б}}$:

$$h_{21\text{б}} = \frac{h_{21\text{э}}}{1+h_{21\text{э}}} = \frac{190}{191} = 0{,}995,$$
$$h_{11\text{б}} = h_{11\text{э}}(1-h_{21\text{б}}) = 350\cdot 0{,}005 = 1{,}8 \text{ Ом},$$

с учетом которых и определяется $C_{\text{э}}$:

$$C_{\text{э}} = \frac{0{,}159}{100\cdot\left[1{,}8+(1-0{,}995)\cdot 300\right]\cdot\sqrt{1{,}06^2-1}} \approx 138\cdot 10^{-6}\,\text{Ф}.$$

Из стандартного ряда выбирается $C_{\text{э}} = 130$ мкФ.

После того как закончен параметрический синтез, необходимо убедиться в адекватности полученных результатов, то есть проанализировать работу устройства путем имитационного моделирования. Это можно сделать в среде любой доступной программы, например моделирование электронных устройств можно проводить в среде Electronics Workbench.

На рисунке 7.3. представлена копия экрана рабочего окна программы Electronics Workbench при моделировании в реальном масштабе времени синтезированного устройства.

Необходимо отметить, что на представленном рисунке выполнено графическое комбинирование – одновременно прибор, предназначенный для получения АЧХ и ФЧХ, представлен в виде трех приборов: на экране Bode Plotter 1 показана фазочастотная характеристика, а на экранах 2 и 3 показаны амплитудно-частотные характеристики. Причем на Bode Plotter 2 показана АЧХ с чертой на начале полосы пропускания для уровня 0,707 и частоты $f_н = 200$ Гц, а на Bode Plotter 3 показана АЧХ с чертой на окончании полосы пропускания для уровня 0,707 и частоты $f_в = 12$ кГц.

Из анализа результатов моделирования (осциллограмм входного и выходного сигналов АЧХ и ФЧХ) видно, что спроектированное устройство по своим параметрам соответствует поставленным требованиям.

7.3.2. АНАЛИЗ СИСТЕМ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

На рисунке 7.3 представлена структурная схема замкнутой системы автоматического управления. Параметры передаточных функций звеньев системы заданы в табл. 7.2.

На рисунке 7.4 введены обозначения: $x(t)$ – полезное задающее воздействие; $\varepsilon(t)$ – ошибка системы; $y(t)$ – выходное воздействие.

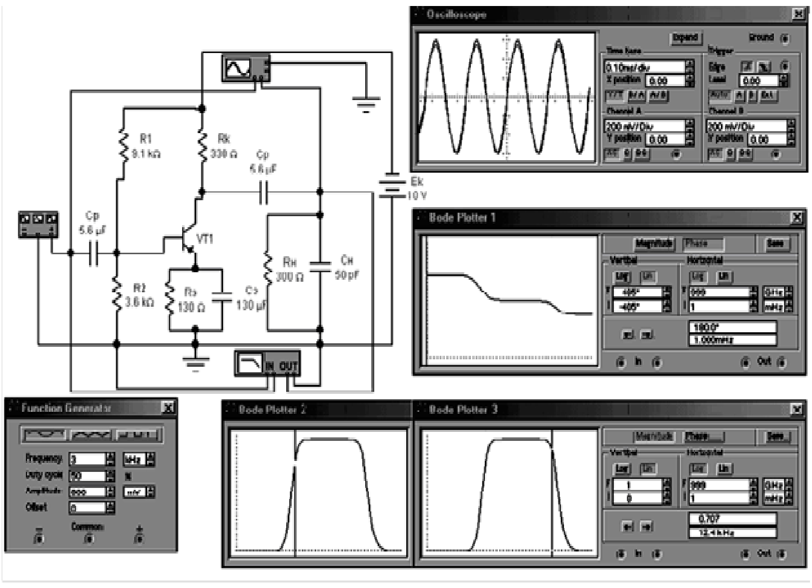


Рис. 7.3. Модель автоматического управления УЗЧ

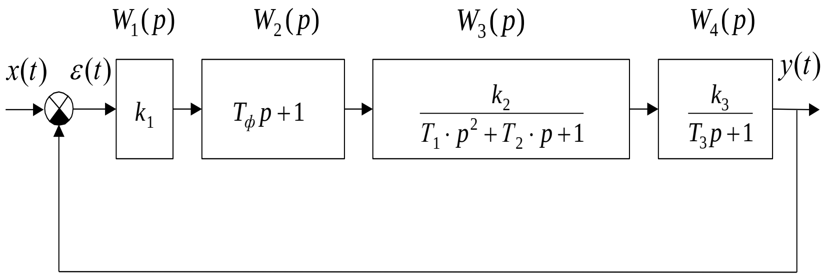


Рис. 7.4. Структурная схема системы автоматического управления

Необходимо:

1. Произвести расчет ошибки системы в динамическом и установившемся режимах для различных входных сигналов $x(t)$.
2. Исследовать устойчивость системы по критериям Гурвица, Ляпунова, Михайлова, Найквиста. Определить запасы устойчивости.
3. Построить график переходного процесса, провести анализ его качества.
4. Сделать выводы по качеству САУ.

7.2. Параметры передаточных функций

k_1	$T_Ф$	k_2	T_1	T_2	k_3	T_3	$x(t)$
3	0,11	3	0,18	0,9	1	1	1,3; 1,3t

Задание вида и параметров передаточных функций звеньев САУ в среде Mathcad:

$$W_1(p) := k_1,$$

$$W_2(p) := T_ф \cdot p + 1,$$

$$W_3(p) := \frac{k_2}{T_1 \cdot p^2 + T_1 \cdot p + 1},$$

$$W_4(p) := \frac{k_3}{T_3 \cdot p + 1},$$

$$k_1 \equiv 3, \quad k_2 \equiv 3, \quad k_3 \equiv 1,$$

$$T_ф \equiv 0{,}11, \quad T_1 \equiv 0{,}18, \quad T_2 \equiv 0{,}9, \quad T_3 \equiv 1.$$

Определение передаточных функций САУ и их упрощение:

– разомкнутой САУ

$$W_р(p) := W_1(p) \cdot W_2(p) \cdot W_3(p) \cdot W_4(p),$$

$$W_р(p) \text{ simplify} \rightarrow 3 \cdot \frac{(0{,}33 \cdot p + 3)}{(0{,}18 \cdot p^3 + 1{,}08 \cdot p^2 + 1{,}9 \cdot p + 1)};$$

– замкнутой САУ

$$\Phi_{xy}(p) := \frac{W_р(p)}{1 + W_р(p)},$$

$$\Phi_{xy}(p) \text{ simplify} \rightarrow \frac{(9{,}9 \cdot 10^{-2} \cdot p + 0{,}9)}{(1{,}8 \cdot 10^{-2} \cdot p^3 + 0{,}108 \cdot p^2 + 0{,}289 \cdot p + 1)};$$

– по ошибке от задающего воздействия $x(t)$

$$\Phi_{x\varepsilon}(p) := 1 - \Phi_{xy}(p),$$

$$\Phi_{x\varepsilon}(p) \text{ simplify} \rightarrow \frac{(1{,}8 \cdot 10^{-2} \cdot p^3 + 0{,}108 \cdot p^2 + 0{,}19 \cdot p + .1)}{(1{,}8 \cdot 10^{-2} \cdot p^3 + 0{,}108 \cdot p^2 + 0{,}289 \cdot p + 1)}.$$

Для упрощения передаточных функций необходимо воспользоваться оператором символьных преобразований simplify панели инструментов Математика/Symbolic.

Расчет динамических ошибок САУ

Если на систему действует задающее воздействие, ошибка системы может быть рассчитана, используя коэффициенты ошибок:

$$\varepsilon(t) = D_0 \cdot x(t) + D_1 \cdot \frac{d}{dt} \cdot x(t) + D_2 \cdot \frac{d^2}{dt^2} \cdot x(t) + \ldots,$$

где $x(t)$ – задающее воздействие; D_0, D_1, D_2 – коэффициенты ошибки по положению, скорости, ускорению соответственно. Эти коэффициенты могут быть вычислены по известным выражениям [6]. Для статической системы с ПФ по ошибке $D(p) = \dfrac{A(p)}{C(p)} = 1 - \Phi_{xy}(p)$ имеют место следующие соотношения:

$$D_0 := \frac{a_0}{c_0},\ D_1 := \frac{1}{c_0}(a_1 - D_0 \cdot c_1),\ D_2 := \frac{1}{c_0}(a_2 - D_1 \cdot c_1 - D_0 \cdot c_2),$$

где a_i – коэффициенты полинома числителя ПФ САУ по ошибке; c_i – коэффициенты полинома знаменателя ПФ САУ по ошибке.

Скопируем знаменатель ПФ по ошибке $\Phi_{xy}(p)$ в выражение $C(p)$:

$$C(p) := 1{,}8 \cdot 10^{-2} \cdot p^3 + 0{,}108 \cdot p^2 + 0{,}289 \cdot p + 1,$$

а числитель ПФ по ошибке $\Phi_{xy}(p)$ в выражение $A(p)$:

$$A(p) := 1{,}8 \cdot 10^{-2} \cdot p^3 + 0{,}108 \cdot p^2 + 0{,}19 \cdot p + 0{,}1.$$

Так как числитель и знаменатель ПФ по ошибке представлены в виде полиномов, легко получить матрицы-столбцы коэффициентов полиномов a и c:

$$A(p)\ \ coeffs\ ,\ p\ \rightarrow \begin{pmatrix} 0{,}1 \\ 0{,}19 \\ 0{,}108 \\ 1{,}8\cdot 10^{-2} \end{pmatrix} \qquad a\ := \begin{pmatrix} 0{,}1 \\ 0{,}19 \\ 0{,}108 \\ 1{,}8\cdot 10^{-2} \end{pmatrix},$$

$$C(p)\ \ coeffs\ ,\ p\ \rightarrow\ 0, \begin{pmatrix} 0{,}1 \\ 0{,}289 \\ 0{,}108 \\ 1{,}8\cdot 10^{-2} \end{pmatrix} \qquad c\ := \begin{pmatrix} 0{,}1 \\ 0{,}289 \\ 0{,}108 \\ 1{,}8\cdot 10^{-2} \end{pmatrix}.$$

При расчете коэффициентов ошибок D_i ограничимся коэффициентами ошибок по положению D_0 и по скорости D_1: D = 0,1, D = 0,161.

Соответственно, при заданном постоянном воздействии $x(t) := 1{,}3$ ошибка будет равна

$$\varepsilon\,(t) := D_0\cdot x(t) + D_1\cdot \frac{d}{dt}x(t)\ ,$$

$$\varepsilon\,(t) = 0{,}13\,,$$

$$\lim_{t\to\infty}\varepsilon\,(t)\ \rightarrow 0{,}13,$$

а при воздействии $x(t) = 1{,}3\cdot t$, где t примем равным, например, 100 с:

$$\varepsilon\,(t) := D_0\cdot x(t) + D_1\cdot \frac{d}{dt}x(t)\,,$$

$$\varepsilon\,(t) = 13{,}209\,,$$

$$\lim_{t\to\infty}\varepsilon(t) \rightarrow \infty.$$

Исследование устойчивости САУ с помощью критерия Гурвица

Формулировка критерия: если характеристическое уравнение системы автоматического управления (САУ) n-й степени имеет вид:

$$C(p) = c_n p^n + c_{n-1}p^{n-1} + \ldots + c_1 p + c_0 = 0,$$

то для устойчивости системы необходимо и достаточно, чтобы при всех $a_i > 0$ все n определителей Гурвица $\Delta_1, \Delta_2, \ldots, \Delta_n$, составленных по определенной схеме, были положительны.

Характеристическое уравнение составляется приравниванием знаменателя передаточной функции (ПФ) нулю:

$$C(p) := 1{,}8 \cdot 10^{-2} \cdot p^3 + 0{,}108 \cdot p^2 + 0{,}289 \cdot p + 1.$$

Вычислим определители, составленные в соответствии с критерием Гурвица, используя функции панели Matrix (матричные вычисления):

$$\Delta 0 := c_2, \ |\Delta 0| = 0{,}108;$$

$$\Delta 1 := \begin{pmatrix} c_0 & 0 \\ c_2 & c_1 \end{pmatrix}, \ |\Delta 1| = 0{,}289;$$

$$\Delta 2 := \begin{pmatrix} c_0 & 0 & 0 \\ c_0 & c_1 & c_2 \\ 0 & c_3 & c_2 \end{pmatrix}, \ |\Delta 2| = 0{,}029.$$

Анализ устойчивости по критерию Гурвица показывает, что замкнутая САУ устойчива, так как:

1) старший коэффициент характеристического уравнения положителен;

2) все определители Гурвица положительны.

Исследование устойчивости САУ с помощью критерия Михайлова

Формулировка критерия: для устойчивой системы необходимо и достаточно, чтобы ее характеристический вектор (годограф) при изменении частоты от нуля до бесконечности повернулся в положительном направлении (против часовой стрелки), начиная с положительной вещественной полуоси, на число квадрантов, равное порядку характеристического уравнения.

Зададим характеристический полином замкнутой САУ путем копирования знаменателя результата символьного преобразования после упрощения:

$$C(p) := 1{,}8 \cdot 10^{-2} \cdot p^3 + 0{,}108 \cdot p^2 + 0{,}289 \cdot p + 1.$$

Зададим комплексную переменную *j*, индекс массива *k*, шаг и пределы изменения частоты ω:

$$j := \sqrt{-1};\ \ k := 0..10000;\ \ \omega_k := 2 \cdot \pi \cdot k \cdot 0{,}01;\ \ p_k := \omega_k \cdot j.$$

Затем построим годограф Михайлова САУ. Для построения годографа воспользуемся встроенным графиком «точка *X*-*Y*». По оси *X* отложим выделенное с помощью функции Re значение вещественной части ПФ, а по оси *Y* – выделенное с помощью функции Im значение мнимой части ПФ.

Анализ устойчивости по критерию Михайлова (рис. 7.5) показывает, что замкнутая САУ устойчива, так как годограф замкнутой системы третьего порядка (судя по высшей степени переменной p характеристического полинома), начинаясь на вещественной полуоси, последовательно проходит три квадранта комплексной плоскости, нигде не обращаясь в нуль (не проходя через точку (0, *j*0)).

Исследование устойчивости САУ с помощью критерия Найквиста

Формулировка критерия для устойчивой разомкнутой системы: замкнутая система будет устойчива, если амплитудно-фазочастотная характеристика (АФЧХ) устойчивой разомкнутой системы при изменении частоты от нуля до бесконечности не охватывает точку с координатами [–1, *j*0].

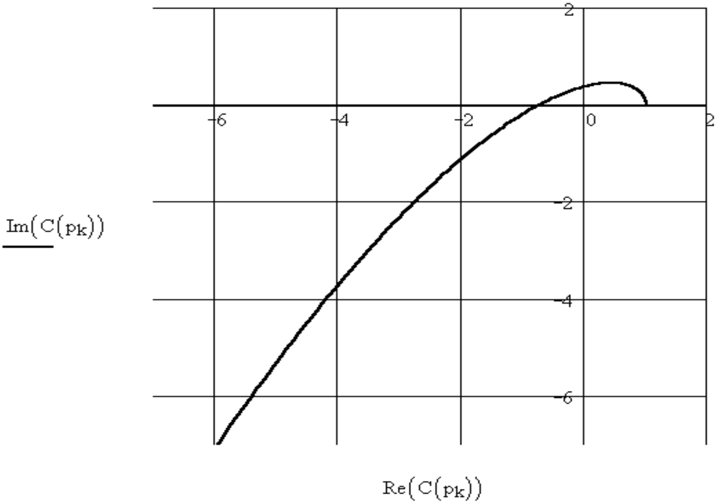


Рис. 7.5. Годограф Михайлова
Рис. 2.5. Годограф Михайлова

Зададим комплексную переменную *j*, индекс массива *k*, шаг и пределы изменения частоты ω, а с помощью функций синуса и косинуса – окружность единичного радиуса:

$$j := \sqrt{-1};\ k := 0..10000;\ \omega_k := 2\cdot\pi\cdot k\cdot 0{,}01;\ p_k := \omega_k \cdot j.$$

$$x(t) := \sin(t);\ y(t) := \cos(t).$$

Затем построим годограф (АФЧХ) разомкнутой САУ (рис. 7.6, *а*, *б*). Для построения годографа воспользуемся встроенным графиком «точечный *X*-*Y*». По оси *X* отложим выделенное с помощью функции Re значение вещественной части ПФ, а по оси *Y* – выделенное с помощью функции Im значение мнимой части ПФ.

Анализ устойчивости по критерию Найквиста показывает, что замкнутая САУ устойчива, так как годограф устойчивой разомкнутой системы при изменении частоты от нуля до бесконечности не охватывает точку с координатами (–1, *j*0).

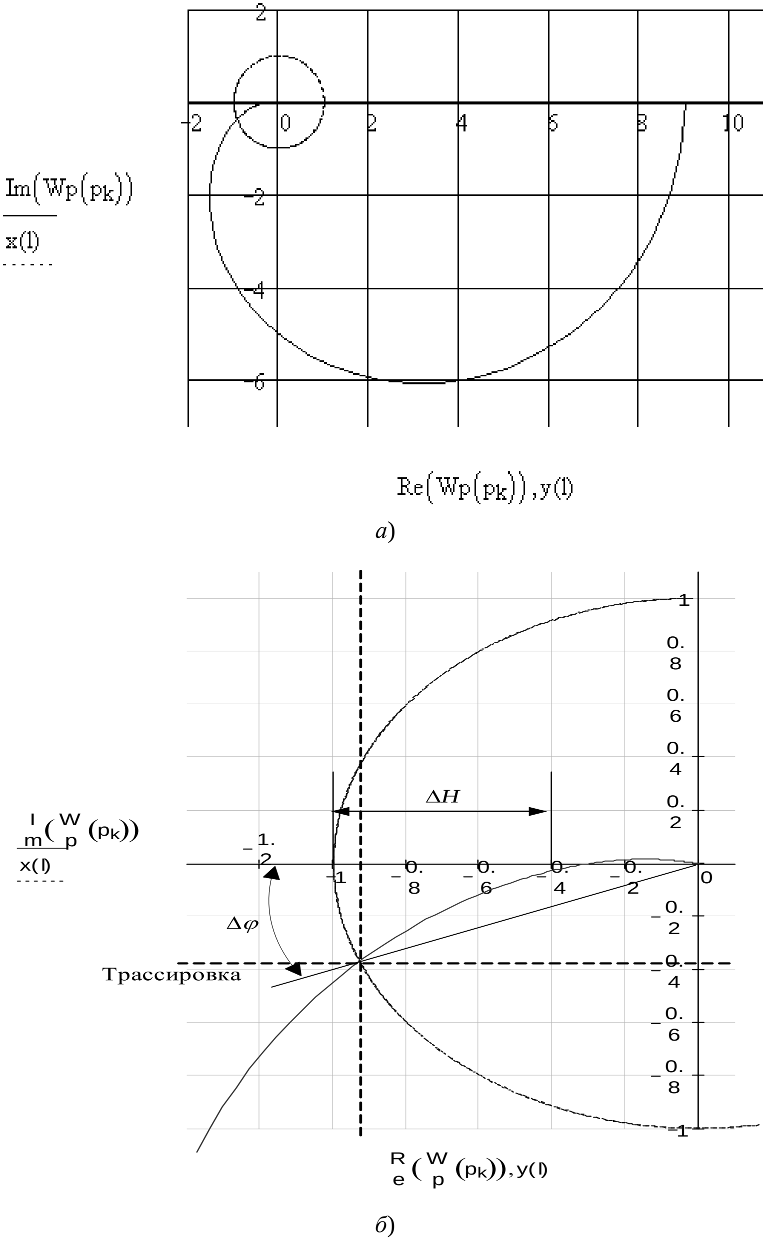


Рис. 7.6. Годограф разомкнутой САУ

Анализируя график, определяем запасы устойчивости САУ по амплитуде ΔH и по фазе $\Delta\varphi$.

Запас устойчивости по амплитуде равен расстоянию между точкой с координатами (–1, j0) и точкой пересечения годографа с вещественной отрицательной полуосью на отрезке [–1; 0]. Его численное значение $\Delta H = 0{,}72067$.

Запас устойчивости по фазе, равный углу между отрицательной вещественной полуосью и линией проведенной через начало координат и точку, где $W(j \cdot w) = 1$, определяется по координате $\mathrm{Re}(W(j \cdot w))$ по оси X в точке пересечения годографа $W(p)$ с окружностью единичного радиуса.

Его численное значение $\Delta\varphi := a\cos(0{,}93043)$, $\Delta\varphi := 0{,}375\ \mathrm{rad}$ (радиан) или $\Delta\varphi := 21{,}486\ \deg$ (градусов).

Для определения координат при определении запасов устойчивости использован инструмент «Трассировка», окно которого вызывается из контекстного меню графика, и функция арккосинуса.

Для подтверждения результатов анализа и получения характеристик исследуемой САУ в реальном масштабе времени можно провести моделирование САУ в среде программного пакета **MATLAB\Simulink**.

На рисунке 7.7 представлена схема моделирования.

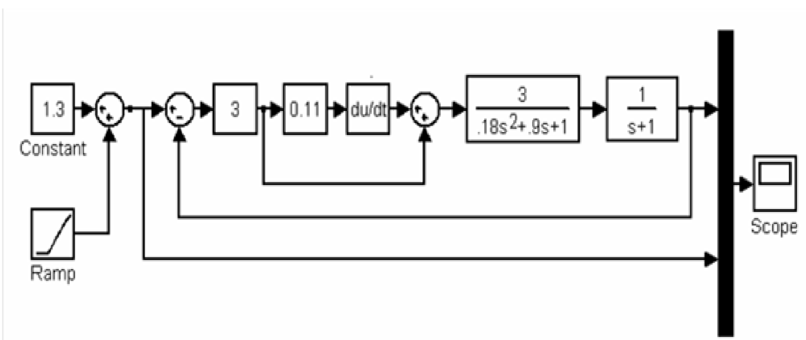
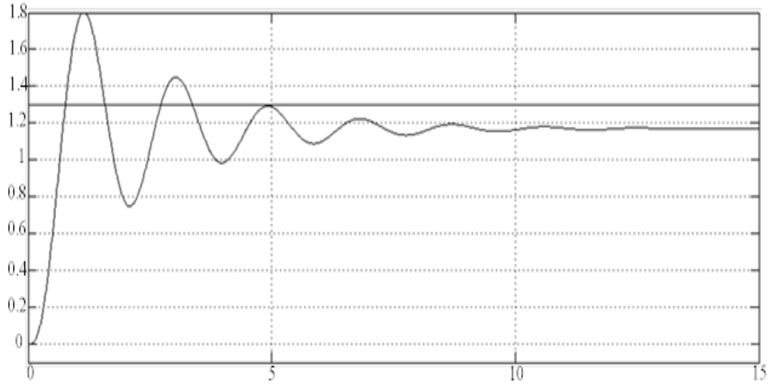
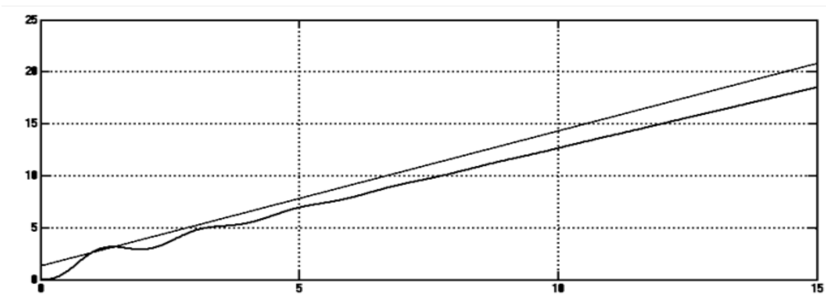


Рис. 7.7. Структурная схема САУ в Simulink



а)



б)

Рис. 7.8. Переходные процессы

Результаты моделирования при различных входных воздействиях представлены на рис. 7.8, *а*, *б*.

7.3.3. МОДЕЛИРОВАНИЕ И РАСЧЕТ ИНСТРУМЕНТАЛЬНЫХ ПОГРЕШНОСТЕЙ ТРЕХОСНОГО ГИРОСТАБИЛИЗАТОРА СИСТЕМЫ УПРАВЛЕНИЯ

В системах управления ракет нашли применение платформенные инерциальные навигационные системы (ИНС), основным элементом которых является трехосный гиростабилизатор [8]. Трехосный гиростабилизатор обеспечивает моделирование на борту инерциальной системы координат, и стабилизацию положения измерительных осей (осей чувствительности) триады акселерометров. Точность решения навигационной задачи системой управления во многом зависит от инструментальных погрешностей трехосного гиростабилизатора и триады акселерометров, установленных на его стабилизированной площадке.

В связи с этим расчет ошибок трехосного гиростабилизатора и триады акселерометров занимает важное место в задачах моделирования и обеспечения требуемых точностных характеристик системы управления.

Получим модели и расчетные соотношения ошибок трехосного гиростабилизатора и триады акселерометров в функции от собственных инструментальных погрешностей, режимов функционирования и параметров движения летального аппарата. Без потери общности за основу возьмем конструктивную схему трехосного гиростабилизатора (ТГС) с ортогональными схемами установки на стабилизированной площадке гироблоков и акселерометров.

Со стабилизированной площадкой (гироплатформой) ТГС свяжем приборный трехгранник (ТГ) $X_{\text{п}}Y_{\text{п}}Z_{\text{п}}$. Положение приборного ТГ в процессе работы ТГС будет отличаться от расчетного из-за ошибок стабилизации, обусловленными собственными инструментальными погрешностями ТГС. За расчетный (модельный, сопровождающий) ТГ примем трехгранник XYZ, описывающий ориентации материализуемой ТГС системы координат (в полете стартовой системы координат). Так как отклонение положения приборного ТГ относительно расчетного ТГ вызвано инструментальными погрешностями ТГС, то для прецизионных ТГС рассогласование между осями приборного и расчетного ТГ за все время работы будет составлять малые углы. Тогда ошибка в положении приборного ТГ относительно расчетного ТГ (ошибка ориентации) может быть описана вектором малого поворота (ВМП), который обозначим $\overline{\psi}(t)$.

Выставка ортогональной триады акселерометров на гироплатформе ТГС относительно приборного трехгранника из-за конечной точности выполнения технологических операций осуществляется с некоторыми ошибками. В результате оси чувствительности акселерометров оказываются отклоненными от соответствующих осей приборного ТГ на некоторые малые углы, называемые ошибками выставки акселерометров. Очевидно, что ошибки выставки оси чувствительности каждого акселерометра могут быть заданы двумя малыми углами, характеризующими отклонения оси чувствительности акселерометра относительно одноименной оси гироплатформы ТГС. Зададим ошибки выставки триады акселерометров шестью малыми углами следующим образом:

– для акселерометра X (1) малые углы – $\gamma_2^1\gamma_3^1$;

– для акселерометра Y (2) малые углы – $\gamma_1^2\gamma_3^2$;

– для акселерометра Z (3) малые углы – $\gamma_1^3\gamma_2^3$.

Здесь верхний индекс – номер акселерометра, а нижний индекс – номер оси гироплатформы, относительно которой осуществлен поворот на данный малый угол.

Примем, что вдоль осей расчетного ТГ действует вектор линейного ускорения $\overline{W}(t)=[W_X(t)W_Y(t)W_Z(t)]^T$, который подлежит измерению триадой акселерометров. Тогда в осях приборного трехгранника с учетом ошибки ориентации $\overline{\psi}(t)$ данный вектор имеет вид:

$$\begin{bmatrix} W_X(t) \\ W_Y(t) \\ W_Z(t) \end{bmatrix}_{\text{п}} = \begin{bmatrix} 1 & \psi_Z(t) & -\psi_Y(t) \\ -\psi_Z(t) & 1 & \psi_X(t) \\ \psi_Y(t) & -\psi_X(t) & 1 \end{bmatrix} \begin{bmatrix} W_X(t) \\ W_Y(t) \\ W_Z(t) \end{bmatrix}, \qquad (7.1)$$

где $\psi_X(t)$, $\psi_Y(t)$, $\psi_Z(t)$ – составляющие вектора малого поворота $\overline{\psi}(t)$.

Учитывая ошибки выставки акселерометров на гироплатформе ТГС аналогично, получим проекции вектора ускорений $\overline{W}(t)$ на оси чувствительности акселерометров:

$$\begin{bmatrix} W_X(t) \\ W_Y(t) \\ W_Z(t) \end{bmatrix}_{\text{а}} = \begin{bmatrix} 1 & \gamma_3^1 & -\gamma_2^1 \\ -\gamma_3^2 & 1 & \gamma_1^2 \\ \gamma_2^3 & -\gamma_1^3 & 1 \end{bmatrix} \begin{bmatrix} W_X(t) \\ W_Y(t) \\ W_Z(t) \end{bmatrix}_{\text{п}}. \tag{7.2}$$

Уравнения измерений триады акселерометров с учетом с собственных инструментальных погрешностей представим в векторно-матричном виде

$$\begin{bmatrix} J_x(t) \\ J_y(t) \\ J_z(t) \end{bmatrix} = \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + \begin{bmatrix} k_x + \Delta k_x & 0 & 0 \\ 0 & k_x + \Delta k_x & 0 \\ 0 & 0 & k_x + \Delta k_x \end{bmatrix} \begin{bmatrix} W_x(t) \\ W_y(t) \\ W_z(t) \end{bmatrix}_{\text{а}} + \begin{bmatrix} u_x(t) \\ u_y(t) \\ u_z(t) \end{bmatrix}, \tag{7.3}$$

где $\bar{J} = [J_X \quad J_Y \quad J_Z]^T$ – вектор выходных сигналов триады акселерометров; $\overline{\Delta} = [\Delta_X \quad \Delta_Y \quad \Delta_Z]^T$ – вектор смещений нулей триады акселерометров; $k_X, k_Y, k_Z, \Delta k_X, \Delta k_Y, \Delta k_Z$ – масштабные коэффициенты и отклонения масштабных коэффициентов соответствующих акселерометров; $\bar{u}(t) = [u_x(t) \quad u_y(t) \quad u_z(t)]$ – вектор измерительных шумов триады акселерометров. Используя выражения (7.1), (7.2) и пренебрегая членами второго порядка малости, после преобразования уравнения (7.3) получим уравнение измерений триады акселерометров в функции от измеряемого ускорения, инструментальных погрешностей триады акселерометров и ошибок ориентации приборного трехгранника:

$$\bar{J} = \overline{\Delta} + [k(E) + \hat{\gamma} + \Psi(t)) + E_k] \cdot \overline{W}(t) + u(t), \tag{7.4}$$

где $k = \begin{bmatrix} k_X & 0 & 0 \\ 0 & k_Y & 0 \\ 0 & 0 & k_Z \end{bmatrix} k = \begin{bmatrix} k_X & 0 & 0 \\ 0 & k_Y & 0 \\ 0 & 0 & k_Z \end{bmatrix}$ – матрица масштабных коэффициентов триады акселерометров; E – единичная матрица размера (3×3);

$\hat{\gamma} = \begin{bmatrix} 0 & \gamma_3^1 & -\gamma_2^1 \\ -\gamma_3^2 & 0 & \gamma_1^2 \\ \gamma_2^3 & -\gamma_1^3 & 0 \end{bmatrix}$ – матрица ошибок выставки триады акселерометров; $\hat{\psi}(t) = \begin{bmatrix} 0 & \psi_Z(t) & -\psi_Y(t) \\ -\psi_Z(t) & 0 & \psi_X(t) \\ \psi_Y(t) & -\psi_X(t) & 0 \end{bmatrix}$ – кососимметрическая матрица ошибок ориентации приборного трехгранника;

$E_k = \begin{bmatrix} \Delta k_X & 0 & 0 \\ 0 & \Delta k_Y & 0 \\ 0 & 0 & \Delta k_Z \end{bmatrix}$ – матрица отклонений в масштабных коэффициентах триады акселерометров.

В уравнение измерений (7.4) ошибки ориентации гироплатформы и инструментальные погрешности триады акселерометров входят как коэффициенты соответствующих матриц, что затрудняет их анализ, и использование при моделировании и расчетах современных методов теории динамических систем. Преобразуем уравнение измерений (7.4) к векторно-матричной форме в пространстве состояний, более наглядной и удобной для анализа влияния инструментальных погрешностей триады акселерометров и ошибок ориентации гироплатформы на результаты измерения вектора ускорения $\overline{W}(t)$ по траектории движения. С этой целью введем составной вектор $\begin{bmatrix} \overline{\psi}(t) \\ \overline{X}_A \end{bmatrix}$, в котором $\overline{\psi}(t)$ – вектор ошибок ориентации гироплатформы;

$$\overline{X}_A = \begin{bmatrix} \overline{\Delta} \\ \Delta\overline{k} \\ \overline{\gamma} \end{bmatrix}$$ – составной вектор инструментальных погрешностей триады акселерометров; $\Delta\overline{k} = [\Delta k_X \quad \Delta k_Y \quad \Delta k_Z]^T$, $[\Delta_X \quad \Delta_Y \quad \Delta_Z]^T$ – вектор смещений нулей триады акселерометров; – вектор отклонений в масштабных коэффициентах триады акселерометров; $\overline{\gamma} = \left[\gamma_2^1 \quad \gamma_3^1 \quad \gamma_1^2 \quad \gamma_3^2 \quad \gamma_1^3 \quad \gamma_2^3\right]^T$ – вектор ошибок выставки триады акселерометров на гироплатформе. Для выделения составляющих векторов $\overline{\psi}(t)$ и $\overline{X}_A$ проведем перестановочные преобразования уравнения измерений (7.4). В результате запишем данное уравнение измерений триады акселерометров в следующей векторно-матричной форме:

$$\overline{J}(t) = k \cdot W(t) \begin{bmatrix} \overline{\psi}(t) \\ \overline{X}(t) \end{bmatrix} + \overline{u}(t), \qquad (7.5)$$

где $H(t) = \left[k \cdot E_\psi \vdots E \vdots E_W \vdots k \cdot E_\gamma\right]$ – блочная матрица размера (3×15);

$$E_\psi = \begin{bmatrix} 0 & -W_Z(t) & W_Y(t) \\ W_Z(t) & 0 & -W_X(t) \\ -W_Y(t) & W_X(t) & 0 \end{bmatrix}$$ – составляющая блочной матрицы, отвечающая вектору $\overline{\psi}(t)$;

$$E_W = \begin{bmatrix} W_X(t) & 0 & 0 \\ 0 & W_Y(t) & 0 \\ 0 & 0 & W_Z(t) \end{bmatrix}$$ – составляющая блочной матрицы, отвечающая вектору Δk;

$$E_\gamma = \begin{bmatrix} -W_Z(t) & W_Y(t) & 0 & 0 & 0 & 0 \\ 0 & 0 & W_Z(t) & -W_X(t) & 0 & 0 \\ 0 & 0 & 0 & 0 & -W_Y(t) & W_X(t) \end{bmatrix}$$ – составляющая блочной матрицы, отвечающая вектору $\overline{\gamma}$.

Первое слагаемое в правой части данного уравнения измерений представляет собой расчетные (идеальные) измерения триады акселерометров ускорения $\overline{W}(t)$, а второе слагаемое описывает ошибки измерения, обусловленные ошибками ориентации (ВМП $\overline{\psi}(t)$) и инструментальными погрешностями триады акселерометров (вектор $\overline{X}_A$).

Элементы матрицы измерений $H(t)$ зависят только от вектора ускорения $\overline{W}(t)$ (вида траектории), что обеспечивает наглядность при анализе полученных уравнений и делает их удобными для моделирования и расчетов ошибок в выходных сигналах акселерометров. Если известны статистические характеристики помех измерений, то измерительный шум $\overline{u}(t)$ – может быть представлен соответствующим описанием, например, с использованием формирующих фильтров.

Следует отметить, что уравнение (7.5) допускает аддитивное расширение инструментальных погрешностей акселерометров, удобно не только для расчета ошибок и действительных показаний триады акселерометров, но также может быть использовано для описания процесса калибровки триады акселерометров, то есть определения составляющих вектора $\overline{X}_A$.

Ошибка ориентации гироплатформы $\overline{\psi}(t)$ может быть описана следующим дифференциальным уравнением [9]:

$$\dot{\overline{\psi}}(t) = \hat{\vartheta}(t)\overline{\psi}(t) + \delta\overline{\vartheta}(t), \tag{7.6}$$

где $\hat{\vartheta}(t) = \begin{bmatrix} 0 & \vartheta_Z(t) & -\vartheta_Y(t) \\ -\vartheta_Z(t) & 0 & \vartheta_X(t) \\ \vartheta_Y(t) & -\vartheta_X(t) & 0 \end{bmatrix}$ – кососимметрическая матрица; $\overline{\vartheta}(t) = [\vartheta_X(t) \quad \vartheta_Y(t) \quad \vartheta_Z(t)]^T$ – вектор угловой скорости приборного

ТГ, заданный в проекциях на собственные оси; $\delta\overline{\vartheta}(t)$ – ошибка в формировании вектора $\overline{\vartheta}(t)$, обусловленная уходами гироплатформы, ошибками выставки гироблоков на гироплатформе и ошибками в задании командной прецессии гироплатформе ТГС.

Запишем на основании прецессионной теории уравнение движения трехосного гиростабилизатора в режиме стабилизации с командной прецессией [8]:

$$\overline{\vartheta}(t) = \overline{\omega} + \gamma_{\Gamma}^{T} \cdot \overline{\vartheta}_{\text{к}}^{\text{р}}, \qquad (7.7)$$

где $\gamma_{\Gamma} = \begin{bmatrix} 1 & \gamma_{\Gamma 3}^{1} & -\gamma_{\Gamma 2}^{1} \\ -\gamma_{\Gamma 3}^{2} & 1 & \gamma_{\Gamma 1}^{2} \\ \gamma_{\Gamma 2}^{3} & -\gamma_{\Gamma 1}^{3} & 1 \end{bmatrix}$ – матрица, характеризующая ошибки выставки гироблоков Г1, Г2, Г3 на гироплатформе, ошибки выставки входных осей гироблоков заданы аналогично как и ошибки выставки осей чувствительности триады акселерометров, только для удобства верхние и нижние индексы переставлены местами и в нижний индекс добавлена «г» (гироблок);

$\overline{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix}$ – вектор угловой скорости ухода гироблоков;

ω_i – угловая скорость ухода i-го гироблока;

$\overline{\vartheta}_k^p = \begin{bmatrix} \vartheta_{kx}^{p} \\ \vartheta_{ky}^{p} \\ \vartheta_{kz}^{p} \end{bmatrix}$ – действительная, реально формируемая угловая скорость командной прецессии гироплатформы ТГС; «т» – знак транспонирования.

Действительная угловая скорость командной прецессии гироплатформы $\overline{\vartheta}_k^p$ отличается от расчетной (заданной) угловой скорости $\overline{\vartheta}_k$ вследствие отклонений коэффициентов передачи в цепи управления командной прецессией:

$$\overline{\vartheta}_k^p = (E+M)\overline{\vartheta}_k, \tag{7.8}$$

где $M = \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix}$ – матрица отклонений коэффициентов в цепи управления командной прецессией; m_i – отклонения коэффициента передачи i-го канала цепи управления командной прецессией гироплатформы; $\overline{\vartheta}_k$ – расчетное значение угловой скорости командной прецессии гироплатформы.

Используя уравнения (7.7), (7.8) определим ошибку в угловой скорости $\delta\overline{\vartheta} = \overline{\vartheta} - \overline{\vartheta}_k$ следующим образом:

$$\delta\overline{\vartheta} = \overline{\omega} + M \cdot \overline{\vartheta}_k + \hat{\gamma}_\Gamma^T \cdot \overline{\vartheta}_k, \tag{7.9}$$

где $\hat{\gamma}_\Gamma^T = \begin{bmatrix} 0 & -\gamma_{\Gamma 2}^3 & \gamma_{\Gamma 3}^2 \\ \gamma_{\Gamma 1}^3 & 0 & -\gamma_{\Gamma 3}^1 \\ -\gamma_{\Gamma 1}^2 & \gamma_{\Gamma 2}^1 & 0 \end{bmatrix}$.

Современные поплавковые гироблоки имеют развитую модель ухода (дрейфа), которую задают в виде полинома от ускорения, действующего по их осям [10]:

$$\omega_i = \omega_0^i + \omega_X^i \cdot W_{xi} + \omega_Y^i \cdot W_{yi} + \omega_Z^i \cdot W_{zi} + \omega_{XY}^i \cdot W_{xi} \cdot W_{yi} +$$
$$+ \omega_{XZ}^i \cdot W_{xi} \cdot W_{zi} + \omega_{YZ}^i \cdot W_{yi} \cdot W_{zi} + \omega_{YY}^i \cdot W_{yi}^2 + \omega_{ZZ}^i \cdot W_{zi}^2, \tag{7.10}$$

где i – номер гироблока; W_{xi}, W_{yi}, W_{zi} – проекции линейного ускорения соответственно на выходную, входную и ось собственного вращения i-го гироблока; ω_0^i – независящая от ускорения составляющая ухода i-го гироблока; $\omega_X^i, \omega_Y^i, \omega_Z^i$ – зависящие от первой степени ускорения составляющие ухода i-го гироблока; $\omega_{XY}^i, \omega_{XZ}^i, \omega_{YZ}^i, \omega_{YY}^i, \omega_{ZZ}^i$ – зависящие от произведений и от квадратов ускорений составляющие ухода i-го гироблока.

Введем вектора составляющих уходов гироблоков (i =1, 2, 3):

$$\overline{\omega}_{di} = \left[\omega_X^i, \omega_Y^i, \omega_Z^i, \omega_{XY}^i, \omega_{XZ}^i, \omega_{YZ}^i, \omega_{YY}^i, \omega_{ZZ}^i\right]^T.$$

Тогда вектор ухода i-го гироблока можно представить следующим выражением:

$$\omega_i = W_i \cdot \overline{\omega}_{di}, \tag{7.11}$$

где I – номер гироблока;

$W_i = \left[1, W_{xi}, W_{yi}, W_{zi}, W_{xi} \cdot W_{yi}, W_{xi} \cdot W_{zi}, W_{yi} \cdot W_{zi}, W_{yi}^2, W_{zi}^2\right]$ – матрица-строка размера (1×9).

Отсюда вектор уходов гироблоков легко задать следующим векторно-матричным уравнением:

$$\overline{\omega} = W \cdot \overline{\omega}_{di}, \tag{7.12}$$

где $W = \begin{bmatrix} W_1 & O_{19} & O_{19} \\ O_{19} & W_2 & O_{19} \\ O_{19} & O_{19} & W_3 \end{bmatrix} W = \begin{bmatrix} W_1 & O_{19} & O_{19} \\ O_{19} & W_2 & O_{19} \\ O_{19} & O_{19} & W_3 \end{bmatrix}$ – матрица размера (3×27), с нулевыми элементами; O_{19} – матрица размера (1×9) с нулевыми элементами; $\overline{\omega}_d = \begin{bmatrix} \overline{\omega}_{d1} \\ \overline{\omega}_{d2} \\ \overline{\omega}_{d3} \end{bmatrix}$ – вектор составляющих ухода гироблоков размерности 27.

Введем вектор инструментальных погрешностей гироблоков $\overline{X}_{\text{гб}}$ (размерности 36):

$$\overline{X}_{\text{гб}} = \begin{bmatrix} \overline{\omega}_{d1} \\ \overline{\omega}_{d2} \\ \overline{\omega}_{d3} \\ \overline{\gamma}_{\text{г}} \\ \overline{m} \end{bmatrix} = \begin{bmatrix} \overline{\omega}_d \\ \overline{\gamma}_{\text{г}} \\ \overline{m} \end{bmatrix}, \qquad (7.13)$$

где $\overline{\gamma}_{\text{г}} = \begin{bmatrix} \gamma_{\text{г}2}^3 & \gamma_{\text{г}3}^2 & \gamma_{\text{г}1}^3 & \gamma_{\text{г}3}^2 & \gamma_{\text{г}1}^2 & \gamma_{\text{г}2}^1 \end{bmatrix}^T$ – вектор ошибок выставки гироблоков на гироплатформе; $\overline{m} = [m_1 \quad m_2 \quad m_3]^T$ – вектор отклонений коэффициентов передачи в цепи управления командной прецессией гироплатформы.

Тогда используя выражения (7.9), (7.12), (7.13), после подстановки в уравнение (7.6) и преобразований, получим следующее дифференциальное векторно-матричное уравнение:

$$\dot{\overline{\Psi}}(t) = A(t)\begin{bmatrix} \overline{\Psi}(t) \\ \overline{X}(t) \end{bmatrix} + u_{\text{г}}, \qquad (7.14)$$

где $A(t) = \left[\hat{\vartheta}(t) \;\vdots\; W \;\vdots\; E_{\text{г}}(t) \;\vdots\; E_{\vartheta}(t)\right]$ – блочная матрица размера 3×39;

$$E_{\text{г}} = \begin{bmatrix} -\vartheta_{ky}(t) & & \vartheta_{kz}(t) & & 0 & & 0 & & 0 & & 0 \\ 0 & \vdots & 0 & \vdots & \vartheta_{kx}(t) & \vdots & -\vartheta_{kz}(t) & \vdots & 0 & \vdots & 0 \\ 0 & & 0 & & 0 & & 0 & & -\vartheta_{kx}(t) & & \vartheta_{ky}(t) \end{bmatrix} -$$

матрица размера (3×6), отвечающая вектору ошибок выставки гироблоков $\bar{\gamma}_{г}$;

$$E_{\vartheta}(t)=\begin{bmatrix}\vartheta_{KX}(t) & 0 & 0\\ 0 & \vartheta_{KY}(t) & 0\\ 0 & 0 & \vartheta_{KZ}(t)\end{bmatrix}$$ – матрица размера (3×3), отвечающая вектору ошибок управления командной прецессией гироплатформы $\bar{m}$;

$\bar{u}_{г}$ – вектор шумов.

Уравнение (14) в общем случае работы ТГС является нестационарным дифференциальным уравнением. Данное уравнение связывает ошибки ориентации гироплатформы (вектор $\bar{\psi}(t)$) с инструментальными погрешностями трехосного гиростабилизатора (вектор $\bar{X}_{гб}$), записано в форме Коши, что позволяет считать его уравнением состояния ТГС в данной задаче моделирования погрешностей ТГС. Матрица состояния $A(t)$ является блочной матрицей, матрицы-компоненты которой зависят от траектории движения летательного аппарата (матрица W) или от закона управления гироплатформой ТГС и режима ее работы (матрицы $E_{г}$, E_{ϑ}, $\hat{\vartheta}$). Уравнение состояния ТГС (14) является универсальным, так как при его выводе предполагалось, что ТГС находится в режиме стабилизации с командной прецессией. В связи с этим данное уравнение состояния пригодно для анализа ошибок ориентации гироплатформы во всех основных режимах: стабилизации, командной прецессии и приведения. Для использования уравнения в различных режимах работы ТГС нужно соответствующим образом задать элементы матрицы состояния $A(t)$. Например, для расчета ошибок ориентации гироплатформы в режиме стабилизации ТГС в процессе полета все матрицы-компоненты в матрице состояния $A(t)$ будут нулевыми, кроме матрицы W, которая зависит от вектора ускорений $\bar{W}(t)$. В режимах приведения и командной прецессии матрицы-компоненты $E_г, E_\vartheta, \hat{\vartheta}$ будут определятся видом применяемого закона управления гироплатформой. Очевидно, что для решения уравнения (7.14) необходимы начальные условия по вектору ошибок ориентирования ($\overline{\psi}(0)$), которые характеризуют ошибки начальной выставки гироплатформы в азимуте и горизонте.

Отметим, что уравнение измерений триады акселерометров (7.5) и уравнение состояния ТГС (7.14) составляют систему, так как связаны через вектор малого поворота $\overline{\psi}(t)$, и в матрицы $H(t)$ и $A(t)$ входят компоненты вектора ускорений $\overline{W}(t)$. Поэтому данную систему можно называть уравнением состояния и измерения ТГС.

Использование уравнений состояния и измерений ТГС (7.5), (7.14) позволяет произвести моделирование ошибок навигационных систем платформенного типа.

Последовательность выполнения моделирования может быть следующей:

– моделирование траектории движения;

– моделирование ошибок ориентации гироплатформы ТГС с использованием уравнения состояния (7.14);

– моделирование ошибок триады акселерометров с использованием уравнений измерений (7.5);

– моделирование ошибок решения навигационной задачи (задачи наведения) на основании результатов исследования ошибок ТГС и триады акселерометров согласно модели ошибок системы навигации (наведения).

ЗАКЛЮЧЕНИЕ

В учебном пособии рассмотрены вопросы, которые помогут специалистам технической сферы найти ответы на вопросы, связанные с постановкой задач исследования, путями их решения, моделирования и обработки полученных результатов.

Основные главы посвящены вопросам математики, теории вероятностей и математической статистики. Последовательно изложены вопросы использования систем поддержки принятия решений, приведена их классификация, сформулированы преимущества.

Пособие заканчивается вопросами моделирования, начиная от классификации моделей и, заканчивая примерами отдельных динамических структур.

Выполняя преобразование алгоритма управления, его приведение к машинному виду, необходимо обеспечить заданную точность системы управления при минимальных затратах таких машинных ресурсов, как быстродействие ЦВК, объем памяти ОЗУ и ПЗУ. В связи с этим наилучшим машинным алгоритмом считается такой, который содержит меньшее число промежуточных и исходных данных, хранимых в ЗУ машины; меньше логических условий, определяющих разветвленность вычислительного процесса.

При разработке машинных алгоритмов в качестве исходных условий учитываются система команд, форма представления чисел и адресность БЦВК.

В процессе преобразования алгоритмов решаются следующие задачи:

1. Выбор численного метода приведения алгоритма управления к последовательности машинных математических и логических операций.
2. Тождественное преобразование алгоритма к последовательному типу.
3. Сжатие информации, представляемой в виде таблиц или сложных зависимостей.
4. Масштабирование переменных.
5. Определение операций, выполняемых над переменными, представляемыми в формате удвоенной разрядности.

Выбор численного метода приведения алгоритма управления и последовательности машинных математических и логических операций осуществляется, исходя из системы команд предварительно выбранного (базового) ЦВУ. Эта система команд в процессе создания математического обеспечения ЦВК может неоднократно уточняться.

Дальнейшее развитие изложенной тематики видится в расширенном рассмотрении вопросов: более детального рассмотрения современной и классической теории управления, базирующееся на современной математике, а также работоспособности алгоритма при действии в возмущенном режиме.

СПИСОК ЛИТЕРАТУРЫ

1. Андрейчиков, А. В. Анализ, синтез, планирование решений в экономике / А. В. Андрейчиков, О. Н. Андрейчикова. – М. : Финансы и статистика, 2000. – 368 с.

2. Андрейчиков, А. В. Методология, методы и цифровые технологии принятия управленческих решений для повышения качества оптимизационных моделей ракетно-космической промышленности : монография / А. В. Андрейчиков, О. Н. Андрейчикова. — М. : РИОР, 2019 – 418 с.

3. Бриллюэн, Л. Наука и теория информации / Л. Бриллюэн ; пер. с англ. А. А. Харкевича. – М. : Гос. изд. физ-мат. лит-ры, 1960. – 391 с.

4. Ван дер Варден, Б. Л. Алгебра. / Б. Л. Ван дер Варден ; под ред. Ю. И. Мерзлякова. – М. : Наука, Главная редакция физико-математической литературы, 1979. – 623 с.

5. ГОСТ Р ИСО 9000–2015. Системы менеджмента качества. Основные положения и словарь. – М. : Стандартинформ, 2018. – 52 с.

6. ГОСТ Р 54521–2011. Статистические методы. Математические символы и знаки для применения в стандартах. – М. : Стандартинформ, 2012. – 32 с.

7. ГОСТ 15467–79. Управление качеством продукции. Основные понятия. Термины и определения (утв. и введен в действие Постановлением Государственного комитета СССР по стандартам от 26 января 1979 г. № 244. Переиздание (июнь 1986 г.) с Изменением № 1, утвержденным в январе 1985 г. (ИУС 4-85).

8. ГОСТ Р 51275–2006. Национальный стандарт Российской Федерации. Защита информации. Факторы, воздействующие на информацию. Общие положения. (Утв. и введен в действие приказом Федерального агентства по техническому регулированию и метрологии от 27 декабря 2006 г. № 374-ст). – М. : Стандартинформ. 2007. – 7 с.

9. ГОСТ Р 51170–98. Качество служебной информации. Термины и определения. Принят и введен в действие Постановлением Госстандарта России от 12 мая 1998 г. № 184. – М. : ИПК Издательство стандартов. 1998. – 7 с.

10. Информатика. Энциклопедический словарь для начинающих / под ред. Д. А. Поспелова. – М. : Педагогика-Пресс, 1994.

11. Князев, В. В. Вопросы формализации задачи научного исследования / В. В. Князев // Тр. ФГУП «НПЦАП» Системы и приборы управления. Научно-технический журнал. – 2019. – № 3. – С. 61 – 69.

12. Князев, В. В. К вопросу о векторном и многокритериальном принятии решений / В. В. Князев // Труды ФГУП «НПЦАП» Системы и приборы управления. Научно-технический журнал. – 2025. – № 1/71. – С. 4 – 8.

13. Князев, В. В. Метод сведения векторного показателя к скалярному с индикацией недопустимых сочетаний значений частных показателей / В. В. Князев // Тр. ФГУП «НПЦАП». Научно-технический журнал. – 2019. – № 2(43).

14. Князев, В. В. Методологические основы формализации обеспечения информационной безопасности отработки качества сложных динамических объектов : монография / В. В. Князев. – М. : МО РФ, 2009. – 307 с.

15. Ловцов, Д. А. Информационная теория эргасистем. Тезаурус / Д. А. Ловцов. – 2-е изд., испр. и доп. – М. : Наука, 2005 (Краснодар). – 245 с.

16. Ловцов, Д. А. Информационная теория эргасистем : монография / Д. А. Ловцов. – М. : РГУП, 2020. – 314 с.

17. Ловцов, Д. А. Системология правового регулирования информационных отношений в инфосфере : монография / Д. А. Ловцов. – М. : РГУП, 2016. – 316 с.

18. Микони, С. В. Теория принятия управленческих решений : учебное пособие / С. В. Микони. – СПб. : Лань, 2015. – 448 с.

19. Письменный, Д. Т. Конспект лекций по теории вероятностей, математической статистике и случайным процессам / Д. Т. Письменный. – М. : АЙРЕСС ПРЕСС, 2010. – 288 с.

20. Надежность и эффективность в технике : справочник : в 10 т. Т. 3. Эффективность технических систем / В. У. Торбин, Г. Н. Охотников, Е. С. Егоров и др. ; под ред. В. Ф. Уткина и Ю. В. Крючкова. – М. : Машиностроение, 1988. – 329 с.

21. Подиновский, В. В. Идеи и методы теории важности критериев в многокритериальных задачах принятия решений / В. В. Подиновский. – М. : Наука, 2019. – 103 с.

22. Подиновский, В. В. Математическая теория выработки решений в сложных ситуациях : учебник / В. В. Подиновский. – М. : МО СССР, 1981. – 211 с.

23. Подиновский, В. В. Многокритериальные задачи принятия решений: теория и методы анализа : учебник для вузов / В. В. Подиновский. – М. : Юрайт, 2023. – 486 с.

24. Постников, В. М. Методы принятия решений в системах организационного управления : учебное пособие / В. М. Постников, В. М. Черненький. – М. : Изд-во МГТУ им. Н. Э. Баумана, 2014. – 205 с.

25. Проектирование систем управления летательными аппаратами : учебное пособие / под ред. А. В. Зайцева. – М. : МО РФ, 2009. – 343 с.

26. Методы принятия решений в задачах оценки качества и технического уровня сложных технических систем / С. С. Семенов, Е. М. Воронов, А. В. Полтавский, А. В. Крянев ; под ред. д-ра техн. наук, профессора Е. Я. Рубиновича. – М. : ООО «ЛЕНАНД», 2015. – 520 с.

27. Соболева, Т. С. Дискретная математика : учебник для студ. вузов / Т. С. Соболева, А. В. Чечкин ; под ред. А. В. Чечкина. – М. : Издательский центр «Академия», 2006. – 256 с. – (Университетский учебник. Сер. Прикладная математика и информатика).

28. Урсул, А. Д. Отражение и информация / А. Д. Урсул. – М. : Мысль, 1973. – 231 с.

29. Об информации, информационных технологиях и о защите информации : федер. закон Российской Федерации от 27 июля 2006 г. № 149-ФЗ.

30. Чернавский, Д. С. Синергетика и информация / Д. С. Чернавский. – М. : Наука, 2001.

31. Шеннон, К. Математическая теория связи / К. Шеннон // Работы по теории информации и кибернетике. – М. : Изд-во ИЛ, 1963. – 830 с.

32. Ловцов, Д. А. Проблема обеспечения информационной безопасности России [Электронный ресурс]. – URL : http://www.nasled.ru/pressa/obozrev/N02_99/2_10.HTM

ПРИЛОЖЕНИЕ

Результаты анализа существующих моделей обеспечения безопасности информации (Артемов. А. В. Информационная безопасность. Курс лекций. – URL : https://tech.wikireading.ru/13009).

Название модели	Основная идея, соотношения с другими моделями	Математические основы и формальная запись	Механизмы изменения состояния и изменения доступа	Принципы реклассификации ресурсов
1	2	3	4	5
Модель Белла-Лападуллы (BLP-модель)	Конечный автомат, описывающий механизм доступа к ресурсам системы	Теория отношений $B=(R,D,W,z_0)$, R - множество запросов; D - множество реакций; W - множество состояний; Z_0 - начальное состояние.	По правилам трансформации состояний. Для управления доступом используется матрица контроля доступа	Реклассификация не рассматривается
Модель Белла для сетей коммутации пакетов (В - модель)	Расширены понятия безопасности, дискретной и *-безопасности, введены правила безопасного выполнения операций одностороннего соединения, терми-нации соединения, удаления прав на соединение. Может быть частично описана в терминах TAC- и SP-моделей	Теория отношений $B=(R,D,W,Z_0)$ (см. BLP-модель)	По правилам трансформации состояний. Для управления доступом используется матрица контроля доступом	Реклассификация не рассматривается
Модель мандатного управления доступом (МАС-модель)	Позволяет рассматривать операции объединения прав пользователей в неоднородных системах. Может быть описана в терминах TAC- и SP-моделей	Теория множеств; МАС== (V,R,T,u_0); V - множ. состояний; R - множ. запросов; T-функция перехода; u_0 - начальное состояние	Используется функция перехода. Управление доступом основывается на мандатном принципе	Реклассификация выполняется администратором безопасности

Модель безопасности военной системы передачи данных (MMS-модель)	Модель с администратором безопасности для управления доступом к данным и устройствам глобальной сети передачи данных	Теория множеств. MMS=(I, S, s_0, TF), I - множество запросов; S - множество состояний; s_0 - начальное состояние; TF - функция трансформации состояний	С помощью функции трансформации. Для управления доступом используется матрицы контроля доступа	Реклассификация выполняется администратором безопасности и уполномоченными пользователями
Модель трансформации прав доступа (TAC-модель)	Модель описывает разграничения прав доступа на основе представления субъекту в данный момент времени только одного права доступа. Может быть описана в терминах SP - модель	Теория множеств. TAC= (To, TS,R,cc,cr, trsns, grant), To, TS - множества типов объектов и субъектов; R - множество прав; ee - функция разрешения создания объекта; trans, grand - функции	Основывается на применении функции трансформации прав доступа	Реклассификация не рассматривается. Классификация ресурсов задается при их создании
Модель с графом защиты (PRG-модель)	Определяет безопасное выполнение операции заимствования, передачи прав в терминах теории графов	Теория графов. PRG=(F,L,P,/,#), V - множество ресурсов; lev -множество терминальных узлов; P - множество продукции; R - множество прав доступа	Состояние изменяется исчислением продукции. Управление доступом реализуется с помощью меток графа защиты	Реклассификация не рассматривается
Схематическая модель (SP-модель)	Модель матрицы доступа со строгой типизацией ресурсов. Может быть описана в терминах TAC-модели	Теория множеств и теория предикатов	Изменение прав доступа достигается аппаратом копирования меток доступа	Реклассификация не рассматривается
Иерархическая модель (H-модель)	Описание управления доступа для параллельных вычислений. Базируется на понятии критического интервала. Может быть описана в терминах TAC-	Теория предикатов $H = (ST, C)$, ST- множество меток; C - множество команд	Изменение матрицы состояния. Управление доступом основывается на вычислении предикатов	Реклассификация не рассматривается
Модель безопасных спецификаций (JJ-модель)	Определяет количество информации, необходимое для раскрытия системы защиты в целом	Аксиоматика Хоара. $JJ = (s_P, s_T)$, S_P - алфавит процесса; S_T - множество трасс процесса	Понятие механизма изменения состояния не применяется. Управление доступом на основе классификации пользователей	Реклассификация не рассматривается
Модель информационных потоков (WT-модель)	Модель в виде совокупности объектов и атрибутов, что позволяет определить информационные потоки	Теория множеств. $WT = (E, A)$, E - множество объектов; A - множество атрибутов	Изменения соотношения между объектами и атрибутами. Управление доступом осуществляется на основе атрибутов объектов	Реклассификация не рассматривается

ОГЛАВЛЕНИЕ

Учебное электронное издание

КНЯЗЕВ Владимир Владимирович
ЗАЙЦЕВ Александр Владимирович

ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ

Учебное пособие

Редактирование Е. С. Мордасовой
Графический и мультимедийный дизайнер Т. Ю. Зотова
Обложка, тиражирование, упаковка Е. С. Мордасовой

ISBN 978-5-8265-2970-6

9 785826 529706

Подписано к использованию 05.12.2025.
Тираж 50 экз. Заказ № 128

Издательский центр ФГБОУ ВО «ТГТУ»
392000, г. Тамбов, ул. Советская, д. 106/5,
помещение 2, к. 14
Тел. 8(4752) 63-81-08;
E-mail: izdatelstvo@tstu.ru