

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

«НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ
ЛИНГВИСТИЧЕСКИЙ УНИВЕРСИТЕТ
ИМ. Н.А. ДОБРОЛЮБОВА»
(НГЛУ)

В.В. Савченко, Д.Ю. Акатьев

АКТУАЛЬНЫЕ ГЛАВЫ
ВЫСШЕЙ МАТЕМАТИКИ

Учебно-методические материалы для студентов
социально-экономических направлений подготовки

Нижний Новгород
2016

Печатается по решению редакционно-издательского совета НГЛУ.
Направления подготовки: 38.03.01 – Экономика, 41.03.05 –
Международные отношения.

Дисциплина: Основы математического анализа.

УДК 51(075.8)

ББК 22

С 137

Савченко В.В., Акатьев Д.Ю. Актуальные главы высшей математики: Учебно-методические материалы для студентов социально-экономических направлений подготовки. – Н. Новгород: НГЛУ, 2016. – 53 с.

В настоящих учебно-методических материалах изложены основные понятия и определения ряда актуальных разделов высшей математики: алгебры матриц и векторов, алгебры множеств и алгебры случайных событий, в которых создается необходимая теоретическая база для последующего изучения основ и возможностей современной информатики.

Материалы предназначены для самостоятельной работы студентов вузов социально-экономических направлений подготовки над математическими учебными курсами.

УДК 51(075.8)

ББК 22

Авторы: В.В. Савченко, д-р техн. наук, профессор

Д.Ю. Акатьев, канд. техн. наук, доцент

Рецензент Ю.В. Воронков, канд. техн. наук, доцент, директор отделения очно-заочного и заочного обучения

© НГЛУ, 2016

© Савченко В.В., Акатьев Д.Ю., 2016

СОДЕРЖАНИЕ

Введение	4
1. Элементы алгебры матриц	6
1.1. Матрицы и действия над ними.....	6
1.2. Определитель квадратной матрицы.....	7
1.3. Обратные матрицы.....	9
2. Линейные векторные пространства	11
2.1. Векторы и действия над ними.....	11
2.2. Линейная комбинация векторов.....	13
2.3. Элементы аналитической геометрии.....	16
3. Абстрактная булева алгебра	21
3.1. Понятие булевой алгебры.....	21
3.2. Алгебра множеств.....	22
3.3. Алгебра событий и алгебра логики.....	23
4. Нормированная алгебра событий и введение в теорию вероятностей	25
4.1. Нормированная алгебра событий.....	25
4.2. Основные теоремы теории вероятностей.....	28
5. Случайная величина и ее закон распределения	31
5.1. Понятие случайной величины.....	31
5.2. Основные числовые характеристики случайной величины.....	33
5.3. Энтропия дискретной случайной величины по Шеннону.....	34
5.4. Система случайных величин.....	35
6. Элементы математической статистики	37
6.1. Теория выборочного метода.....	37
6.2. Задача проверки статистических гипотез.....	39
7. Математические методы оптимизации	41
7.1. Введение в проблему оптимальности.....	41
7.1.1. Критерии оптимальности.....	42
7.1.2. Ограничения.....	43
7.1.3. О методах решения оптимизационных задач.....	44
7.2. Идея итеративных методов.....	45
7.3. Разновидности итеративных методов.....	47
7.4. Методы стохастической аппроксимации.....	48
Список рекомендуемой литературы	52

ВВЕДЕНИЕ

Общеизвестная информатизация всех без исключения областей человеческой деятельности предъявляет повышенные требования к математической подготовке дипломированных специалистов. Сказанное в полной мере относится и к специалистам гуманитарного профиля, в том числе к экономистам и международникам, современная профессиональная деятельность которых неразрывно связана с обработкой разнообразных баз данных и применением новейших способов и средств информационных коммуникаций. Между тем возможности и особенности современной вычислительной техники могут быть доступны и раскрыты в полной мере для широкого круга пользователей лишь при условии их достаточной базовой подготовки по ряду актуальных разделов современной высшей математики. Например, профессиональная работа с базами данных предполагает предварительное изучение пользователями некоторых ключевых элементов алгебры логики. Другой пример: широко распространенные в современных программных продуктах методы статистической обработки информации базируются на математическом аппарате теории вероятностей. И алгебра логики, и теория вероятностей – это два актуальных раздела высшей математики, непосредственно связанных с задачами и возможностями прикладной информатики. В этот же ряд безусловно следует включить и теорию матриц и векторов, или линейную алгебру, а также алгебру множеств и алгебру событий, в терминах которых формулируются все основные аксиомы теории вероятностей. Проблема состоит в том, что их изложение в существующей специальной литературе, как правило, не рассчитано на неподготовленных читателей, к числу которых, в основном, и относятся студенты социально-экономических направлений подготовки. Стремление как-то систематизировать и по возможности упростить изложение необходимого

учебного материала, сделать его доступным и понятным для широкого круга читателей, сохраняя при этом достаточную степень широты и глубины теоретического обзора, и послужило для авторов главным стимулом для написания данной книги.

Включенный в УММ материал по своему содержанию и объему согласован с требованиями ФГОС по направлениям подготовки 38.03.01 – *Экономика* и 41.03.05 – *Международные отношения*.

1. ЭЛЕМЕНТЫ АЛГЕБРЫ МАТРИЦ

1.1. Матрицы и действия над ними

Прямоугольная таблица чисел общего вида, имеющая m строк и n столбцов, называется **матрицей** размера $m \times n$. Обозначается как

$$A = \begin{bmatrix} a_{11}a_{12}\dots a_{1n} \\ a_{21}a_{22}\dots a_{2n} \\ \dots\dots\dots\dots\dots\dots \\ a_{m1}a_{m2}\dots a_{mn} \end{bmatrix}.$$

Обычно в практике применения матриц используются более простые обозначения: $[a_{ij}]$, $\|a_{ij}\|$ и другие, где i – номер строки, j – номер столбца $i < m; j \leq n$.

В случае $m = 1$ получаем матрицу $A = [a_{11}, a_{12}, \dots, a_{1n}] = \text{str}(a_{1i})$ – вектор-строка длиной n . Это частный случай общего понятия матрицы. В случае

$n=1$ получаем матрицу $A = \begin{bmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{bmatrix} = \text{col}(a_{j1})$ – это вектор-столбец размера m .

Если $m = n$, то имеем **квадратную матрицу** $A = [a_{ij}]$ размера (порядка) n . Квадратная матрица специального вида называется **единичной матрицей**.

$$I_n = E = \begin{bmatrix} 1 & 0 & 0 \dots 0 \\ 0 & 1 & 0 \dots 0 \\ 0 & 0 & 1 \dots 0 \\ 0 & 0 & 0 \dots 1 \end{bmatrix}.$$

Единицы здесь располагаются на **главной диагонали**. **Нулевая матрица** состоит из одних нулей: $O_n = [0]$.

Действия над матрицами выполняются по следующим правилам:

1. **Суммой двух матриц** A и B называется такая матрица $C = [c_{ij}]$, каждый элемент которой $[c_{ij}] = [a_{ij}] + [b_{ij}]$ равен сумме одноименных (на одних и тех же позициях) элементов матриц A и B . Сумма обозначается как $C = A + B$.

2. Произведением матриц \mathbf{A} и \mathbf{B} называется матрица \mathbf{C} , каждый элемент которой определяется следующим образом:

$$c_{ij} = \sum a_{ik} * b_{kj} = a_{i1} * b_{1j} + a_{i2} * b_{2j} + \dots + a_{ip} * b_{pj}$$

При этом предполагается, что матрицы \mathbf{A} и \mathbf{B} согласованы по своим размерам: \mathbf{A} ($m \times p$) и \mathbf{B} ($p \times n$), т. е. число столбцов матрицы \mathbf{A} равно числу строк матрицы \mathbf{B} . Указанное определение произведения сводится, таким образом, к поэлементному умножению и суммированию элементов i -ой строки матрицы \mathbf{A} и элементов j -го столбца матрицы \mathbf{B} . Произведение матриц обозначается как $\mathbf{C} = \mathbf{AB}$.

Рассмотренные действия характеризуются следующими **свойствами**:

1. Коммутативность: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, но в общем случае $\mathbf{AB} \neq \mathbf{BA}$.
2. Ассоциативность: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$, $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.
3. $\mathbf{A} + \mathbf{O} = \mathbf{O} + \mathbf{A} = \mathbf{A}$, т. е. сложение с нулевой матрицей не меняет вида исходной матрицы \mathbf{A} .
4. Пусть \mathbf{A} – квадратная матрица размера $n \times n$. Тогда $\mathbf{A} * \mathbf{I}_n = \mathbf{I}_n * \mathbf{A}$, т. е. умножение матрицы на единичную не меняет ее вида.

Матрица \mathbf{B} называется транспонированной по отношению к \mathbf{A} , если в этой матрице \mathbf{B} строки соответствуют столбцам матрицы \mathbf{A} , т. е. в транспонированной матрице используются те же элементы. В частном случае равенства $m = 1$, когда $\mathbf{A} = \vec{a} = str(a_j)$ – вектор-строка размера n , получаем $\mathbf{A}^T = \vec{a}^T = col(a_j)$ – вектор-столбец того же размера n . В другом случае равенства $n = 1$ имеем $\mathbf{A} = \vec{a} = col(a_j)$ – вектор-столбец размера m , а $\mathbf{A}^T = \vec{a}^T str(a_j)$ – вектор-строка того же размера n . Таким образом, транспонирование вектор-строки \vec{a} дает вектор-столбец \vec{a}^T , и наоборот.

Транспонированная матрица обозначается как \mathbf{A}^T .

1.2. Определитель квадратной матрицы

Определителем (детерминантом) квадратной матрицы \mathbf{A} размера $n \times n$

называется число, равное сумме произведений элементов первой строки на их алгебраические дополнения:

$$\det \mathbf{A} = |\mathbf{A}| = a_{11} * A_{11} + a_{12} * A_{12} + \dots + a_{1n} * A_{1n}.$$

Здесь A_{ij} – алгебраическое дополнение элемента первой строки и j -го столбца.

Алгебраическим дополнением произвольного элемента a_{ij} матрицы \mathbf{A} называется определитель матрицы \mathbf{A} меньшего порядка $n-1$, полученной из исходной матрицы \mathbf{A} вычеркиванием из нее i -ой строки и j -го столбца одновременно, взятый со знаком «+», если число $i + j$ четное, и со знаком «-» в обратном случае.

Пример.

Пусть $n = 2$, т. е. матрица \mathbf{A} имеет 2 строки и 2 столбца: $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. Тогда

$$\det \mathbf{A} = a_{11}A_{11} + a_{12}A_{12} = a_{11}a_{22} * (-1)^{1+1} + a_{12}a_{21} * (-1)^{1+2} = a_{11}a_{22} - a_{12}a_{21}.$$

Матрица \mathbf{A} , определитель которой равен нулю, т. е. $\det \mathbf{A} = 0$, называется **вырожденной** (особенной) матрицей. Если $\det \mathbf{A}$ не равен нулю, то матрица \mathbf{A} **не вырождена**.

Определитель матрицы обладает следующими свойствами:

а) вне зависимости от номера строки $i = 1, 2, \dots, n$, выполняется равенство:

$$\det \mathbf{A} = \sum A_{ij} * a_{ij},$$

т. е. определитель раскладывается в сумму произведений алгебраических дополнений и элементов **любой строки** матрицы \mathbf{A} ;

$$\text{б) } \det \mathbf{A}^T = \det \mathbf{A},$$

т.е. определитель транспонированной матрицы равен определителю матрицы \mathbf{A} ;

в) при перемене местами в матрице \mathbf{A} любых двух строк или столбцов определитель матрицы \mathbf{A} умножается на (-1) ;

$$\text{г) } \det \mathbf{AB} = \det \mathbf{B} * \det \mathbf{A},$$

т. е. определитель произведения двух матриц \mathbf{A} и \mathbf{B} равен произведению их определителей;

д) определитель единичной матрицы равен единице, т. е. $\det \mathbf{I}_n = 1$;

е) необходимым и достаточным условием равенства определителя матрицы \mathbf{A} нулю является линейная зависимость ее строк или столбцов.

Говорят, что i -я строка $a_i = (a_{i1}, a_{i2}, \dots, a_{in})$ матрицы \mathbf{A} **линейно зависит** от ее других строк, если существуют числа k_1, k_2, \dots, k_n , одновременно не равные нулю, для которых выполняется равенство

$$k_1 * (\text{1 строка}) + k_2 * (\text{2 строка}) + \dots + k_n * (\text{n-я строка}) = (0, 0, \dots, 0).$$

Пример. Пусть квадратная матрица \mathbf{A} составлена из элементов

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix},$$

т. е. имеет порядок $n = 3$. Тогда выполняется равенство $1 * (\text{1 стр.}) - 2 * (\text{2 стр.}) + 1 * (\text{3 стр.}) = (1, 2, 3) - 2 * (4, 5, 6) + (7, 8, 9) = (1, 2, 3) - (8, 10, 12) + (7, 8, 9) = (0, 0, 0)$, т. е. строки этой матрицы линейно связаны друг с другом. Здесь применено сокращение: стр. – строка.

ж) определитель квадратной матрицы \mathbf{A} , содержащий хотя бы одну строку или столбец, равные нулю, равен нулю, т. к., раскладывая этот определитель по произведениям алгебраических дополнений и элементов нулевой строки, мы всегда получим в сумме 0.

Следствие: любая прямоугольная матрица \mathbf{A} ($m \times n$) является вырожденной.

1.3. Обратные матрицы

Квадратная матрица \mathbf{B} размера $n \times n$ называется **обратной** к матрице \mathbf{A} , если выполняется следующая система равенств:

$$\mathbf{BA} = \mathbf{AB} = \mathbf{I}_n, \text{ т. е. } \mathbf{B} = \mathbf{A}^{-1}.$$

При этом обратная матрица от обратной $\mathbf{B} = (\mathbf{A}^{-1})^{-1} = \mathbf{A}$ есть сама матрица \mathbf{A} .

Для того, чтобы матрица \mathbf{A} имела обратную, необходимо и достаточно, чтобы матрица \mathbf{A} была невырожденной (вырожденные матрицы не имеют обратных).

По определению обратной матрицы можно записать

$$\mathbf{B} = \mathbf{A}^{-1} = \begin{bmatrix} \frac{A_{11}}{\Delta} & \frac{A_{21}}{\Delta} & \dots & \frac{A_{n1}}{\Delta} \\ \dots & \dots & \dots & \dots \\ \frac{A_{1n}}{\Delta} & \frac{A_{2n}}{\Delta} & \dots & \frac{A_{nn}}{\Delta} \end{bmatrix}, \text{ где } \Delta = \det \mathbf{A} - \text{определитель матрицы } \mathbf{A};$$

A_{ji} – алгебраическое дополнение элемента a_{ji} , стоящего на пересечении j -ой строки и i -го столбца. Обратные матрицы широко применяются **при решении систем линейных уравнений** относительно нескольких неизвестных величин. Рассмотрим такую систему общего вида:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2, \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m. \end{cases}$$

В матричных обозначениях будем иметь: $\mathbf{Ax} = \mathbf{b}$, где

$\mathbf{A} = \| a_{ji} \|$ – матрица коэффициентов;

$\mathbf{x} = (x_1, \dots, x_n)^T$ – вектор-столбец неизвестных величин порядка n ;

$\mathbf{b} = (b_1, \dots, b_m)^T$ – вектор-столбец свободных коэффициентов порядка m .

При заданных матрице \mathbf{A} и векторе \mathbf{b} требуется найти вектор \mathbf{x} . В зависимости от соотношения между числом переменных n и числом уравнений m возникают следующие ситуации:

1. Если $n > m$, то рассматриваемая задача не достаточно определена и в общем случае имеет множество (а не единственное, как это обычно требуется) решений;

2. Если, напротив, $n < m$, то задача переопределена и в общем случае не имеет ни одного решения;

3. Если $n = m$, т. е. матрица \mathbf{A} – квадратная, то искомое решение задачи записывается в виде: $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Это решение вполне определено, если матрица \mathbf{A} не вырождена. В противном случае задача также не имеет единственного решения.

Пример. Рассмотрим систему двух уравнений относительно двух неизвестных:

$$\begin{cases} X_1 + X_2 = 5, \\ 2X_1 - X_2 = 3. \end{cases}$$

Ее матричная запись имеет вид:

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}.$$

Решая эту систему, получаем:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \frac{1}{[-3]} \begin{bmatrix} -1 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix} = -\frac{1}{3} \begin{bmatrix} -8 \\ -7 \end{bmatrix} = \begin{bmatrix} 2\frac{2}{3} \\ 2\frac{1}{3} \end{bmatrix}, \text{ т. е.}$$

$$X_1 = 2\frac{2}{3}; \quad X_2 = 2\frac{1}{3}.$$

Проверка: при $X_1 = 2\frac{2}{3}$ и $X_2 = 2\frac{1}{3}$ получаем:

$$X_1 + X_2 = 2\frac{2}{3} + 2\frac{1}{3} = 5,$$

$$2X_1 - X_2 = 4\frac{4}{3} - 2\frac{1}{3} = 3,$$

что и требовалось доказать.

2. ЛИНЕЙНЫЕ ВЕКТОРНЫЕ ПРОСТРАНСТВА

Понятие **вектор**, с которым мы впервые встретились в рамках линейной алгебры или алгебры матриц, имеет множество различных интерпретаций и определений. Наиболее строго это понятие определяется в терминах теории линейных векторных пространств.

2.1. Векторы и действия над ними

В общем случае **вектором** называется любая упорядоченная последовательность символов (a_1, a_2, \dots, a_n) , которые в дальнейшем

условимся считать вещественными числами из множества $\mathbf{R} = (-\infty; \infty)$. Иными словами, для любого $i = 1, \dots, n$: $a_i \in \mathbf{R}$. Число n определяет здесь размер или, говорят, *размерность* вектора, а числа a_1, a_2, \dots, a_n – его *координаты*. Сам вектор обычно обозначается малой буквой латинского алфавита и выделяется жирным шрифтом или стрелкой над символом, например,

$$\vec{a} = \mathbf{a} = (a_1, \dots, a_n); \quad \vec{b} = \mathbf{b} = (b_1, \dots, b_n).$$

Если имеются два вектора \mathbf{a} и \mathbf{b} , то **суммой** этих векторов называется вектор $\mathbf{c} = \mathbf{a} + \mathbf{b}$, с координатами $c_1 = a_1 + b_1; c_2 = a_2 + b_2 \dots c_n = a_n + b_n$.

Если имеется вектор \vec{a} , а также любой возможный скаляр (то есть просто число) $\lambda \in \mathbf{R}$, то **произведением вектора на скаляр** называется вектор $\vec{c} = \lambda \vec{a}$ с координатами: $c_1 = \lambda a_1; c_2 = \lambda a_2; \dots c_n = \lambda a_n$.

Понятие суммы двух векторов и произведения вектора на скаляр образуют две основные аксиомы алгебры векторов.

Нулевым вектором размерности n называется вектор $\mathbf{0} = (0, \dots, 0)$, составленный из n нулей.

Вектором, противоположным вектору \mathbf{a} , называется вектор $\mathbf{c} = -\mathbf{a}$, с координатами $c_1 = -a_1, \dots, c_n = -a_n$.

На множестве всех допустимых координат $a_i \in \mathbf{R}$, (т. е. $-\infty < a_i < +\infty$) понятие вектора $\mathbf{a} = (a_1, \dots, a_n)$ определяет множество или пространство векторов $\mathbf{R}^n = \{\mathbf{a}\}$. Здесь n – размерность векторного пространства (число его координат). При соблюдении всех перечисленных выше аксиом пространство \mathbf{R}^n называется линейным векторным пространством или арифметическим n -мерным пространством. Таким образом, любой вектор \mathbf{a} – это один из элементов арифметического пространства, т. е. $\mathbf{a} \in \mathbf{R}^n$.

Действия или операции над векторами в пространстве \mathbf{R}^n характеризуются следующими свойствами:

$$1) \left. \begin{array}{l} \vec{a} + \vec{b} = \vec{b} + \vec{a}; \\ \lambda \vec{a} = \vec{a} \lambda; \end{array} \right\} \text{коммутативность (доказывается по определению);}$$

$$2) \left. \begin{array}{l} (\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c}); \\ (\lambda\beta) \vec{a} = \lambda(\beta \vec{a}); \end{array} \right\} \text{ассоциативность;}$$

$$3) \left. \begin{array}{l} (\lambda + \beta) \vec{a} = \lambda \vec{a} + \beta \vec{a}; \\ \lambda(\vec{a} + \vec{b}) = \lambda \vec{a} + \lambda \vec{b}; \end{array} \right\} \text{распределительный закон;}$$

$$4) 1 \mathbf{a} = \mathbf{a};$$

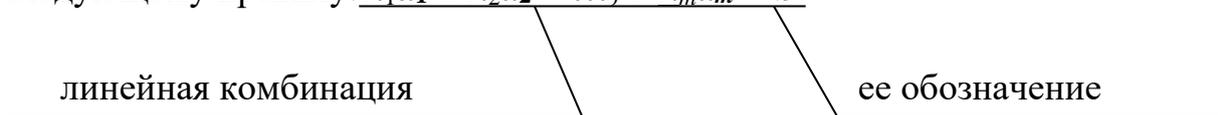
$$0 \mathbf{a} = \mathbf{0}.$$

2.2. Линейная комбинация векторов

Пусть задана любая совокупность из $m > 1$ разных векторов одной размерности n :

$$\begin{aligned} \vec{a}_1 &= (a_{11}, \dots, a_{1n}); \\ \vec{a}_2 &= (a_{21}, \dots, a_{2n}); \\ &\dots\dots\dots \\ \vec{a}_m &= (a_{m1}, \dots, a_{mn}). \end{aligned}$$

Тогда общий вид их линейной комбинации определяется по следующему правилу: $\lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + \dots + \lambda_m \mathbf{a}_m = \mathbf{b}$



Где вектор $\mathbf{b} = (b_1, \dots, b_n)$ составлен из координат:

$$b_i = \lambda_1 a_{1i} + \lambda_2 a_{2i} + \dots + \lambda_m a_{mi} \text{ для всех } i = 1, 2, \dots, n.$$

В рассматриваемом случае говорят, что вектор \mathbf{b} линейно связан с векторами $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ в совокупности. Указанная совокупность векторов называется линейно зависимой, если существуют такие варианты

множителей $\lambda_1, \lambda_2, \dots, \lambda_m$, одновременно не равных нулю, при которых вектор $\mathbf{b} = \mathbf{0}$, т. е. равен нулю вектору размерности n .

Напротив, если при любом наборе множителей $\lambda_1, \lambda_2, \dots, \lambda_m$ вектор $\mathbf{b} \neq \mathbf{0}$ (исключая тривиальный случай $\lambda_i = 0 \forall i \neq m$), то указанная совокупность векторов **линейно независима**.

Теорема 2.1.

Пусть $\mathbf{A} = \begin{bmatrix} a_{11}a_{12}\dots a_{1n} \\ a_{21}a_{22}\dots a_{2n} \\ \dots\dots\dots \\ a_{m1}a_{m2}\dots a_{mn} \end{bmatrix}$ – матрица ($m \times n$), составленная из элементов

рассматриваемой совокупности векторов. Тогда эта система линейно независима в совокупности, если ранг матрицы $\text{rang } \mathbf{A} = m$ – точно равен числу векторов m .

Определение 1. Рангом прямоугольной матрицы \mathbf{A} называется максимальный порядок k отличного от нуля минора этой матрицы \mathbf{M}_k , $k = 1, 2, \dots, m$.

Определение 2. Определитель порядка $k \det \mathbf{A}_k$ ($k < m; k < n$) квадратной матрицы \mathbf{A} , составленный для любых k строк и любых k столбцов исходной матрицы \mathbf{A} , называется ее минором k -го порядка.

Пример. Рассмотрим матрицу вида

$$\mathbf{A} = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix}, \text{ где } m = n = 3. \text{ Ее ранг равен двум: } \text{rang } \mathbf{A} = 2, \text{ т. к. } \det \mathbf{A} = 0,$$

но определитель второго порядка не равен нулю:

$$\begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} = M_2 = 5 - 8 = -3 \neq 0.$$

Следствие 1. Необходимым условием линейной независимости m векторов является нестрогое неравенство вида $m < n$, где m – число рассматриваемых векторов $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$, n – их размерность или число координат.

Теорема 2.2.

Пусть $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ – произвольный базис в \mathbf{R}^n , тогда любой другой вектор $\mathbf{b} = (b_1, b_2, \dots, b_n)$ размерности n может быть разложен в заданном базисе согласно следующему выражению:

$$\mathbf{b} = \beta_1 \mathbf{a}_1 + \beta_2 \mathbf{a}_2 + \dots + \beta_n \mathbf{a}_n.$$

Здесь $\beta_1, \beta_2, \dots, \beta_n$ – координаты вектора \mathbf{b} в рассматриваемом базисе $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$.

Следствие. В применении к стандартному базису формула разложения вектора \mathbf{b} переписывается следующим образом:

$$\mathbf{b} = b_1 \mathbf{e}_1 + b_2 \mathbf{e}_2 + \dots + b_n \mathbf{e}_n.$$

Таким образом, можно утверждать, что элементы произвольного вектора $\mathbf{b} = (b_1, b_2, \dots, b_n)$ являются его координатами в стандартном базисе пространства \mathbf{R}^n .

В качестве иллюстрации рассмотрим следующий рисунок. Здесь $n = 2$, а векторное пространство представляет собой множество точек на плоскости.

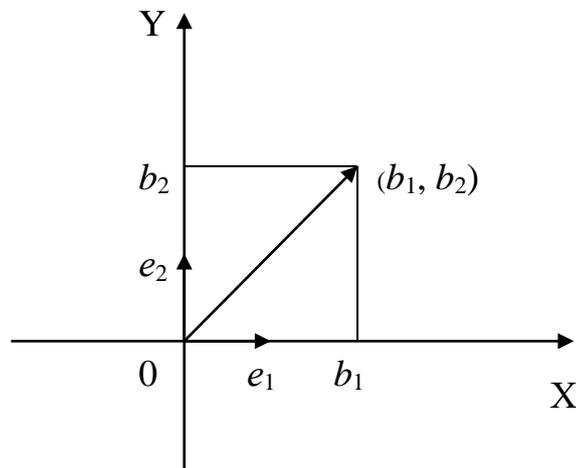


Рис. 2

2.3. Элементы аналитической геометрии

В терминах аналитической геометрии понятие **вектор** приобретает геометрический смысл. Во многих случаях оно используется для иллюстрации и объяснения некоторых основных понятий линейной алгебры.

Вектор – это **направленный отрезок прямой** на плоскости или в трехмерном пространстве координат. Его обычно обозначают **a**, **b** и т. д.

Начало вектора, по договоренности, обычно размещается в начале системы координат. Для иллюстрации рассмотрим три варианта задания вектора: на числовой оси \mathbf{R}^1 (рисунок 3), на плоскости \mathbf{R}^2 (рисунок 4) и в трехмерном пространстве \mathbf{R}^3 (рисунок 5).

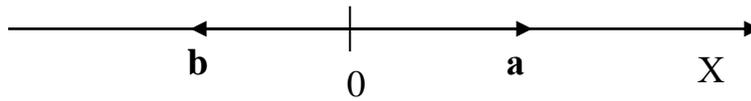


Рис. 3

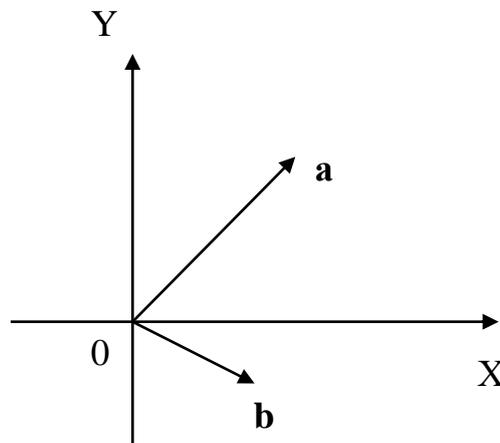


Рис. 4

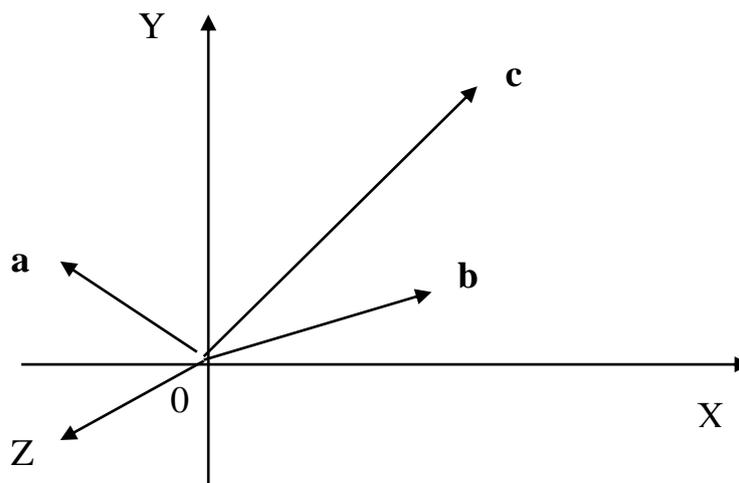


Рис. 5

В любом варианте своего задания вектор исчерпывающе определяется координатами своего конца (указан на рисунке стрелками). Важнейшей характеристикой вектора \mathbf{a} является его **длина**, или норма, $|\mathbf{a}|$.

Вектор нулевой длины $|\mathbf{a}| = 0$ называется **нулевым, нуль вектором**, и обозначается как $\mathbf{0}$. Его основное свойство состоит в следующем равенстве: $\mathbf{0} + \mathbf{a} = \mathbf{a}$.

Над векторами действия выполняются по следующим правилам:

1. Сложение двух векторов $\mathbf{a} + \mathbf{b} = \mathbf{c}$ – по правилу параллелограмма (рисунок 6).

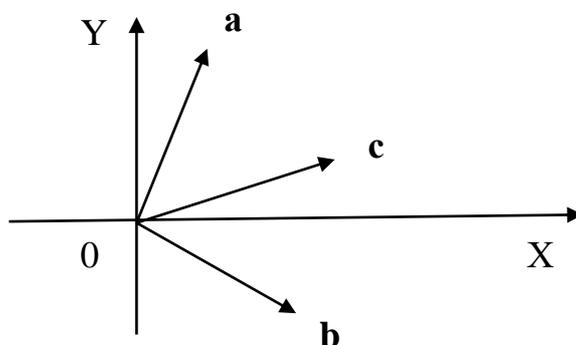


Рис. 6

2. Умножение вектора на скаляр $\mathbf{c} = \lambda \mathbf{a}$ не меняет направление вектора \mathbf{a} , если $\lambda > 0$, и меняет его на противоположное, если $\lambda < 0$ (рисунок 7).

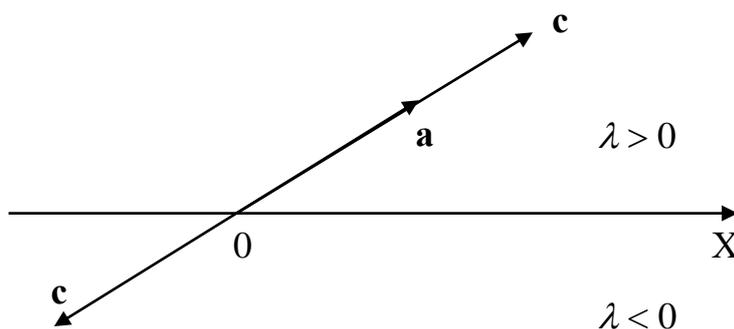


Рис. 7

Два вектора \mathbf{a} и \mathbf{c} , лежащие на одной прямой, называются **коллинеарными**.

Отметим, что рассмотренные операции являются стандартными не

только в рамках аналитической геометрии, но и линейной алгебры в целом. Совсем другой смысл имеет еще одна важная операция над векторами, а именно **скалярное произведение векторов**.

3. В качестве скалярного произведения векторов **a** и **b** понимают скалярную величину или число $c = (\mathbf{a}, \mathbf{b}) = |\mathbf{a}| |\mathbf{b}| \cos(\mathbf{a}, \mathbf{b})$, где $\cos(\mathbf{a}, \mathbf{b})$ обозначает косинус угла между рассматриваемыми векторами на плоскости.

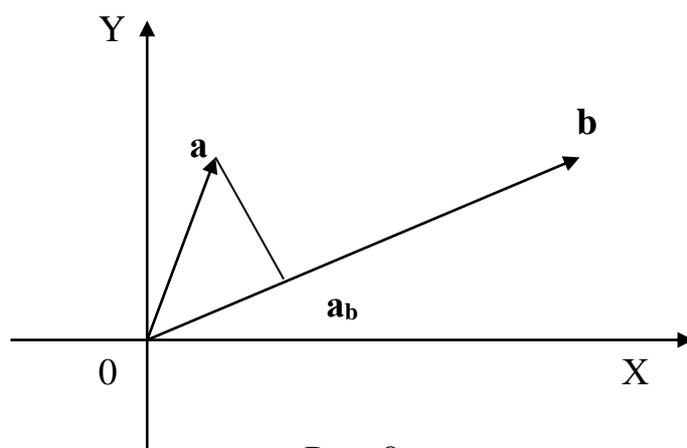


Рис. 8

Физический смысл введенного понятия объясняется следующим образом: $c = |\mathbf{b}| a_b$, где a_b – проекция вектора **a** на вектор **b** (рисунок 8). Чем больше проекция, тем больше скалярное произведение. Напротив, в случае $\mathbf{b} \perp \mathbf{a}$ получаем $c = 0$ (\perp – символ ортогональности векторов).

Таким образом, два ортогональных вектора **b** и **a** всегда имеют нулевое скалярное произведение.

Определение 4. В линейном векторном пространстве \mathbf{R}^n аналогом понятию скалярного произведения аналитической геометрии служит величина:

$$(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n a_i b_i = \mathbf{a} \mathbf{b}^T = c. \quad (*)$$

Здесь T – операция транспонирования, т. е. преобразования вектора строки $\mathbf{b} = (b_1, b_2, \dots, b_n)$ в вектор-столбец. Данное определение охватывает как частный случай любое фиксированное значение размерности $n = 1, 2, \dots$.

Определение 5. Арифметическое пространство любой размерности \mathbf{R}^n , в котором определена операция скалярного умножения векторов (*), называется **n -мерным евклидовым пространством.**

Длина или норма вектора в этом пространстве определяется по следующей формуле: $|\vec{a}| = \sqrt{|\vec{a}|^2} = \sqrt{\vec{a}\vec{a}^T} = \sqrt{\sum_{i=1}^n a_i^2}$.

Последнее выражение распространяется на произвольное значение размерности, включая все три рассмотренных варианта задания векторов в рамках аналитической геометрии (рисунки 3–5).

Все векторы стандартного базиса $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ в пространстве \mathbf{R}^n взаимно (попарно) ортогональны, т. к.

$$(\mathbf{e}_i, \mathbf{e}_j) = 0*0 + \dots + 0*1 + 0*0 + \dots + 1*0 + 0*0 + \dots + 0*0 = 0,$$

а их длины равны единице:

$$|\mathbf{e}_j| = (0*0 + \dots + 1*1 + 0*0 + \dots + 0*0)^{1/2} = 1.$$

Определение 6. Система векторов $\mathbf{e}_1, \dots, \mathbf{e}_n$ размерности n каждый называется **ортонормированным базисом n -мерного векторного пространства.**

Теорема 2.3.

Для любого вектора $\mathbf{b} \in \mathbf{R}$ из n -мерного векторного пространства выполняется равенство:

$$\mathbf{b} = (\mathbf{b}, \mathbf{e}_1) \mathbf{e}_1 + (\mathbf{b}, \mathbf{e}_2) \mathbf{e}_2 + \dots + (\mathbf{b}, \mathbf{e}_n) \mathbf{e}_n,$$

т. е. каждая i -я координата $b_i = (\mathbf{b}, \mathbf{e}_i)$ вектора \mathbf{b} в стандартном базисе определяется его скалярным умножением на соответствующий вектор \mathbf{e}_i из этого базиса: $(\mathbf{b}, \mathbf{e}_i) = b_i*0 + \dots + b_i*1 + \dots + b_n*0 = b_i$.

Сделанный вывод распространяется на любой другой базис $\mathbf{a}_1, \dots, \mathbf{a}_n$ в пространстве \mathbf{R}^n (см. предыдущую теорему 2.2): $b_i = (\mathbf{b}, \mathbf{a}_i) \forall i = \overline{1, n}$.

3. АБСТРАКТНАЯ БУЛЕВА АЛГЕБРА

3.1. Понятие булевой алгебры

Булевой алгеброй называется класс \mathbf{K} объектов A, B, C, \dots , в котором определены две бинарные операции: сложение и умножение, обладающие следующими свойствами:

I. Для всех A, B, C из \mathbf{K} :

1) \mathbf{K} содержит $A + B$ и AB (замкнутость);

2) $A + B = B + A, AB = BA$ (коммутативные законы);

3) $A + (B + C) = (A + B) + C, A(BC) = (AB)C$ (ассоциативные законы);

4) $A(B + C) = AB + AC, A + BC = (A + B)(A + C)$ (дистрибутивные законы);

5) $A + A = AA = A$ (свойства идемпотентности);

6) $A + B = B$ в том и только в том случае, если $AB = A$ (свойство совместимости).

II. Кроме того,

7) \mathbf{K} содержит элементы \mathbf{I} и \mathbf{O} такие, что для всякого элемента A из \mathbf{K}
 $A + \mathbf{O} = A, A\mathbf{I} = A, A\mathbf{O} = \mathbf{O}, A + \mathbf{I} = \mathbf{I}$;

8) для каждого элемента A класс \mathbf{K} содержит элемент \bar{A} (дополнение элемента A), такой, что

$$A + \bar{A} = \mathbf{I}, \quad A\bar{A} = \mathbf{O}.$$

В каждой булевой алгебре справедливы соотношения:

$$A(A + B) = A + AB = A \text{ – законы поглощения};$$

$$\overline{(A + B)} = \bar{A}\bar{B}, \quad \overline{(AB)} = \bar{A} + \bar{B} \text{ – законы де Моргана};$$

$$\overline{\bar{A}} = A; \quad \bar{\mathbf{I}} = \mathbf{O}; \quad \bar{\mathbf{O}} = \mathbf{I}.$$

$$A + \bar{A}B = A + B, \quad \bar{A}\bar{B} + AC + BC = \bar{A}\bar{C} + BC.$$

Булевыми такие алгебры называются по имени известного английского математика и логика XIX в. Джорджа Буля, впервые

применившего такую алгебру в своих исследованиях. Самыми распространенными примерами булевых алгебр являются алгебра множеств, алгебра событий и алгебра логики.

3.2. Алгебра множеств

Множеством называется совокупность объектов любой природы.

Сумму $A + B$ двух множеств A и B естественно определить как их объединение (рисунок 9, а). При этом, очевидно, будем иметь

$$A + B = B + A \quad \text{и} \quad (A + B) + C = A + (B + C).$$

Роль нуля здесь будет играть так называемое «пустое» множество O , вовсе не содержащее элементов; для такого множества имеем

$$A + O = A.$$

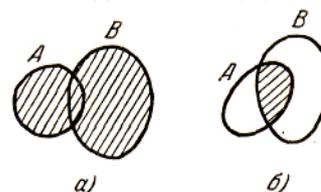


Рис. 9

Определим теперь **произведение** AB двух множеств A и B как их общую часть или пересечение (рисунок 9, б). В таком случае, очевидно,

$$AB = BA \quad \text{и} \quad (AB)C = A(BC).$$

Роль единицы здесь играет универсальное множество I , которое является исходным «общим» множеством для конкретной задачи. Действительно, для любого множества A имеем $AI = A$.

$$\text{Очевидно также, что} \quad AO = O, \quad A + I = I.$$

Для определенной таким образом «алгебры множеств» имеет место также распределительный или дистрибутивный закон:

$$(A + B)C = AC + BC.$$

Кроме того, выполняется и второй дистрибутивный закон:

$$A + BC = (A + B)(A + C).$$

Справедливы также следующие свойства:

$$A + A = AA = A;$$

$$A + B = B \quad \text{в том и только в том случае, если} \quad AB = A.$$

Существенное отличие алгебры множеств от алгебры чисел заключается в наличии в алгебре множеств еще одной операции, ставящей

в соответствие каждому множеству A новое множество \bar{A} (дополнение A). Эта операция определяется следующим образом: \bar{A} состоит из всех элементов универсального множества I , не принадлежащих множеству A . При этом $A + \bar{A} = I$ и $A\bar{A} = O$.

Таким образом, мы показали, что алгебра множеств действительно является булевой алгеброй, т. к. для них выполняется один и тот же набор определяющих аксиом.

3.3. Алгебра событий и алгебра логики

Существует также множество совокупностей разных объектов, для которых естественно определяются понятия суммы, произведения, а также «дополнения» \bar{A} , удовлетворяющие всем перечисленным выше свойствам. Одним из примеров таких совокупностей является совокупность **событий**.

Введем несколько основных понятий **алгебры событий**.

Будем называть **опытом**, или **испытанием**, всякое осуществление определенных условий или действий, при которых наблюдается изучаемое явление. Например, опытом является стрельба по мишени, бросание монеты, выборка одного изделия из множества готовых изделий и т. д. Все, что может быть исходом опыта, называют **событием**.

Достоверным называется событие I , которое обязательно должно произойти в данном опыте. **Невозможным** называется событие O , которое в данном опыте произойти не может. Промежуточное положение между событиями I и O занимают **случайные события**, которые по результатам данного опыта, наблюдения или испытания могут как произойти, так и не произойти. Например, рост завтра цен на бирже – это случайное, т. е. заранее не определенное событие.

События называют **равновозможными**, если нет оснований считать, что одно событие является более возможным, чем другие. Например, при бросании монеты события A (появление цифры) и B (появление герба) равновозможны при условии, что монета изготовлена из однородного

материала, имеет правильную цилиндрическую форму и наличие чеканки не влияет на то, окажется герб или цифра на верхней стороне монеты.

Над объектами типа событий при анализе выполняются определенные операции. Рассмотрим некоторые (базовые) из них.

Суммой (или объединением) событий $A + B$ называется событие, которое заключается в том, что происходит хотя бы одно из данных событий.

Произведением (или пересечением) событий AB называется событие, которое заключается в том, что происходят оба данных события.

Дополнением (или противоположным событием) к событию A называется событие \bar{A} , которое заключается в том, что не происходит событие A .

Нетрудно убедиться, что алгебра событий обладает всеми теми же свойствами, что и алгебра множеств, и, соответственно также является булевой алгеброй. Еще один важный пример из того же ряда, что и предыдущие две алгебры, составляет множество всех логических предложений, которое составляет предмет изучения математической логики.

Договоримся под логическими предложениями понимать элементарные (бинарные) высказывания, относительно каждого из которых можно лишь заключить, что оно истинно или ложно. Обозначим их как A, B, C, \dots .

Тогда под суммой $A + B$ предложений A и B следует понимать предложение «или A , или B » – это логическая операция **дизъюнкции**. Под произведением AB – соответственно предложение «и A , и B » – это операция **конъюнкции**. А под обозначением \bar{A} будем понимать **отрицание** предложения A (предложение «не A »).

Истинность или ложность высказывания, образованного из других высказываний с помощью операций дизъюнкции, конъюнкции и отрицания, определяют с помощью **таблиц истинности** логических операций:

A	B	A+B		A	B	AB		A	\bar{A}
И	И	И		И	И	И		И	Л
И	Л	И		И	Л	Л		Л	И
Л	И	И		Л	И	Л			
Л	Л	Л		Л	Л	Л			

Под элементами **I** и **O** в алгебре логики понимают предложение, которое заведомо истинно, и предложение, которое заведомо ложно.

Видим, что все аксиомы булевой алгебры в алгебре логики сохраняют свою справедливость. Так, например, $A + \bar{A} = \mathbf{I}$ – известный логический закон исключенного третьего: во всех случаях высказывание **A** либо истинно, либо ложно; соотношение $A\bar{A} = \mathbf{O}$ – есть закон непротиворечия: ни в каком случае суждение **A** не может быть одновременно и истинным, и ложным.

4. НОРМИРОВАННАЯ АЛГЕБРА СОБЫТИЙ И ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ

4.1. Нормированная алгебра событий

Элементы алгебры Буля, вообще говоря, не являются числами. Однако часто удается поставить в соответствие каждому элементу **A** некоторое число $|A|$ или $p(A)$, удовлетворяющее следующим условиям:

$$0 \leq p(A) \leq 1; p(\mathbf{O}) = 0, p(\mathbf{I}) = 1;$$

$$\text{если } A \subseteq B, \text{ то } p(A) \leq p(B);$$

$$\text{если } AB = \mathbf{O}, \text{ то } p(A + B) = p(A) + p(B).$$

Это число называют абсолютной величиной элемента **A** или его **нормой**, а саму алгебру Буля в этом случае называют **нормированной**. Совокупность всех предложений математической логики можно

рассматривать как нормированную алгебру Буля, если условиться считать абсолютную величину (норму) предложения равной 1, если это предложение истинно, и равной 0, если оно ложно. Примером нормированной алгебры Буля является и алгебра событий, где роль абсолютной величины или нормы события A играет **вероятность** $p(A)$ этого события.

Связь теории вероятностей с алгебрами Буля может быть положена в основу общего определения самого предмета этой науки. А именно, можно утверждать, что теория вероятностей изучает совокупности объектов, образующие нормированную алгебру Буля; эти объекты называются **событиями**, а норма $p(A)$ события A называется **вероятностью**.

Отсюда вытекает, что в любой относящейся к этой теории задаче исходная алгебра Буля обязательно должна быть задана заранее (т. е. так или иначе указана в самом условии задачи). **Основным предметом исследования** теории вероятностей служат **случайные события**. А **основной задачей** теории вероятностей при этом следует считать нахождение вероятностей сложных, или составных, случайных событий по заданным вероятностям элементарных или первоначальных событий A , B ,

Рассмотрим некоторые ключевые понятия теории вероятностей.

Два события A и B называются **совместными** в данном опыте, если появление одного из них не исключает появления другого. Так, при бросании двух монет события A (цифра на верхней стороне первой монеты) и B (герб на верхней стороне второй монеты) являются совместными.

Два события называются **несовместными** в данном опыте, если появление одного из них исключает появление другого в этом опыте. Примеры несовместных событий: попадание и промах при одном выстреле, «герб» и «цифра» при бросании одной монеты.

Множество событий A_1, A_2, \dots, A_n образуют **полную группу событий**, если они попарно-несовместны и появление одного и только одного (любого) из них является достоверным событием. Например, при бросании игральной кости событие «Выпало 1 очко» обозначим через A_1 , событие «Выпало 2 очка» обозначим A_2 и т. д. Тогда события $A_1, A_2, A_3, A_4, A_5, A_6$ образуют полную группу попарно-несовместных событий: появление одного и только одного из них будет достоверным событием.

Противоположными событиями называются два несовместных события, образующих полную группу событий. Примеры противоположных событий: попадание и промах при стрельбе, выпадение и невыпадение одного очка при бросании игральной кости и т. д.

Назовем каждое событие, которое может наступить в итоге опыта, **элементарным исходом** или **шансом**. Например, события: A_1 – выпадение одного очка, A_2 – выпадение двух очков, ..., A_6 – выпадение шести очков – будут элементарными исходами при бросании игральной кости. Элементарные исходы, при которых рассматриваемое событие наступает, называются **благоприятствующими** этому событию (или **благоприятными шансами**). Так, при бросании игральной кости элементарные исходы A_3 и A_6 являются благоприятствующими событию «Число очков делится на 3».

Вероятностью события (по Лапласу) называется отношение числа элементарных исходов, благоприятствующих данному событию, к числу всех равновозможных элементарных исходов опыта, в котором может появиться это событие. Вероятность события A обозначим как $P(A)$, тогда по определению можно записать

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

где m – число элементарных исходов, благоприятствующих событию A ; n – число всех равновозможных, образующих полную группу

элементарных исходов опыта, в котором может появиться событие А.

Это определение вероятности называется **классическим**. Его достоинством является то, что вероятность можно вычислить до фактического проведения опыта и сделать соответствующие выводы.

Из определения вероятности события можно заключить следующее:

- вероятность достоверного события всегда равна 1.
- вероятность невозможного события равна 0.
- вероятность случайного события выражается числом, лежащим в диапазоне от 0 до 1.

4.2. Основные теоремы теории вероятностей

Решение основной задачи теории вероятностей базируется на ряде ее основных теорем. Но прежде чем мы их сформулируем, рассмотрим некоторые вспомогательные понятия.

Два события А и В называются **зависимыми** (между собой), если вероятность одного из них зависит от появления (или не появления) другого.

Разъясним это понятие на следующем примере. В ящике имеется 100 готовых изделий, из них 90 стандартных и 10 нестандартных. Из ящика наудачу извлекается одно изделие и не возвращается обратно, потом извлекается второе изделие. Рассмотрим следующие события:

А – при первом испытании извлекается стандартное изделие;

В – при втором испытании извлекается также стандартное изделие.

Очевидно, что вероятность события В зависит от того, появилось ли событие А. Действительно, если событие А наступило, то $P(B) = 89/99$; если при первом испытании событие А не наступило, то $P(B) = 90/99 = 10/11$. Следовательно, рассматриваемые события А и В являются зависимыми в статистическом смысле.

События А и В называются **независимыми**, если вероятность одного из них не зависит от появления (или не появления) другого. Приведем

пример независимых событий. В урне 6 красных шаров и 4 синих шара. Извлекается один шар и возвращается обратно, потом снова извлекается один шар. Рассмотрим следующие события:

А – извлекается красный шар при первом испытании;

В – извлекается красный же шар и при втором испытании.

Здесь очевидно, что вероятность события В не зависит от того, произойдет или не произойдет событие А, в любом случае она составляет 6/10. Итак, рассмотренные события А и В являются независимыми.

Рассмотрим любые два зависимых события А и В. **Условной вероятностью** $P(B/A)$ называется вероятность события В при условии, что событие А произошло.

Поясним понятие условной вероятности на следующем примере. В урне 5 красных и 5 синих шаров. Из урны наугад извлекают один шар, затем другой. Какова вероятность того, что при втором испытании появится красный шар (событие В), если при первом испытании также был извлечен красный (событие А)? Если событие А уже произошло, то в урне осталось 9 шаров, из них 4 красных, поэтому $P(B/A) = 4/9$.

Для независимых событий А и В условные вероятности совпадают с безусловными, т. е. $P(B/A) = P(B)$, $P(A/B) = P(A)$.

Теорема 4.1. Умножение вероятностей.

Вероятность совместного появления двух событий равна произведению вероятности одного из них на условную вероятность другого, или

$$P(AB) = P(A)P(B/A) = P(B)P(A/B).$$

Для случая двух независимых событий теорема умножения вероятностей принимает следующий упрощенный вид:

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

Пример. Вероятность поражения цели первым стрелком (событие А) равна 0,8, а вероятность поражения цели вторым стрелком (событие В)

равна 0,7. Какова вероятность того, что цель будет поражена хотя бы одним стрелком?

Решение. Пусть событие $C = \{\text{цель поражена хотя бы одним стрелком}\}$, противоположное событие $\bar{C} = \{\text{оба стрелка промахнулись}\}$. Таким образом, $\bar{C} = \bar{A}\bar{B}$. По теореме умножения вероятностей, с учётом того, что событий \bar{A} и \bar{B} независимы, найдем

$$P(\bar{C}) = P(\bar{A})P(\bar{B}) = [1 - P(A)] \cdot [1 - P(B)] = (1 - 0,8) \cdot (1 - 0,7) = 0,2 \cdot 0,3 = 0,06.$$

Вероятность того, что цель будет поражена хотя бы одним стрелком:

$$P(C) = 1 - P(\bar{C}) = 1 - 0,06 = 0,94.$$

Теорема 4.2. Сложение вероятностей.

Вероятность суммы двух событий равна сумме вероятностей этих событий за вычетом вероятности их совместного появления, т. е.

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

Если события A и B статистически связаны или зависимы друг от друга, то $P(AB) = P(A)P(B/A)$ и теорема сложения вероятностей принимает вид

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

Если A и B независимы, то $P(AB) = P(A)P(B)$ и

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

И, наконец, если события A и B несовместны, то $AB = \mathbf{O}$, где \mathbf{O} – невозможное событие. В этом случае $P(AB) = 0$ и теорема сложения вероятностей принимает вид

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

Пусть событие A может наступить лишь при условии появления одного из попарно-несовместных событий B_1, B_2, \dots, B_n , образующих полную группу. Так как заранее неизвестно, какое из этих событий наступит, их называют гипотезами. Вероятности этих событий, а также условные вероятности события A : $P(A/B_1), P(A/B_2), \dots, P(A/B_n)$ заданы.

Вопрос о том, как найти в этом случае вероятность события A , решает следующая теорема.

Теорема 4.3. Формула полной вероятности.

Вероятность события A , которое может появиться при условии наступления одного из несовместных событий B_1, B_2, \dots, B_n , образующих полную группу, равна

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \dots + P(B_n)P(A/B_n).$$

Рассмотренные теоремы во многих случаях позволяют существенно упростить процедуру вычислений сложных (составных) случайных событий, пользуясь заданными вероятностями их составляющих событий, следуя универсальному принципу познания «от простого к сложному».

5. СЛУЧАЙНАЯ ВЕЛИЧИНА И ЕЕ ЗАКОН РАСПРЕДЕЛЕНИЯ

5.1. Понятие случайной величины

Для описания многих случайных событий при их статистическом (вероятностном) анализе широко применяется понятие **случайной величины (СВ)**. **Случайной** называется величина, которая в результате опыта, наблюдения или испытания принимает одно из множества своих возможных значений, причем какое именно – заранее неизвестно. Случайные величины обозначаются заглавными буквами X, Y, Z конца латинского алфавита, а их возможные значения – соответствующими строчными буквами x, y, z . Среди случайных величин различают дискретные и непрерывные величины.

Дискретной называется случайная величина, которая может принимать конечное или счетное множество значений (счетным называют множество, элементы которого можно пронумеровать).

Непрерывной называют случайную величину, возможные значения которой сплошь заполняют некоторый числовой интервал.

В задачах экономического анализа обычно оперируют как с дискретными, так и непрерывными СВ. Рост или спад котировки акций конкретной компании на следующий торговый день является примером дискретной случайной величины, принимающей два значения. При анализе письменных текстов дискретной случайной величиной может служить случайная величина X длины каждого очередного предложения при чтении нового текста. Напротив, при анализе звучащей речи предпочтение отдают непрерывным случайным величинам. Подробнее о них можно прочитать в учебном пособии В.В. Савченко «Теория вероятностей и математическая статистика» (Н. Новгород: НГЛУ, 2003). Мы же здесь с целью экономии времени ограничимся рассмотрением прежде всего дискретных случайных величин.

Дискретная СВ может принимать возможные значения с различными вероятностями. Чтобы охарактеризовать дискретную случайную величину в статистическом смысле, необходимо указать вероятности всех ее значений.

Законом распределения вероятностей дискретной случайной величины называется таблица соответствия между возможными значениями этой величины и их вероятностями. Такую таблицу называют **рядом распределения** дискретной случайной величины. Ее качественный вид показан ниже.

СВ	x_1	x_2	x_3	...	x_n
Вероятность	p_1	p_2	p_3	...	p_n

Поскольку в одном испытании случайная величина X может принять одно и только одно значение x_k , то события $X=x_1, X=x_2, \dots, X=x_n$ образуют полную группу событий. Следовательно, сумма вероятностей этих

событий равна единице, т. е. $\sum_{i=1}^n p_i = 1$.

5.2. Основные числовые характеристики случайной величины

К числу основных числовых характеристик случайной величины X принадлежат, прежде всего, ее математическое ожидание и дисперсия. **Математическим ожиданием** (МО) дискретной случайной величины X называется сумма произведений ее всех возможных значений на соответствующие вероятности. Математическое ожидание обозначается через $\mathbf{M}(X)$ или m_x . Если случайная величина принимает значения x_1, x_2, \dots, x_n соответственно с вероятностями p_1, p_2, \dots, p_n , то по определению МО будем иметь

$$\mathbf{M}(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i .$$

Отметим, что математическое ожидание случайной величины является величиной постоянной. Его часто называют средним (статистическим) значением случайной величины, а также центром распределения, т. к. около него группируются отдельные значения случайной величины.

Однако математическое ожидание характеризует случайную величину не полностью: зная математическое ожидание, нельзя сказать, какие конкретно значения принимает случайная величина и как отклоняются они от среднего значения. Чтобы знать, как рассеяны значения случайной величины вокруг ее математического ожидания, вводят другую числовую характеристику, называемую **дисперсией**.

Дисперсией (рассеянием) случайной величины X называется математическое ожидание квадрата ее отклонения от среднего статистического значения и обозначается через $\mathbf{D}(X)$, т. е.

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

Принимая во внимание определение математического ожидания, получаем расчетную формулу вида

$$\mathbf{D}(X) = \sum_{k=1}^n [x_k - \mathbf{M}(X)]^2 p_k .$$

Для вычисления дисперсии применяется также еще одна формула

$$D(X) = M(X^2) - [M(X)]^2,$$

которую можно вывести из определения.

Пример. Случайная величина X задана законом распределения:

X	0	1	2	3	4
p	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Найти математическое ожидание и дисперсию СВ X .

Решение. Вычислим математическое ожидание СВ X :

$$M(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = 2.$$

Аналогично найдем $M(X^2) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 4 \cdot \frac{6}{16} + 9 \cdot \frac{4}{16} + 16 \cdot \frac{1}{16} = 5.$

Вычислим дисперсию по формуле $D(X) = M(X^2) - [M(X)]^2 = 5 - 2^2 = 1.$

5.3. Энтропия дискретной случайной величины по Шеннону

Рассмотренные выше основные статистические характеристики (закон распределения, МО и дисперсия) при практическом анализе сложных случайных явлений органично сочетаются с рядом основных **информационных характеристик** случайных величин. Основной информационной характеристикой дискретной случайной величины является ее **энтропия**, вычисляемая по известной формуле К. Шеннона:

$$H(X) = -\sum_{i=1}^n p_i \log p_i.$$

Энтропия измеряется в **битах**, если основание логарифма равно двум, и **нитах**, если применяется натуральный логарифм. Величина 1 бит определяет энтропию случайного объекта с двумя равновероятными состояниями (типа элементарной ячейки памяти в ПК). Энтропия характеризует меру неопределенности поведения дискретной случайной величиной при ее многократных (повторных) наблюдениях. Чем больше энтропия, тем более непредсказуемой выглядит анализируемая СВ.

5.4. Система случайных величин

На практике исследователей интересуют не столько свойства отдельных случайных величин – их законы распределения, числовые характеристики, – сколько связи между разными случайными величинами. Например, в экономике интересуются зависимостью между производительностью труда и прибылью предприятия, в сельском хозяйстве – между количеством внесенного удобрения и величиной урожая, в здравоохранении – между количеством прививок и размерами эпидемии и т. д. В лингвистике математический подход распространился широко в связи, главным образом, с его уникальными возможностями для **сравнительного анализа** сложных, часто уникальных явлений.

Для решения такого рода сложных задач в теории вероятностей разработан специальный аппарат, рассчитанный на анализ нескольких случайных величин, рассматриваемых не изолированно друг от друга, а в системе. На данном этапе для экономии времени ограничимся обзорным рассмотрением простейшей системы из двух случайных величин (X, Y) дискретного типа.

Система двух случайных величин (X, Y) образует понятие **двумерной случайной величины (ДСВ)**. Закон распределения ДСВ (X, Y) дискретного типа можно задать таблицей распределения, в которой указаны значения x_i, y_k ; компонент X, Y и вероятности p_{ik} принятия двумерной случайной величиной (X, Y) ее значений $(x_i, y_k); i = 1, \dots, m; k = 1, \dots, n$:

	y_1		y_k	...	y_n
x_1	p_{11}		p_{1n}
x_2	p_{21}		p_{2n}
...
x_i			p_{ik}		
...					
x_m	p_{m1}		p_{mn}

Заметим, что события $(x_i, y_k), i = 1, \dots, m; k = 1, \dots, n$ попарно несовместны и образуют полную группу. Поэтому выполняется тождество

$$\sum_{i=1}^m \sum_{k=1}^n p_{ik} = 1.$$

В качестве важной числовой характеристики статистической связи между случайными величинами X , Y применяется **коэффициент взаимной корреляции** вида

$$\rho_{XY} = \frac{K_{XY}}{\sigma_X \sigma_Y}.$$

Величина K_{XY} называется **ковариацией**, или **корреляционным моментом**. Ее можно вычислить по формуле

$$K_{XY} = \mathbf{M}[XY] - m_X m_Y = \sum_{i=1}^m \sum_{k=1}^n x_i y_k p_{ik} - m_X m_Y.$$

Важное практическое значение имеют следующие **свойства коэффициента корреляции**:

1. $-1 \leq \rho_{XY} \leq 1$, т. е. ρ_{XY} – нормированная числовая характеристика;
2. Если случайные величины X , Y независимы, то $\rho_{XY} = 0$.

Отмеченные свойства позволяют утверждать, что коэффициент корреляции ρ_{XY} является показателем степени статистической зависимости между случайными величинами X и Y .

Другим показателем степени связи между двумя рассматриваемыми случайными величинами может служить **величина взаимной энтропии** или информационного рассогласования двух распределений в метрике Кульбака – Лейблера. Взаимная энтропия (ВЭ) двух случайных величин X , Y вычисляется по формуле

$$H(X/Y) = \sum_{i=1}^n p_i \log (p_i / q_i) \geq 0,$$

где p_i, q_i – это вероятности i -го значения случайной величины X и Y соответственно. Равенство ВЭ нулю достигается лишь при условии идентичности двух распределений. Чем больше различия в законах распределений, тем больше значение ВЭ.

6. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Математическая статистика занимается разработкой методов сбора, описания и обработки опытных данных, т. е. результатов наблюдений, с целью получения научных и практических выводов.

К числу основных задач математической статистики относятся:

- 1) оценка функции распределения случайной величины (или ее параметров) по результатам статистических наблюдений;
- 2) проверка статистических гипотез.

6.1. Теория выборочного метода

На практике для изучения признаков, характеризующих некоторую случайную величину, используют выборочный метод. Для этого опыт или наблюдения над реализациями случайной величины X проводят многократно и независимо n раз. В результате получают так называемую выборку (x_1, x_2, \dots, x_n) объема n , содержащую n чисел.

Рассмотрим выборку, в которой среди n чисел имеется m различных значений, т. е. значение x_1 наблюдается n_1 раз, $x_2 - n_2, \dots, x_m - n_m$ раз, причем

$$\sum_{k=1}^m n_k = n.$$

Наблюдаемые значения x_k называют **вариантами**. Последовательность всех m вариантов, записанных в возрастающем порядке, называют **вариационным рядом**. Числа n_k называют частотами вариантов, а их отношения к объему выборки n – относительными частотами, которые обозначают как

$$\omega_k = n_k/n. \quad (*)$$

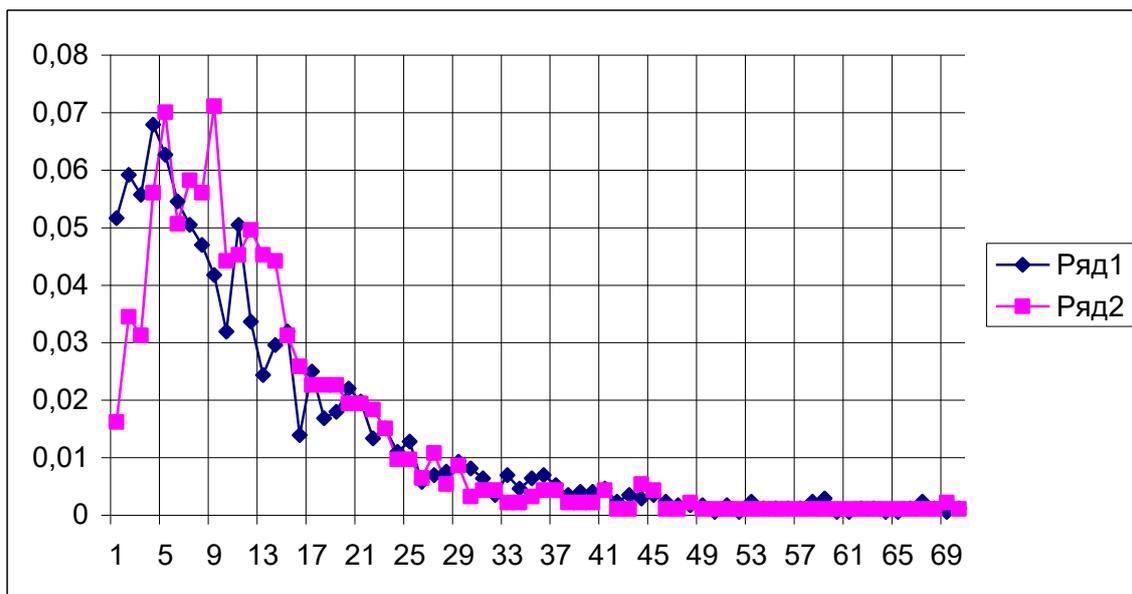
Последовательность частот всех вариантов x_k на множестве проведенных наблюдений определяет важное для статистики понятие **дискретного вариационного ряда (ДВР)**. Нетрудно увидеть, что в форме ДВР мы получаем, по сути, статистическую оценку (по выборке)

неизвестного закона распределения дискретной случайной величины X .

Чем больше объем выборки n , тем точнее такая оценка.

Наглядное представление о ДВР дает его график в системе координат «номер варианта (ось x) – относительная частота (ось y)». Указанный график называется **гистограммой распределения СВ**.

В качестве примера на рисунке ниже по тексту показаны в одних осях две гистограммы: для случайных величин длины предложений в текстах «Идиот» и «Судьба человека». Для их построения использовалась формула (*), а также фрагменты текстов объемом 150 и 70 тысяч знаков или, в пересчете к числу предложений, выборки объемом $n_1 = 1650$ и $n_2 = 860$ соответственно. При этом число вариантов $m = 70$ на гистограммах определяется максимальной длиной предложений в обоих текстах.



Видно, что различия между обоими графиками не носят качественного характера. Однако их количественная характеристика: величина взаимной энтропии – по формуле Кульбака – Лейблера из п. 5.4

достигает значения $H(X / Y) = \sum_{i=1}^n p_i \log_2(p_i / q_i) = 0,090$.

Для сравнения два аналогичных фрагмента, но взятых из одного

текста «Идиот», характеризуются величиной ВЭ порядка 0,03. Это в три раза меньше, чем для разных текстов.

В задачи статистического анализа входит не только определение неизвестного закона распределения наблюдаемой случайной величины X по имеющейся выборке $\{x_k\}$ конечного объема n . Такая задача тесно связана также с оцениванием параметров распределения (т. е. его числовых характеристик). Числовые характеристики случайной величины, полученные в результате обработки выборки, называются **выборочными** или **статистическими оценками**. Они являются приближением (в некотором смысле) к теоретическим характеристикам, которые чаще всего нам при анализе не известны.

Так, выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$

является **оценкой математического ожидания**. Аналогично, **оценка дисперсии** по выборке определяется следующим выражением:

$$s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2.$$

В теории показано, что при неограниченном увеличении объема выборки обе оценки имеют своими пределами истинные значения соответственно МО и дисперсии СВ. В таком случае говорят, что полученные оценки **сходятся** к неизвестным истинным значениям параметров распределений.

6.2. Задача проверки статистических гипотез

При решении многих научных и практических задач приходится делать предположения о законах распределения случайных величин. Такие предположения называют **статистическими гипотезами**. Из принятой гипотезы получают определенные следствия и проверяют, подтверждаются ли они на практике. Задача проверки статистических

гипотез является важной составной частью математической статистики.

Статистическая гипотеза называется **простой**, если соответствующее суждение позволит полностью восстановить закон распределения. В противном случае гипотеза называется **сложной**. Например, гипотеза о равенстве МО некоторой величине – сложная, т.к. она оставляет открытым вопрос о виде закона распределения.

Вместе с выдвинутой гипотезой (*нулевой*, или *основной*) рассматривается и противоречащая ей гипотеза, которую называют *конкурирующей* (*альтернативной*). Нулевую гипотезу обозначим H_0 , а конкурирующую – H_1 .

Выдвинутая статистическая гипотеза может оказаться правильной или, напротив, ошибочной, поэтому возникает необходимость ее проверки. Проверка осуществляется с помощью статистических методов и называется статистической. Возможны следующие результаты проверки:

- 1) принята правильная гипотеза;
- 2) отвергнута правильная гипотеза;
- 3) принята неправильная гипотеза;
- 4) отвергнута неправильная гипотеза.

Очевидно, верное решение осуществлено в первом и четвертом случаях, ошибочное – во втором и третьем.

При этом различают ошибки первого и второго рода. **Ошибкой первого рода** называют ошибку, состоящую в том, что отвергнута нулевая гипотеза, когда она была верна, а **ошибкой второго рода** – ошибку, состоящую в том, что принята альтернативная гипотеза, когда в действительности она была ошибочной.

Случайная величина K , служащая для проверки нулевой гипотезы, называется **статистическим критерием** (или просто критерием).

Значение K_n критерия K , вычисленное по выборкам, называется **наблюдаемым**.

Основной принцип проверки статистических гипотез состоит в следующем: если наблюдаемое значение выбранного критерия принадлежит некоторой области W (критической области), то нулевую гипотезу отвергают, в противном случае ее принимают.

7. МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОПТИМИЗАЦИИ

7.1. Введение в проблему оптимальности

Проблема оптимальности – это центральная проблема теоретической информатики. Любые сколько-нибудь обоснованные выводы, решения, алгоритмы применительно к конкретной задаче могут рассматриваться как оптимальные, т. е. наилучшие в определенном смысле, если именно **им отдано предпочтение** на практике.

Проблема оптимальности решается путем **оптимизации** применяемых алгоритмов обработки информации. Любая оптимизация преследует вполне определенную цель: достижение наивысших показателей точности, надежности и эффективности алгоритмов в условиях поставленной задачи. Выбор и формулировка цели – это **первый** (исходный) **этап** в решении проблемы оптимальности.

На втором этапе выбранная цель согласуется с имеющимися возможностями или ресурсами, которыми располагает исследователь при решении проблемы оптимальности. Главное содержание этого этапа – учет ограничивающих условий (или проще – ограничений) на используемые ресурсы, а также выбор соответствующего математического аппарата.

На завершающем **третьем этапе** решения проблемы оптимальности синтезируется оптимальный при рассматриваемых ограничениях алгоритм обработки информации и решаются вопросы его практической реализации. Именно здесь выясняется истинная ценность различных методов

оптимизации, их сравнительные преимущества и недостатки.

Суммируя все вышеизложенное, можно утверждать, что решение проблемы оптимальности в общем случае **сводится к решению трех перечисленных задач:**

- 1) выбор и формулировка цели;
- 2) введение ограничений на используемые ресурсы;
- 3) синтез и практическая реализация оптимального алгоритма.

7.1.1. Критерии оптимальности

Любая задача оптимизации может быть **сведена к выбору** наилучшего варианта алгоритма или устройства обработки сигналов из множества конкурирующих альтернативных вариантов. Каждый вариант характеризуется **вектором своих параметров** $\mathbf{c} = (c_1, c_2, \dots, c_N)$, где N – число контролируемых параметров, зависящее от сложности рассматриваемых алгоритмов. Качество того или иного варианта определяется некоторым показателем $J(\mathbf{c})$ – численной характеристикой, которая называется **целевой функцией** поставленной задачи.

Наилучшему (т. е. оптимальному) варианту отвечает вектор параметров \mathbf{c}^* , доставляющий экстремум (*max* или *min*) целевой функции $J(\mathbf{c})$, т. е. $J(\mathbf{c}^*) = \text{extr } J(\mathbf{c})$.

По своему характеру зависимость $J(\mathbf{c})$ есть функция многих переменных или, говорят, **функционал** вектора параметров \mathbf{c} . Функционал $J(\mathbf{c})$ часто называют функционалом качества оптимизационной задачи или, по-другому, **целевым функционалом**.

Таким образом, для нахождения оптимального варианта решения поставленной задачи, во-первых, определяют целевой функционал $J(\mathbf{c})$ на множестве всех возможных векторов \mathbf{c} и после этого находят вектор \mathbf{c}^* , отвечающий требованию:

$$\mathbf{c}^*: J(\mathbf{c}) = \text{extr} . \quad (1)$$

Последнее выражение называют **критерием оптимальности** искомого решения поставленной задачи.

7.1.2. Ограничения

Если на вектор параметров \mathbf{c} и вид функционала $J(\mathbf{c})$ не накладывают никаких ограничений, то рассматриваемая оптимизационная задача в формулировке (1) становится тривиальной, а собственно проблема оптимальности утрачивает всякий смысл.

Проблема оптимальности возникает именно тогда, когда существуют **противоречия в требованиях** к оптимальному варианту решения. Источником таких противоречий обычно служат **ограничения на область допустимых решений** или множество рассматриваемых вариантов \mathbf{c} . Решение оптимизационной задачи в этих условиях сводится не просто к поиску оптимального варианта \mathbf{c}^* , а к поиску наилучшего варианта в **пределах рассматриваемого множества**. Задачи такого ряда называются **оптимизационными задачами с ограничениями**.

Указанные ограничения могут быть заданы в виде равенств $g_r(\mathbf{c}) = 0$, $r = \overline{1, M_1}$, неравенств $g_r(\mathbf{c}) \leq 0$, $r = \overline{1, M_2}$ или логических соотношений между различными элементами вектора \mathbf{c} (оптимизируемого вектора параметров \mathbf{c}). По своей сути они разделяются на **ограничения I и II рода**.

Ограничения I рода – это системы уравнений или неравенств, связывающие выходные сигналы с входными сигналами и его параметрами. Например, линейный трансверсальный фильтр N -го порядка описывается в дискретном времени $n = 0, 1, 2, \dots$ линейным разностным уравнением вида

$$y_n = x_n + c_1 x_{n-1} + \dots + c_N x_{n-N}.$$

Здесь вектор оптимизируемых параметров $\mathbf{c} = (c_1, \dots, c_N)$ составляется из N весовых коэффициентов фильтра c_i , $i = 1, N$.

Представленное разностное уравнение определяет математическую модель трансверсального фильтра. Задать систему ограничений первого рода – это значит **задать математическую модель объекта оптимизации**.

Ограничения II рода накладываются на элементы вектора параметров $\mathbf{c} = (c_1, \dots, c_N)$ в пределах заданной математической модели. Например, полагая в трансверсальном фильтре $c_2 = 0, c_3 = 0, \dots, c_N = 0$, получаем разностное уравнение пониженного 1-го порядка:

$$y_n = x_n + c_1 x_{n-1}, \quad n = 0, 1, \dots$$

Таким образом, ограничения второго рода затрагивают в основном имеющиеся ресурсы при решении оптимизационной задачи (в рассмотренном примере фильтр N -го порядка при $N > 1$ – существенно более гибкая структура по сравнению с фильтром 1-го порядка).

При учете выбранного критерия оптимальности задача в общем виде формулируется как вариационная: $J(\mathbf{c}) = \text{extr}$,

$$g_v(\mathbf{c}) = 0, \quad v = \overline{1, M_1},$$

$$g_r(\mathbf{c}) \leq 0, \quad r = \overline{1, M_2},$$

с учётом ограничений в виде равенств и неравенств.

7.1.3. О методах решения оптимизационных задач

Стандартный подход к решению оптимизационных задач использует классические методы вариационного исчисления. Известные методы можно условно разделить на две большие группы:

- а) аналитические методы;
- б) алгоритмические или итеративные методы.

Первая группа методов привлекательна тем, что приводит к явному аналитическому решению оптимизационной задачи. Эти методы являются предметом изучения стандартного курса математического анализа. Основным их недостатком можно считать весьма ограниченную область применения: решение в явном виде существует лишь для относительно простых задач с высокой степенью идеализации условий и по этой причине обычно далеких от практики современной информатики. Так, формулы для вычисления корней алгебраических уравнений имеют

простой вид для уравнения 1-й и 2-й степени. Такие формулы можно записать для уравнений 3-й и 4-й степени, однако сложность их резко в этом случае возрастает. И, наконец, подобных формул просто не существует для уравнений выше 4-й степени.

Альтернативой аналитическим могут служить так называемые алгоритмические или итеративные методы оптимизации систем. Алгоритмические методы не нацелены на получение искомого результата в виде законченной математической формулы, а лишь указывают алгоритмы, т. е. последовательность операций, осуществление которых приводит к приемлемому по точности решению за ограниченное число шагов или итераций.

Алгоритмические методы возникли на почве численного решения с применением ЭВМ алгебраических уравнений и систем высокого порядка. Благодаря своим высоким эксплуатационным свойствам – простоте, универсальности, надежности, – а также в связи с повсеместным распространением быстродействующей цифровой вычислительной техники именно алгоритмические методы находят наиболее широкое применение при анализе и синтезе современных информационных систем. Более того, современная теория и практика анализа таких систем, или, говорят, **системного анализа**, – это в значительной своей части теория и практика применения алгоритмических методов оптимизации. Современный системный анализ как направление научных исследований и одновременно учебная дисциплина образует методологическую основу для решения актуальной проблемы оптимальности.

7.2. Идея итеративных методов

При анализе сложных систем **уравнение оптимальности** (1) представляет собой сложное **нелинейное уравнение** относительно неизвестного вектора параметров \vec{c} , и надежда получить его аналитическое решение отсутствует почти всегда, за исключением случая

квадратичной целевой функции (функционала). Если учесть, кроме этого, что размерность вектора \vec{c} , т. е. число скалярных уравнений:

$$\frac{dJ(\mathbf{c})}{dc_i} = 0, \quad i = \overline{1, N}, \quad (1^*)$$

при анализе сложных систем много превышает единицу, то сразу становится очевидной необходимость отказа от аналитических методов оптимизации.

Основная идея решения N -мерного оптимизационного уравнения (1*) с помощью регулярных итеративных методов состоит в следующем. Перепишем его в эквивалентной форме: $\nabla_c J(\mathbf{c}) = 0$ или

$$\mathbf{c} = \mathbf{c} - \gamma \nabla_c J(\mathbf{c}),$$

где $\nabla J(\mathbf{c})$ обозначает градиент, а γ – некоторый скаляр, т. е. число, и будем искать корень последнего матричного уравнения $\mathbf{c} = \mathbf{c}^*$ путем последовательных приближений или **итераций**:

$$\mathbf{c}(l) = \mathbf{c}(l-1) - \gamma_l \nabla J(\mathbf{c}(l-1)), \quad l = 1, 2, \dots, \quad (2)$$

где $\{\mathbf{c}(l)\}$ – l -е приближение вектора \mathbf{c}^* , или его приближение на l -ом шаге, $\nabla_c J(\mathbf{c}(l-1)) = 0$ – значение градиента на предыдущем шаге; γ_l – переменный в общем случае шаг итераций ($\mathbf{c}(0) = \mathbf{const}$ – начальное значение вектора \mathbf{c}).

В теории показано, что при правильном выборе шага итераций для любого начального приближения $\mathbf{c}(0)$ и произвольной формы функционала $J(\mathbf{c})$ последовательность итераций (2) **асимптотически сходится** к оптимальному решению, т. е.

$$\lim_{l \rightarrow \infty} \mathbf{c}(l) = \mathbf{c}^*.$$

В таком случае уравнение (2) определяет искомый **алгоритм последовательных приближений**. Методы решения уравнения

оптимальности (1*), основанные на таком алгоритме и отличающиеся только выбором шага γ_l , называются **итеративными методами оптимизации**. Их основное достоинство – удобство программирования на ЭВМ.

7.3. Разновидности итеративных методов

В общем случае в итеративном алгоритме вместо скалярной переменной γ_l используется матрица переменных усиления:

$$\mathbf{c}(l) = \mathbf{c}(l-1) - \mathbf{\Gamma}_l \nabla J(\mathbf{c}(l-1)), \quad l = 1, 2, \dots,$$

где $\mathbf{\Gamma}_l = \|\gamma_{lj} \mathbf{e}\|$, $i, j = \overline{1, N}$.

Если $\mathbf{\Gamma}_l = \mathbf{\Gamma} = \mathbf{const}$, говорят об алгоритмах с **постоянным шагом итераций**; в случае же $\mathbf{\Gamma}_l \neq \mathbf{const}$ получают алгоритмы с **переменным шагом**.

Простейшие алгоритмы со скалярным постоянным шагом $\mathbf{\Gamma} = \gamma_0 \mathbf{I}_N$, где $\mathbf{I}_N = \mathbf{diag}_N \{1\}$ – единичная матрица, нашли широкое применение в так называемых итеративных методах **градиентного типа**.

Для таких методов можно записать:

$$\mathbf{c}(l) = \mathbf{c}(l-1) - \gamma_0 \nabla J(\mathbf{c}(l-1)), \quad l = 1, 2, \dots$$

Примером алгоритмов с переменным матричным шагом может служить известный метод **Ньютона**:

$$\mathbf{c}(l) = \mathbf{c}(l-1) - [\nabla^2 J(\mathbf{c}(l-1))]^{-1} \nabla J(\mathbf{c}(l-1)), \quad l = 1, 2, \dots,$$

где $\nabla^2 J(\mathbf{c}(l-1))$ – матрица вторых производных функционала $J(\mathbf{c})$ размера $(N \times N)$.

Модифицированный метод Ньютона

$$\mathbf{c}(l) = \mathbf{c}(l-1) - [\nabla^2 J(\mathbf{c}(0))]^{-1} \nabla J(\mathbf{c}(l-1)), \quad l = 1, 2, \dots$$

использует алгоритм с матричным постоянным шагом $\mathbf{\Gamma} = \nabla^2 J(\mathbf{c})$ при $\mathbf{c} = \mathbf{c}(0)$.

Таким образом, различные виды итеративных методов отличаются друг от друга только выбором шага итераций γ_l .

7.4. Методы стохастической аппроксимации

Характерной особенностью функционирования сложных информационных систем является стохастическая (т. е. случайная) природа протекающих в них процессов, не поддающихся традиционным регулярным методам анализа. При выборе и формулировке функционала качества таких систем используются стохастические схемы математического описания, такие как модели конечных автоматов со случайными переходами, системы массового обслуживания, агрегативные модели и др. Математический синтез и анализ стохастических систем нуждается в специальных, статистических методах оптимизации.

Особенности этих методов явно присутствуют уже на этапе выбора и математической формулировки критерия оптимальности.

Пусть задан некоторый функционал $Q(\mathbf{c}, \mathbf{x})$, описывающий зависимость выбранного показателя качества системы от вектора ее оптимизируемых (внутренних) параметров \mathbf{c} и вектора внешних, не подлежащих оптимизации воздействий \mathbf{x} . Если природа вектора \mathbf{x} случайная (не поддается достоверному предсказанию), то и функционал $Q(\mathbf{c}, \mathbf{x})$ также является случайной величиной. Оценить эффективность функционирования системы с помощью такого показателя можно только в некотором усредненном или статистическом смысле. При этом говорят о статистическом подходе к задаче оптимизации.

Классической реализацией статистического подхода является вычисление функционала качества системы $J(\mathbf{c})$ по формуле математического ожидания зависимости $Q(\mathbf{c}, \mathbf{x})$ на множестве реализаций вектора \mathbf{x} с заданным законом распределения $f_n(\mathbf{x})$, т. е.

$$J(\mathbf{c}) = \int Q(\mathbf{c}, \mathbf{x}) f_n(\mathbf{x}) d\mathbf{x}, \quad (3)$$

или более кратко

$$J(\mathbf{c}) = \mathbf{M}_x \{Q(\mathbf{c}, \mathbf{x})\}.$$

Последняя формула определяет среднее статистическое значение показателя качества $Q(-)$ и выполняет поэтому роль статистического критерия оптимальности.

Если многомерная плотность распределения случайного вектора $f_n(\mathbf{x})$ задана, то задача поиска оптимального вектора параметров \mathbf{c}^* решается классическими методами математического анализа, в том числе итеративными методами.

Принципиальные же трудности возникают тогда, когда плотность $f_n(\mathbf{x})$ заранее неизвестна или она настолько сложна в формулировке, что интеграл (3) в элементарных функциях не вычисляется. Очевидно, что классические методы не охватывают ни одну из указанных ситуаций. Задача оптимизации в этих условиях нуждается, следовательно, в специальном математическом аппарате. Одной из современных разновидностей такого аппарата являются методы стохастической аппроксимации (СА).

Методы стохастической аппроксимации являются статистической разновидностью итеративных методов оптимизации, поэтому в их терминологии и ряде определений имеется много общего. Тем не менее, надо всегда помнить и об их принципиальном различии, а именно: если в регулярных (детерминистических) итеративных методах вид функционала качества $J(\mathbf{c})$ аналитически точно задан, то объектом методов СА являются системы с не известным в явном виде функционалом качества.

Для того чтобы изложить идею методов СА, обратимся к условию оптимальности вектора параметров системы вида

$$\mathbf{c}^* = \text{Arg}\{\nabla_{\mathbf{c}} J(\mathbf{c}) = 0\},$$

которое при учете (3) перепишем следующим образом:

$$\mathbf{c}^* = \text{Arg}\{\mathbf{M}_{\mathbf{x}}\{\nabla_{\mathbf{c}} Q(\mathbf{c}, \mathbf{x})\} = 0\}.$$

Здесь $\nabla_c Q(\mathbf{c}, \mathbf{x}) = \left\{ \frac{dQ(\mathbf{c}, \mathbf{x})}{dc_i}, i = \overline{1, M} \right\}$ – градиент функционала $Q(\mathbf{c}, \mathbf{x})$; c_i –

i -ый элемент оптимизируемого вектора \mathbf{c} , $i = \overline{1, M}$; M – размер вектора \mathbf{c} .

Матричное уравнение $\mathbf{M}_x \{ \nabla_c Q(\mathbf{c}, \mathbf{x}) \} = 0$ называется по аналогии с (1*) **уравнением оптимальности**.

В регулярном случае, когда плотность $f_n(\mathbf{x})$ задана, для его решения можно воспользоваться итеративной процедурой общего вида:

$$\mathbf{c}(l) = \mathbf{c}(l-1) - \Gamma_l \mathbf{M}_x \{ \nabla_c Q(\mathbf{c}(l-1), \mathbf{x}(l)) \}, \quad l = 1, 2, \dots, \quad (4)$$

где $\{\mathbf{c}(l)\}$ – последовательность приближений оптимального вектора \mathbf{c}^* ; $\{\mathbf{x}_l\}$ – последовательность реализаций случайного вектора \mathbf{x} на l -ом шаге оптимизаций; Γ_l – переменный матричный шаг итераций.

В стохастическом случае функция $f_n(\mathbf{x})$ неизвестна и, следовательно, правая часть процедуры (4) не может быть задана в явном виде. Известны лишь конкретные реализации случайного вектора \mathbf{x}_l , $l = 1, 2, \dots$. Центральная идея методов СА заключается в том, что при надлежащем выборе матрицы усиления Γ_l в процедуре (4) можно отказаться от операции статистического усреднения случайного градиента $\nabla_c Q(\mathbf{c}, \mathbf{x})$, используя вместо математического среднего $\mathbf{M}_x \{ \nabla_c Q(\mathbf{c}) \}$ текущее значение или реализацию такого градиента. При этом итеративный алгоритм определения оптимального вектора \mathbf{c} сводится к следующему разностному уравнению:

$$\mathbf{c}(l) = \mathbf{c}(l-1) - \Gamma_l \nabla_c Q(\mathbf{c}(l-1), \mathbf{x}(l)), \quad l = 1, 2, \dots \quad (5)$$

Учитывая, что переменная l в данном случае имеет смысл дискретного времени (это номер текущего отсчета вектора \mathbf{x}), монотонно возрастающего до бесконечности, из сопоставления уравнения (4) и (5) находим объяснение основной идеи методов СА: в них операция статистического усреднения случайного градиента $\nabla_c Q(\mathbf{c}, \mathbf{x})$ заменяется

неявным образом операцией усреднения во времени его последовательных реализаций $\nabla_c Q(\mathbf{c}(0), \mathbf{x}(1)), \nabla_c Q(\mathbf{c}(1), \mathbf{x}(2)), \dots$.

Пример. Рассмотрим трансверсальный фильтр первого порядка:

$$y(l) = x(l) + c_1 x(l-1), \quad (6)$$

на вход которого подается случайный процесс $x(1), x(2), \dots, x(l)$ с нулевым математическим ожиданием $M(X) = 0$ и постоянной дисперсией $D(X) = M(X^2(l)) = \sigma_x^2$. Здесь $l = 1, 2, \dots$ – дискретное время, c_1 – коэффициент или параметр фильтра. Оптимизируем его значение c^* из условия минимума дисперсии случайного отклика на выходе: $D(Y) = M(Y^2(l)) \rightarrow \min$. Фильтр (6) с таким свойством называется оптимальным фильтром предсказания ошибки (ФПО) первого порядка, который находит широкое применение в задаче прогнозирования случайных временных рядов.

Для решения поставленной задачи предварительно вычисляем:

- целевой функционал: $J(c_1) = M(Y^2(l))$;
- его текущее n -е значение: $Q(c_1, x) = y^2(l) = [x(l) + c_1 x(l-1)]^2 = x^2(l) + 2c_1 x(l) x(l-1) + c_1^2 x^2(l-1)$;
- градиент (первую производную) функционала качества: $\nabla_c Q(c_1, x) = 2x(l)x(l-1) + 2c_1 x^2(l-1)$.

В соответствии с выражением (5) получаем искомый алгоритм адаптивной оптимизации ФПО вида:

$c_1(l) = c_1(l-1) - 2\gamma x(l-1)[x(l) + c_1 x(l-1)] = c_1(l-1) - \tilde{\gamma} x(l-1)y(l), l=1, 2, \dots$,
 где $\tilde{\gamma} = 2\gamma = \text{const}$ – шаг адаптации. При этом для любого текущего момента времени $l = 1, 2, \dots$, величина: $x(l+1) = c_1(l)x(l)$ определяет, в свою очередь, адаптивный прогноз неизвестного $(l+1)$ -го отсчета анализируемого случайного процесса по данным его ретроспективных наблюдений $x(1), x(2), \dots, x(l)$ на интервале длиной l . При $l \rightarrow \infty$ полученный прогноз становится строго оптимальным в смысле минимума дисперсии его отклонения $\sigma_x^2 = M[x(l+1) - x(l+1)]^2$ относительно истинного (неизвестного заранее) значения $x(l+1)$.

Список рекомендуемой литературы

1. Высшая математика для экономического бакалавриата: Учебник и практикум: углубл. курс / Н.Ш. Кремер и др.; под ред. Н.Ш. Кремера. 4-е изд., перераб. и доп. М.: Юрайт, 2015. С. 909.
2. Савченко В.В. Теория вероятностей и математическая статистика: Конспект лекций. Н. Новгород: НГЛУ, 2003. С. 132.
3. Никольская В.А. Высшая математика. Часть I. Математический анализ Н. Новгород: НГЛУ, 2013. С. 122.
4. Никольская В.А., Родькина О.Я. Математика и информатика для студентов экономических специальностей. Н. Новгород: НГЛУ, 2008. С. 200.
5. Савченко В.В., Никольская В.А. Математика и информатика для студентов лингвистических специальностей Н. Новгород: НГЛУ, 2011. С. 137.
6. Никольская В.А. Учебное пособие по статистике для студентов экономических специальностей. Н. Новгород: НГЛУ, 2007. С. 74.
7. Савченко В.В. Основы теории вероятностей: Конспект лекций. Н. Новгород: НГЛУ, 2011. С. 127.
8. Фридман Л.М. Что такое математика. 2-е изд. М.: URSS, 2010. С. 191.
9. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика: Учебник. 3-е изд., перераб. и доп. М.: Кнорус, 2009. С. 375.
10. Шапкин А.С., Шапкин В.А. Задачи с решениями по высшей математике, теории вероятностей, математической статистике, математическому программированию: Учебное пособие. Электронный ресурс Интернет: URL: <http://biblioclub.ru/index.php?page=book&id=115811&sr=1>.

Владимир Васильевич Савченко,
Дмитрий Юрьевич Акатьев

АКТУАЛЬНЫЕ ГЛАВЫ ВЫСШЕЙ МАТЕМАТИКИ

**Учебно-методические материалы для студентов
социально-экономических направлений подготовки**

Редакторы: А.О. Кузнецова
Д.В. Носикова
А.С. Паршаков

Лицензия ПД № 18-0062 от 20.12.2000

Подписано к печати			Формат 60 x 90 1/16
Печ. л.	Тираж	экз.	Заказ
Цена договорная			

Типография НГЛУ
603155, Н. Новгород, ул. Минина, 31а