

**Федеральное государственное образовательное бюджетное  
учреждение высшего образования  
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ  
РОССИЙСКОЙ ФЕДЕРАЦИИ»  
(Финансовый университет)  
Новороссийский филиал  
Кафедра «Информатика, математика и общегуманитарные  
науки»**

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ПОДГОТОВКЕ  
К ГОСУДАРСТВЕННОЙ ИТОГОВОЙ АТТЕСТАЦИИ**

для студентов, обучающихся по направлению подготовки:

09.03.03 - Прикладная информатика, ОП «Инженерия данных»,

Профиль: «Инженерия данных»

*Рекомендовано Ученым советом Новороссийского филиала  
Финуниверситета (протокол № 20 от 28 февраля 2025 г.)*

*Одобрено кафедрой «Информатика, математика и  
общегуманитарные науки»  
(протокол № 7 от 28 февраля 2025 г.)*

**Новороссийск 2025**

## Содержание

1. Функции нескольких переменных и их экстремумы. Выпуклые множества и выпуклые функции нескольких переменных. Функции нескольких переменных в экономической теории.....	8
2. Определенный интеграл Римана. Свойства определенного интеграла. Применение определенного интеграла в экономических моделях.....	15
3. Степенные ряды. Теорема Абеля. Область сходимости степенного ряда. Интегрируемость и дифференцируемость суммы степенного ряда на интервале сходимости. Ряд Тейлора. Разложение функций в ряд Тейлора.....	20
4. Линейное пространство над полем действительных чисел. Базис и размерность линейного пространства. Подпространства линейного пространства.....	24
5. Алгебра матриц. Линейные преобразования. Собственные значения и собственные векторы.....	28
6. Евклидово пространство. Выпуклые множества. Отделяющая гиперплоскость. Разделяющая гиперплоскость.....	32
7. Метрические пространства. Теорема Банаха о неподвижной точке сжимающего отображения.....	36
8. Логика высказываний. Логика предикатов. Логические законы.....	38
9. Булевы функции. Важнейшие классы булевых функций. Теорема о полноте (теорема Поста).....	43
10. Устойчивость численных методов интегрирования систем дифференциальных уравнений. Проявления неустойчивости. Способы решения неустойчивых задач.....	48
11. Вероятностные пространства. Независимость событий. Формула полной вероятности. Формула Байеса.....	53

12. Центральная предельная теорема. Закон больших чисел. Применение центральной предельной теоремы.....	56
13. Линейные дифференциальные уравнения первого порядка и методы их решения. Задача Коши. Существование и единственность решения задачи Коши.....	58
14. Линейные уравнения $n$ -го порядка. Пространство решений линейного однородного уравнения.....	61
15. Модель множественной линейной регрессии: спецификация, предпосылки и их тестирование.....	65
16. Точечные и интервальные оценки параметров и эндогенной переменной, их применение в статистическом анализе эконометрических моделей.....	69
17. Динамические регрессионные модели: модели с распределенными лагами, авторегрессионные модели (спецификация, интерпретация параметров, методы оценки параметров).....	73
18. Системы одновременных уравнений: структурная и приведенная формы модели, проблема оценки структурных параметров, проблема идентификации, методы оценки структурных параметров.....	78
19. Антагонистические игры, их применение.....	83
20. Игры с природой. Матрица рисков. Принятие решений в условиях риска и неопределенности.....	87
21. Бескоалиционные неантагонистические игры, их применение.....	91
22. Производственная функция и ее свойства. Производственная функция с постоянной эластичностью замещения факторов производства и ее частные случаи.....	95
23. Задача управления производственным сектором экономики и структурная модель Леонтьева «затраты – выпуск».....	100
24. Реляционная модель данных. Нормализация отношений. Функциональные зависимости. Нормальные формы. Алгоритмы	

нормализации. Влияние нормализации на эффективность работы с базой данных.....	104
25. Оптимизация выполнения запросов на языке SQL. Непроцедурность запросов на языке SQL. Критерий и аргументы оптимизации запросов. Информация в базе данных для процесса оптимизации запросов. Средства отображения оптимальных планов в СУБД.....	123
26. Точечные оценки и их свойства (несмещенность, состоятельность и эффективность). Метод максимального правдоподобия и метод моментов.....	175
27. Проверка гипотез об определенных значениях параметров нормальных распределений. Проверка гипотезы о равенстве средних при известной дисперсии. Проверка гипотезы о равенстве средних при неизвестной дисперсии. Р-значение критерия. Определение и способ его вычисления.....	185
28. Критерии согласия. Проверка гипотезы о соответствии эмпирических данных теоретическому закону с данной функцией распределения по критерию Пирсона. Проверка гипотезы об однородности нескольких выборок по критериям Пирсона и Колмогорова – Смирнова.....	193
29. Сравнение нескольких средних методом дисперсионного анализа.....	199
30. Денежные потоки и их числовые характеристики. Чистая приведенная стоимость. Ренты. Объединение и замена рент. Инвестиционные потоки. Внутренняя норма доходности.....	201
31. Облигации и их числовые характеристики. Рыночная цена облигации и доходность к погашению. Курс облигации, премия и дисконт. Цена облигации как функция доходности к погашению и срока ее обращения.....	205

32. Дюрация потока платежей. Дюрация облигации и ее свойства. Выпуклость облигаций. Хеджирование риска изменений процентной ставки. Теорема об иммунизации.....	208
33. Портфель ценных бумаг, его доходность и риск. Эффективное множество портфелей. Оптимальный портфель при наличии безрискового актива.....	210
34. Биномиальная модель ценообразования активов. Оценка опционов с использованием биномиальной модели. Модель Кокса-Росса-Рубинштейна.....	213
35. Модель Блэка-Шоулза. Уравнение Блэка-Шоулза. Оценка производных финансовых инструментов с использованием модели Блэка-Шоулза.....	216
36. Договоры страхования с вычетом, лимитом и франшизой. Основные вероятностные и числовые характеристики.....	218
37. Индивидуальные модели риска. Вероятность разорения. Принципы назначения страховых премий.....	223
38. Коллективные модели риска. Составное пуассоновское распределение.....	227
39. Эмпирическое распределение для полных и сгруппированных данных. Огива и гистограмма. Оценка Нельсона-Алена кумулятивной функции доли риска.....	232
40. Эмпирическое распределение для цензурированных и усеченных данных. Оценка Каплана-Мейера.....	237
41. Оценка параметров распределений для полных данных. Метод моментов, метод эмпирических квантилей и метод максимального правдоподобия. Метод моментов для цензурированных данных.....	242
42. Оценка максимального правдоподобия для полных сгруппированных, усеченных и цензурированных данных.....	246
43. Модель ценообразования финансовых активов (CAPM), ее предпосылки и роль в анализе доходности акций.....	249

44. Касательный портфель в модели ценообразования финансовых активов (CAPM).....	251
45. Мера риска актива в модели ценообразования финансовых активов (CAPM).....	255
46. Градиентные методы поиска локального экстремума функции многих переменных: метод наискорейшего подъема (или спуска), метод проекции градиента, метод условного градиента.....	259
47. Постановка задачи целочисленного программирования. Метод Гомори. Метод ветвей и границ.....	263
48. Постановка задачи многокритериальной оптимизации. Метод целевого программирования. Метод идеальной точки.....	265
49. Постановка взаимно двойственных задач линейного программирования. Экономический смысл двойственности. Основная теорема двойственности и ее следствия.....	268
50. Общая постановка многокритериальной оптимизации. Парето-эффективная граница. Методы решения многокритериальной оптимизации.....	271
Практико-ориентированные задания.....	278
Рекомендуемая литература.....	340

1. *Функции нескольких переменных и их экстремумы. Выпуклые множества и выпуклые функции нескольких переменных. Функции нескольких переменных в экономической теории*

Пусть  $D \subset R^n$  и  $Y \subset R$  – числовые множества. Пусть указано правило, по которому каждому элементу  $(x_1; x_2; \dots; x_n) \in D$  сопоставляется некоторый (единственный) элемент  $y \in Y$ . Тогда говорят, что задано отображение или функция  $n$  переменных  $y = f(X)$  (используется также обозначение  $y = f(x_1; x_2; \dots; x_n)$ ). При этом множество  $D$  называется областью определения функции.

Пусть функция  $z = f(X)$  определена на множестве  $D \subset R^n$  и пусть  $X_0 \in D$ .

Точка  $X_0$  называется *точкой локального максимума* функции  $z = f(X)$ , если для каждой точки  $X$  из некоторой окрестности точки  $X_0$  выполняется неравенство  $f(X_0) \geq f(X)$ .

Точка  $X_0$  называется *точкой локального минимума* функции  $z = f(X)$ , если для каждой точки  $X$  из некоторой окрестности точки  $X_0$  выполняется неравенство  $f(X_0) \leq f(X)$ .

Точки локального максимума и локального минимума называются точками *экстремума* функции  $z = f(X)$ .

*Необходимое условие экстремума.* Если функция  $z = f(X)$  имеет в точке  $M(m_1, m_2, \dots, m_n)$  экстремум, то все частные производные первого порядка в этой точке равны нулю.

Таким образом, точки экстремума удовлетворяют системе уравнений

$$\frac{\partial z}{\partial x_1} = 0; \frac{\partial z}{\partial x_2} = 0; \dots; \frac{\partial z}{\partial x_n} = 0. \quad (1.1)$$

Условие (1.1) не является достаточным условием экстремума. Поэтому точки, удовлетворяющие системе уравнений (1.1), называются *точками возможного экстремума* или *стационарными*.

*Достаточное условие локального экстремума.* Пусть функция  $n$  переменных  $z = f(X)$  дифференцируема в некоторой окрестности точки  $M$

и дважды дифференцируема в самой точке  $M$ . Пусть, кроме того, точка  $M$  является стационарной точкой функции  $z = f(X)$ , то есть  $dz(M) = 0$ . Тогда, если  $d^2z(M)$ , второй дифференциал функции в точке  $M$ , представляет собой положительно определенную (отрицательно определенную) квадратичную форму от переменных  $dx_1, dx_2, \dots, dx_n$ , функция  $z = f(X)$  имеет в точке  $M$  локальный минимум (соответственно, локальный максимум). Если же второй дифференциал  $d^2z(M)$  является знакопеременной квадратичной формой, то функция  $z = f(X)$  не имеет локального экстремума в точке  $M$ .

Пусть функция  $z = f(X) = f(x_1; x_2; \dots; x_n)$  определена на множестве  $D \subset R^n$ , а  $P \subset D$  – подмножество в  $D$ .

Точка  $M \in P$  называется точкой *условного локального максимума* функции  $f(X)$ , если для любой точки  $X$ , достаточно близкой к точке  $M$  из подмножества  $P$ , выполняется условие:  $f(M) \geq f(X)$ .

Точка  $M \in P$  называется точкой *условного локального минимума* функции  $f(X)$ , если для любой точки  $X$ , достаточно близкой к точке  $M$  из подмножества  $P$ , выполняется условие:  $f(M) \leq f(X)$ .

Точки *условного локального максимума* и *условного локального минимума* называются точками *условного экстремума* функции  $z = f(X)$ .

Отличие *условного экстремума* от *обычного экстремума* состоит в том, что точка  $X$ , близкая к точке  $M$ , удовлетворяет дополнительному условию  $X \in P$ .

Чаще всего в приложениях подмножество  $P$  задается системой уравнений

$$g_1(X) = 0, g_2(X) = 0, \dots, g_s(X) = 0. \quad (1.2)$$

Уравнения (1.2) в этом случае называются *уравнениями связи*.

*Необходимое условие условного экстремума.* Пусть функции  $f(X)$ ,  $g_1(X), g_2(X), \dots, g_s(X)$  определены и имеют непрерывные частные производные в окрестности точки  $M$ , причем градиенты функций



$g_1(X), g_2(X), \dots, g_s(X)$  в точке  $M$  линейно независимы. Тогда, если  $M$  – точка условного экстремума функции  $f(X)$  при условии (1.2), то найдутся числа  $\lambda_1, \lambda_2, \dots, \lambda_s$ , для которых точка  $M$  – стационарная точка функции

$$L(X) = f(X) + \lambda_1 g_1(X) + \lambda_2 g_2(X) + \dots + \lambda_s g_s(X). \quad (1.3)$$

Функция (1.3) называется *функцией Лагранжа*, а числа  $\lambda_1, \lambda_2, \dots, \lambda_s$  – *множителями Лагранжа*.

*Достаточное условие условного экстремума.* Предположим, что в точке  $M$  выполнено необходимое условие условного экстремума. Пусть функции  $f(X)$  и  $g_k(X)$   $k = 1, \dots, s$  дважды дифференцируемы в точке  $M$ . Рассмотрим квадратную матрицу  $H$  порядка  $n + s$ :  $H = \begin{pmatrix} O & G \\ G^T & L \end{pmatrix}$ , где  $O$  –  $s \times n$  - матрица из нулей,  $G = \left( \frac{\partial g_i}{\partial x_k}(M) \right)$ ,  $L = \left( \frac{\partial^2 L}{\partial x_i \partial x_j}(M) \right)$  – матрица Гессе функции Ларанжа.

Обозначим угловые миноры матрицы  $H$  через  $H_1, H_2, \dots, H_{n+s}$ .

1) если знаки угловых миноров  $H_{2s+1}, H_{2s+2}, \dots, H_{n+s}$  совпадают со знаком числа  $(-1)^s$ , то  $M$  – точка условного минимума функции  $f(X)$  при условиях связи (1.2);

2) если знаки угловых миноров  $H_{2s+1}, H_{2s+2}, \dots, H_{n+s}$  чередуются, причем знак минора  $H_{2s+1}$  совпадает со знаком числа  $(-1)^{s+1}$ , то  $M$  – точка условного максимума функции  $f(X)$  при условиях связи (1.2).

Множество  $D \subset R^n$  называется *выпуклым*, если вместе с любыми двумя своими точками оно содержит и отрезок, соединяющий их: из того, что  $A, B \in D$ , следует, что точка  $X = \alpha A + \beta B$ , где  $\alpha, \beta \in [0; 1]$  и  $\alpha + \beta = 1$ , принадлежит множеству  $D$ . Важным свойством выпуклых множеств является то, что пересечение любого числа выпуклых множеств является выпуклым множеством.

Пусть функция  $z = f(X)$ , определена на выпуклом множестве  $D \subset R^n$ . Функция  $f(X)$  называется *выпуклой*, если все точки произвольного отрезка

с концами в точках графика функции  $z = f(X)$  лежат выше графика или на самом графике.

Формально это определение можно сформулировать так: функция  $z = f(X)$  называется *выпуклой* на множестве  $D$ , если для любых точек  $A, B \in D$  и любых двух чисел  $\alpha, \beta \in [0; 1]$ , таких что  $\alpha + \beta = 1$ , выполняется неравенство  $f(\alpha A + \beta B) \leq \alpha f(A) + \beta f(B)$ .

Если в тех же условиях, для любых двух различных точек  $A, B \in D$  выполняется неравенство  $f(\alpha A + \beta B) < \alpha f(A) + \beta f(B)$ , то функция  $z = f(X)$  называется *строго выпуклой* на множестве  $D$ .

Функция  $z = f(X)$  называется *вогнутой* (соответственно *строго вогнутой*) на выпуклом множестве  $D \subset R^n$ , если функция  $z = -f(X)$  является выпуклой (соответственно строго выпуклой) на множестве  $D$ . Вогнутые и строго вогнутые функции характеризуются неравенствами:

$$f(\alpha A + \beta B) \geq \alpha f(A) + \beta f(B) \quad \text{и} \quad f(\alpha A + \beta B) > \alpha f(A) + \beta f(B),$$
 соответственно. Для вогнутых функций все точки произвольного отрезка с концами в точках графика функции лежат ниже графика функции или на самом графике.

Выпуклые (вогнутые) функции обладают следующими свойствами:

- сумма выпуклых функций – выпуклая функция;
- сумма вогнутых функций – вогнутая функция;
- сумма строго выпуклых функций – строго выпуклая функция;
- сумма строго вогнутых функций – строго вогнутая функция.

Сформулируем условие, позволяющее устанавливать выпуклость или вогнутость функций.

Пусть функция  $z = f(X)$  имеет на выпуклом множестве  $D \subset R^n$  непрерывные частные производные второго порядка.

Если второй дифференциал  $d^2(X)$  в каждой точке  $X \in D$  выпуклого множества  $D$  является положительно определенной квадратичной формой, то функция  $f(X)$  является на  $D$  выпуклой. Если  $d^2(X)$  в каждой точке  $X \in$

$D$  является отрицательно определенной квадратичной формой, то функция  $f(X)$  на  $D$  вогнута. Отметим, что это утверждение не требует невырожденности второго дифференциала, а требует только его знакоопределенности в каждой точке множества. Если же второй дифференциал в каждой точке множества знакоопределенный и невырожденный, то функция строго выпуклая или строго вогнутая.

Для дважды дифференцируемой функции многих переменных можно сформулировать следующее достаточное условие строгой выпуклости и строгой вогнутости.

Пусть  $D \subset R^n$  – выпуклое открытое множество,  $z = f(x_1; x_2; \dots; x_n)$  – функция, имеющая в  $D$  непрерывные частные производные второго порядка,  $C = (C_{ij}) = \left( \frac{\partial^2 z}{\partial x_i \partial x_j} \right)$  – матрица Гессе (являющаяся симметричной, в силу теоремы о равенстве смешанных производных). Функция  $z = f(X)$  является строго выпуклой на  $D$ , если для каждой точки области  $D$  матрице  $C$  соответствует положительно определенная квадратичная форма в  $R^n$ , то есть (в силу критерия Сильвестра):

$$\Delta_1 = c_{11} > 0, \Delta_2 = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} > 0, \Delta_3 = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} > 0, \dots, \\ \Delta_n = \det C > 0. \quad (1.4)$$

Строгую вогнутость функции  $z = f(X)$  можно проверить, доказав строгую выпуклость функции  $u = -f(X)$ . Тогда условие (1.4) приобретает следующий вид:

$$\Delta_1 = c_{11} < 0, \Delta_2 = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} > 0, \Delta_3 = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} < 0, \dots, \\ (-1)^n \Delta_n = (-1)^n \det C > 0.$$

*Единственность локального минимума (максимума) у строго выпуклой (вогнутой) функции.* Если функция  $z = f(X)$  дифференцируема и строго выпукла (вогнута) на выпуклом множестве  $D \subset R^n$ , то она может

иметь локальный минимум (максимум) только в одной точке этого множества.

Экстремумы выпуклых и вогнутых функций носят глобальный характер в следующем смысле.

Точкой *глобального минимума (глобального максимума)* функции  $z = f(X)$  на множестве  $D \subset R^n$  называется такая точка  $M \in D$ , в которой функция  $z = f(X)$  принимает наименьшее (наибольшее) значение на множестве  $D$ . Точки глобального минимума и точки глобального максимума называют *точками глобального экстремума*.

*Глобальный характер экстремума выпуклой функции.* Если  $M$  – точка локального минимума (локального максимума) выпуклой (соответственно вогнутой) функции  $z = f(X)$  на выпуклом множестве  $D \subset R^n$ , то  $f(M)$  – наименьшее (наибольшее) значение функции  $f(X)$  на  $D$ . При этом, если функция  $f(X)$  строго выпукла (строго вогнута), то  $M$  – единственная точка глобального экстремума.

*О достижении выпуклой функцией глобального экстремума в стационарной точке.* Пусть  $f(X)$  – выпуклая (вогнутая функция) на выпуклом множестве  $D \subset R^n$  и пусть точка  $M \in D$  – стационарная точка функции  $f(X)$ . Тогда  $M$  – точка глобального минимума (максимума) функции  $f(X)$  на множестве  $D$ .

Основные виды функций нескольких переменных, встречающиеся в экономических задачах, – это функция полезности, производственная функция, функция спроса и другие.

*Производственные функции.* Производственные функции используются для анализа, планирования и прогноза деятельности отдельной организации, корпорации, отрасли, региона и даже целой страны. Производственная функция  $Q = Q(x_1; x_2; \dots; x_n)$  характеризует объем продукции, который может быть изготовлен на данном производстве при использовании ресурсов  $X_1; X_2; \dots; X_n$  в количествах

$x_1; x_2; \dots; x_n$  соответственно. Если  $n = 1$ , то эта функция называется однофакторной, если же  $n > 1$ , то – многофакторной. Область определения производственной функции в силу ее экономического смысла – это множество точек из  $R^n$  с неотрицательными координатами  $x_m$ . Значения производственной функции также неотрицательны.

Приведем примеры некоторых производственных функций.

Линейная производственная функция для случая  $n$  факторов  $(x_1; x_2; \dots; x_n)$  имеет вид

$$Q(x_1; x_2; \dots; x_n) = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n,$$

где  $a_0 > 0, a_1 > 0, a_2 > 0, \dots, a_n > 0$ .

Мультипликативные производственные функции для случая  $n$  факторов  $(x_1; x_2; \dots; x_n)$  имеют следующий вид:

$$Q(x_1; x_2; \dots; x_n) = a_0 x_1^{a_1} x_2^{a_2} \dots x_n^{a_n},$$

где  $a_0 > 0, a_1 > 0, a_2 > 0, \dots, a_n > 0$ .

Важным частным случаем двухфакторной производственной функции является функция Кобба–Дугласа. Объем производства  $Q$  связан с объемом используемого капитала  $K$  и затратами труда  $L$  следующим соотношением:  $Q(K, L) = aK^\beta L^{1-\beta}$ , где  $a > 0, 0 < \beta < 1$ . Производственная функция Кобба–Дугласа обладает тем свойством, что в производственной системе, описываемой этой функцией, возможно почти полное замещение труда капиталом и, наоборот, капитала трудом. Имеется в виду, что для любого положительного (даже очень малого) количества одного ресурса можно подобрать такое количество другого ресурса, что возможно произвести любой требуемый объем продукции. В общем случае это не так. Например, для производственной функции CES:  $Q(K, L) = (c_1 K^{-\rho} + c_2 L^{-\rho})^{-\beta/\rho}$ , где  $c_1 > 0, c_2 > 0, \rho > -1$ , полное или даже почти полное замещение одного ресурса другим невозможно.

*Функция полезности.* Пусть на рынке представлены товары нескольких видов. Функция полезности определена на множестве

возможных наборов товаров и принимает большее значение на том из двух наборов, который является более предпочтительным для потребителя. Если для потребителя два набора равноценны, то функция полезности принимает на них равные значения. Ясно, что функция полезности у каждого потребителя индивидуальна, поэтому на практике рассматривают усреднённую функцию полезности, характерную для более или менее широкой категории потребителей. Пусть  $x_m$  – количество товара вида  $m$  в наборе товаров. Тогда функция полезности представляет собой функцию  $n$  переменных:  $U(x_1; x_2; \dots; x_n)$ . Областью определения функции полезности является множество точек из  $R^n$  с неотрицательными координатами  $x_m$ .

Неоклассическая функция полезности имеет вид:

$$U(x_1; x_2) = x_1^\alpha x_2^\beta, \text{ где } \alpha > 0, \beta > 0, \alpha + \beta = 1.$$

Неоклассическая функция полезности общего вида:

$$U(x_1; x_2; \dots; x_n) = (x_1 - a_1)^{a_1} (x_2 - a_2)^{a_2} \dots (x_n - a_n)^{a_n},$$

где  $x_i > a_i > 0, i = 1, 2, \dots, n$ . Здесь вектор  $(a_1, a_2, \dots, a_n)$  может пониматься как минимально допустимый набор товаров.

*Логарифмическая функция полезности.* Эта функция полезности является результатом логарифмирования неоклассической функции полезности общего вида:

$$\begin{aligned} U_1(x_1; x_2; \dots; x_n) &= \ln U = \\ &= a_1 \ln(x_1 - a_1) + a_2 \ln(x_2 - a_2) + \dots + a_n \ln(x_n - a_n), \end{aligned}$$

где  $x_i > a_i > 0, i = 1, 2, \dots, n$ .

## **2. Определенный интеграл Римана. Свойства определенного интеграла. Применение определенного интеграла в экономических моделях.**

Пусть функция  $y = f(x)$  определена на отрезке  $[a; b]$ . Разобьем отрезок  $[a; b]$  точками  $a = x_0; x_1; \dots; x_n = b$  на  $n$  частичных отрезков  $[x_0; x_1], [x_1; x_2], \dots, [x_{n-1}; x_n]$ .

Пусть  $\Delta_k = x_k - x_{k-1}$  – длина частичного отрезка  $[x_{k-1}; x_k]$ . В каждом отрезке  $[x_{k-1}; x_k]$  выберем точку  $\xi_k$  и найдем значение  $f(\xi_k)$ .

Составим сумму

$$f(\xi_1)\Delta_1 + f(\xi_2)\Delta_2 + \dots + f(\xi_n)\Delta_n = \sum_{k=1}^n f(\xi_k)\Delta_k = \sigma_n.$$

Число  $\sigma_n$  называется *интегральной суммой* функции  $f(x)$  на отрезке  $[a; b]$ . Число  $\sigma_n$  определяется неоднозначно, а зависит от выбора точек разбиения  $x_k$  и от выбора точек  $\xi_k$  внутри частичных отрезков.

Пусть  $\lambda = \max\{\Delta_k\}$  – длина наибольшего отрезка разбиения.

Если множество интегральных сумм для  $f(x)$  на отрезке  $[a; b]$  имеет конечный предел при  $\lambda \rightarrow 0$ , то этот предел называется *определенным интегралом* от  $f(x)$  на отрезке  $[a; b]$  и обозначается

$$\int_a^b f(x)dx = \lim_{\lambda \rightarrow 0} \sigma_n.$$

При этом числа  $a$  и  $b$  называются нижним и верхним пределами интегрирования, а  $f(x)$  – подынтегральной функцией.

*Достаточное условие интегрируемости.* Если  $f(x)$  непрерывна на отрезке  $[a; b]$ , то она интегрируема на  $[a; b]$ , то есть интеграл  $\int_a^b f(x)dx$  существует.

*Геометрический смысл определенного интеграла.* Пусть  $f(x)$  непрерывна и положительна на отрезке  $[a; b]$ . Тогда интеграл  $\int_a^b f(x)dx$  равен площади криволинейной трапеции, ограниченной сверху графиком функции  $f(x)$ , снизу – осью  $Ox$ , слева – прямой  $x = a$ , справа – прямой  $x = b$ .

*Свойства определенного интеграла*

1.  $\int_a^b dx = b - a;$
2.  $\int_a^a f(x)dx = 0;$
3.  $\int_a^b f(x)dx = -\int_b^a f(x)dx;$
4. Для любых трех чисел  $a$ ,  $b$  и  $c$  справедливо равенство

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx;$$

$$5. \quad \int_a^b (f(x) \mp g(x))dx = \int_a^b f(x)dx \mp \int_a^b g(x)dx;$$

$$6. \quad \int_a^b c \cdot f(x)dx = c \cdot \int_a^b f(x)dx \quad (c = \text{const});$$

$$7. \quad \text{Если } f(x) \geq 0, \text{ при } x \in [a; b], \text{ то } \int_a^b f(x)dx \geq 0;$$

$$8. \quad \text{Если функции } f(x) \text{ и } g(x) \text{ интегрируемы на } [a; b] \text{ и } f(x) \geq g(x), \text{ при всех } x \in [a; b], \text{ то } \int_a^b f(x)dx \geq \int_a^b g(x)dx;$$

9. *Теорема об оценке определенного интеграла.* Если функция  $f(x)$  интегрируема на отрезке  $[a; b]$  и  $m \leq f(x) \leq M$ , при  $x \in [a; b]$ , то

$$m(b-a) \leq \int_a^b f(x)dx \leq M(b-a);$$

10. *Теорема о среднем значении.* Если функция  $f(x)$  непрерывна на отрезке  $[a; b]$ , то существует точка  $c \in [a; b]$ , такая что

$$\int_a^b f(x)dx = f(c)(b-a);$$

11. Определенный интеграл не зависит от переменной интегрирования.

*Определенный интеграл с переменным верхним пределом.* Пусть в определенном интеграле  $\int_a^t f(x)dx$  нижний предел интегрирования  $a$  фиксирован, а верхний предел  $t$  является переменной величиной из отрезка  $[a; b]$ . Придавая верхнему пределу интегрирования различные значения, будем получать соответствующие значения интеграла. Таким образом, интеграл  $\int_a^t f(x)dx$  является функцией своего верхнего предела, то есть  $\int_a^t f(x)dx = \Phi(t)$ .

*Свойства функции  $\Phi(t)$ :*

1) Если  $f(x)$  непрерывна на отрезке  $[a; b]$ , то функция  $\Phi(t) = \int_a^t f(x)dx$  непрерывна на отрезке  $[a; b]$ , то есть  $\lim_{t \rightarrow t_0} \Phi(t) = \Phi(t_0)$ .



2) Если  $f(x)$  непрерывна на отрезке  $[a; b]$ , то функция  $\Phi(t) = \int_a^t f(x)dx$  дифференцируема на  $[a; b]$ , причем  $\Phi'(t) = \left(\int_a^t f(x)dx\right)' = f(t)$ .

*Формула Ньютона-Лейбница.* Если  $f(x)$  непрерывна на отрезке  $[a; b]$  и  $F(x)$  – любая первообразная для  $f(x)$ , то справедлива формула Ньютона-Лейбница:

$$\int_a^b f(x)dx = F(b) - F(a) = F(x)|_a^b$$

*Доказательство.* Так как  $f(x)$  непрерывна на отрезке  $[a; b]$ , то  $f(x)$  интегрируема на  $[a; b]$ . Пусть  $t \in [a; b]$ , тогда  $f(x)$  интегрируема на  $[a; t]$ . Рассмотрим функцию  $\Phi(t) = \int_a^t f(x)dx$ . По свойству интеграла с переменным верхним пределом,  $\Phi'(t) = f(t)$ . Значит  $\Phi(t)$  – первообразная для  $f(t)$ . Пусть  $F(t)$  – любая другая первообразная для  $f(t)$ , тогда  $F(t) = \Phi(t) + C$ , где  $C = \text{const}$  (по теореме об общем виде первообразной). Имеем:

$$F(a) = \Phi(a) + C = \int_a^a f(x)dx + C = 0 + C = C, \text{ то есть } F(a) = C.$$

С другой стороны,

$$F(b) = \Phi(b) + C = \int_a^b f(x)dx + C = \int_a^b f(x)dx + F(a),$$

$$\text{то есть } F(b) = \int_a^b f(x)dx + F(a).$$

Значит,  $\int_a^b f(x)dx = F(b) - F(a)$ , что и требовалось доказать.

*Замена переменной в определенном интеграле.* Пусть функция  $f(x)$  непрерывна на отрезке  $[a; b]$ , а функция  $x = \varphi(t)$  дифференцируема на отрезке  $[\alpha; \beta]$ , и ее производная  $\varphi'(t)$  непрерывна на отрезке  $[\alpha; \beta]$ . Тогда, если множеством значений функции  $x = \varphi(t)$  является отрезок  $[a; b]$ , и  $\varphi(\alpha) = a$ ,  $\varphi(\beta) = b$ , то справедлива формула

$$\int_a^b f(x)dx = \int_a^b f(\varphi(t))\varphi'(t)dt.$$

*Формула интегрирования по частям для определенного интеграла.*  
Если функции  $u = u(x)$  и  $v = v(x)$  имеют непрерывные производные на отрезке  $[a; b]$ , то

$$\int_a^b u(x)v'(x)dx = u(x)v(x)|_a^b - \int_a^b v(x)u'(x)dx.$$

В экономических исследованиях изучают так называемые предельные величины, то есть для данной величины, выражаемой функцией  $f(x)$  рассматривают ее производную  $f'(x)$ . Например, если  $C(q)$  есть функция издержек в зависимости от объема  $q$  выпускаемого товара, то  $MC = C'(q)$  называют предельными издержками. Экономический смысл этой величины – это издержки на производство дополнительной единицы выпускаемого товара. Задача заключается в том, чтобы по функции предельных издержек  $MC$  необходимо найти функцию издержек  $C(q)$ . Функцию издержек находим интегрированием:  $C(q) = \int_{q_0}^q MCdq + C(q_0)$ .

Еще *примеры*. Пусть объем капитала крупной компании меняется непрерывно, то есть для каждого момента времени  $t_0 \leq t \leq T$  задана плотность потока инвестиций  $I(t)$  – сумма, поступающая на счет компании за единицу времени (то есть величина денежного потока за промежуток времени от  $t$  до  $t + dt$  приближенно равна  $I(t)dt$ ). Тогда общая сумма капитала в компании в момент времени  $t$  вычисляется по формуле

$$K = \int_{t_0}^T I(t) dt = K(T) - K(t_0).$$

Заметим, что поток  $I(t)$  может принимать и отрицательные значения, если вместе с инвестициями рассматривать и расходы компании.

При *непрерывном начислении процентов* годовая процентная ставка называется также силой роста и обозначается  $\delta$ . При начальной сумме  $S_0$ , сумма  $S(t)$ , накопленная к моменту времени  $t$ , вычисляется по формуле

$S(t) = S_0 e^{\delta t}$ . Если сила роста мало меняется по времени, то имеет место приближенное равенство  $\Delta S = \delta S \cdot \Delta t$ . Переходя к пределу, получаем  $\frac{dS}{dt} = S\delta$ . или  $\frac{dS}{S} = \delta dt$ . Интегрируя, получаем:

$$\int_{t_0}^t \frac{dS}{S} = \int_{t_0}^t \delta dt, \ln S(t) - \ln S(t_0) = \int_{t_0}^t \delta dt,$$

то есть  $\ln \frac{S(t)}{S(t_0)} = \int_{t_0}^t \delta dt$ . Следовательно,  $\frac{S(t)}{S(t_0)} = e^{\int_{t_0}^t \delta dt}$ , или

$$S(t) = S(t_0) \cdot e^{\int_{t_0}^t \delta dt}.$$

Величина  $e^{\int_{t_0}^t \delta dt}$  называется *коэффициентом наращения* и показывает, во сколько раз вырастает начальная сумма  $S_0$  за время  $t$ .

**3. Степенные ряды. Теорема Абеля. Область сходимости степенного ряда. Интегрируемость и дифференцируемость суммы степенного ряда на интервале сходимости. Ряд Тейлора. Разложение функций в ряд Тейлора.**

Ряд вида

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n \quad (3.1)$$

называется *степенным рядом*. Числа  $a_0, a_1, \dots, a_n, \dots$  называются *коэффициентами* степенного ряда. Придавая переменной  $x$  конкретные значения, получим различные числовые ряды, для каждого из которых можно ставить вопрос о его сходимости. *Областью сходимости* ряда (3.1) называется множество всех значений  $x$ , при которых ряд (3.1) сходится. Область сходимости степенного ряда – непустое множество, так как степенной ряд всегда сходится, по крайней мере, в точке  $x = 0$ .

*Теорема Абеля.* Если ряд (3.1) сходится при  $x = x_0 \neq 0$ , то он сходится, и притом абсолютно, для всех  $x$ , удовлетворяющих условию  $|x| < |x_0|$ . Если ряд (3.1) расходится при  $x = x_1$ , то он расходится при всех  $x$ , удовлетворяющих условию  $|x| > |x_1|$ .

Из теоремы Абеля следует, что существует такое положительное число  $r$ , что ряд (3.1) абсолютно сходится при  $x \in (-r; r)$  и расходится при  $x \notin [-r; r]$ . Число  $r$  называется *радиусом сходимости* степенного ряда (3.1), а интервал  $(-r; r)$  называется *интервалом сходимости*.

Используя признаки Даламбера и Коши, можно получить следующие формулы для нахождения радиуса сходимости степенного ряда:

$$r = \lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right|, r = \lim_{n \rightarrow \infty} \frac{1}{\sqrt[n]{|a_n|}}.$$

Область сходимости степенного ряда всегда состоит из его интервала сходимости и, быть может, граничных точек этого интервала. Исследовать степенной ряд на сходимость – значит найти его интервал сходимости и выяснить, сходится или расходится ряд в граничных точках интервала сходимости. При  $r = 0$ , область сходимости – единственная точка  $x = 0$ . При  $r = +\infty$ , область сходимости – вся числовая прямая.

Ряд вида

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n = a_0 + a_1 (x - x_0) + a_2 (x - x_0)^2 + \dots + a_n (x - x_0)^n \quad (3.2)$$

является обобщением ряда (3.1) и называется степенным рядом с центром в точке  $x_0$ . В частности, (3.1) – это степенной ряд с центром в точке 0. После замены  $t = x - x_0$  ряд (3.2) сводится к ряду вида (3.1):

$$\sum_{n=0}^{\infty} a_n t^n = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n. \quad (3.3)$$

Если  $t \in (-r; r)$  – интервал сходимости ряда (3.3), то  $x \in (-r + x_0; r + x_0)$  – интервал сходимости ряда (3.2).

В области сходимости степенного ряда (3.1) его сумма становится функцией переменной  $x$ . Поэтому в области сходимости степенного ряда можно записать  $f(x) = \sum_{n=0}^{\infty} a_n x^n$ . Про функцию  $f(x)$  в этом случае говорят, что она разложена в степенной ряд.

Пусть функция  $f(x)$  является суммой степенного ряда

$$f(x) = \sum_{n=0}^{\infty} a_n x^n, \quad (3.4)$$

с интервалом сходимости  $(-r; r)$ .

Перечислим основные свойства степенных рядов.

1. Если функция  $f(x)$  на интервале  $(-r; r)$  является суммой ряда (3.1), то она дифференцируема на интервале  $(-r; r)$  и ее производная  $f'(x)$  вычисляется почленным дифференцированием ряда (3.1), то есть,

$$f'(x) = \frac{d}{dx} (\sum_{n=0}^{\infty} a_n x^n) = \sum_{n=0}^{\infty} \frac{d}{dx} (a_n x^n).$$

Аналогично могут быть вычислены производные любого порядка. При этом все они имеют тот же интервал сходимости.

2. Если функция  $f(x)$  на интервале  $(-r; r)$  является суммой ряда (3.1), то она интегрируема на интервале  $(-r; r)$  и ее интеграл может быть найден почленным интегрированием:

$$\int_a^b f(x) dx = \int_a^b (\sum_{n=0}^{\infty} a_n x^n) dx = \sum_{n=0}^{\infty} \int_a^b a_n x^n dx.$$

При этом ряд, полученный интегрированием ряда (3.1) по отрезку  $[0; x]$ , где  $x \in (-r; r)$ , имеет тот же интервал сходимости, что и ряд (3.1).

3. Если функция  $f(x)$  на интервале  $(-r; r)$  разлагается в степенной ряд, то это разложение единственное. Более того, если это разложение существует, то оно имеет вид:

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n + \dots \quad (3.5)$$

Ряд, стоящий в правой части равенства (3.5), называется *рядом Маклорена* функции  $f(x)$ .

Для любой бесконечно дифференцируемой функции  $f(x)$  имеет место *формула Маклорена*:

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n + r_n(x), \quad (3.6)$$

где  $r_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}x^{n+1}$  – остаточный член, а  $c$  – число, находящееся между 0 и  $x$ .

Если обозначить через  $S_n(x)$  частичную сумму ряда Маклорена, то формулу (3.6) можно записать в следующем виде:

$$f(x) = S_n(x) + r_n(x). \quad (3.7)$$

*Необходимое и достаточное условие разложимости функции в ряд.*

Для того чтобы ряд Маклорена сходиллся на интервале  $(-r; r)$  и имел своей суммой функцию  $f(x)$ , необходимо и достаточно, чтобы последовательность  $\{r_n(x)\}$  стремилась к нулю для любого  $x \in (-r; r)$ , то есть  $\lim_{n \rightarrow \infty} r_n(x) = 0$  для любого  $x \in (-r; r)$ .

*Достаточное условие разложимости функции в ряд Маклорена).*

Пусть функция  $f(x)$  определена и бесконечно дифференцируема в интервале  $(-r; r)$ . Если существует такая постоянная  $M$ , что для любого  $x \in (-r; r)$  выполняются неравенства

$$|f^{(n)}(x)| < M, n = 0, 1, 2, \dots$$

то в этом интервале ряд Маклорена сходится к функции  $f(x)$ .

Формула Маклорена (3.6) дает разложение функции  $f(x)$  в степенной ряд с центром в точке 0. Если требуется разложить функцию  $f(x)$  по степеням  $(x - x_0)$ , то следует разложить  $f(x)$  в степенной ряд с центром в точке  $x_0$ . Соответствующий ряд называется *рядом Тейлора* и имеет следующий вид:

$$f(x) = f(0) + f'(0)(x - x_0) + \frac{f''(0)}{2!}(x - x_0)^2 + \dots + \frac{f^{(n)}(0)}{n!}(x - x_0)^n. \quad (3.8)$$

При практическом разложении функции в степенные ряды применяют следующие приемы.

1. *Непосредственное разложение функции  $f(x)$  в ряд Тейлора.* Вычисляют производные всех порядков функции  $f(x)$  в точке  $x_0$  и подставляют в разложение (3.8), то есть формально составляют ряд Тейлора для функции  $f(x)$ . Затем находят область сходимости полученного ряда.

2. *Использование табличных разложений.*

3. *Использование сложения и вычитания рядов и умножения ряда на число.* Иногда разложение функции в ряд достигается суммированием

табличных или ранее найденных разложений, а также умножением известного разложения на число.

4. *Использование дифференцирования и интегрирования рядов.* При разложении функции в ряд используют почленное дифференцирование и интегрирование рядов.

#### **4. Линейное пространство над полем действительных чисел. Базис и размерность линейного пространства. Подпространства линейного пространства.**

*Арифметическим  $n$ -мерным вектором* называется упорядоченный набор  $n$  чисел  $a_1, a_2, \dots, a_n$ , который обозначается через  $\bar{a} = (a_1, a_2, \dots, a_n)$ , а входящие в набор числа называют координатами вектора  $\bar{a}$ .

Иногда удобно записывать координаты вектора вертикально, или в виде *вектора-столбца*. Применим такую запись для пояснения двух операций над арифметическими векторами:

$$1. \text{ Сложение векторов: } \bar{a} + \bar{b} = \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \dots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \dots \\ a_n + b_n \end{pmatrix}.$$

$$2. \text{ Умножение вектора на число: } \lambda \cdot \bar{a} = \lambda \cdot \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \dots \\ \lambda a_n \end{pmatrix}, \text{ где } \lambda \in \mathbb{R}.$$

При этом вектор  $-\bar{a} = (-1) \cdot \bar{a}$  называется противоположным вектору  $\bar{a}$ , а вектор  $\bar{0} = (0, 0, \dots, 0)$  – нулевым.

Основных свойств операций над векторами – восемь, они напоминают аналогичные свойства операций с действительными числами  $\mathbb{R}$ :

$$B1. \forall \bar{a}, \bar{b}, \bar{c}: (\bar{a} + \bar{b}) + \bar{c} = \bar{a} + (\bar{b} + \bar{c});$$

$$B2. (\exists \bar{0})(\forall \bar{a}): \bar{a} + \bar{0} = \bar{a};$$

$$B3. (\forall \bar{a})(\exists -\bar{a}): \bar{a} + (-\bar{a}) = \bar{0};$$

$$B4. \forall \bar{a}, \bar{b}: \bar{a} + \bar{b} = \bar{b} + \bar{a};$$

$$B5. (\forall \bar{a}) \forall \lambda, \mu \in \mathbb{R}: \lambda \cdot (\mu \cdot \bar{a}) = (\lambda \cdot \mu) \cdot \bar{a};$$

$$B6. \forall \bar{a}: 1 \cdot \bar{a} = \bar{a};$$

$$B7. (\forall \bar{a}) \forall \lambda, \mu \in \mathbb{R}: (\lambda + \mu) \cdot \bar{a} = \lambda \cdot \bar{a} + \mu \cdot \bar{a};$$

$$B8. (\forall \bar{a}, \bar{b}) \forall \lambda \in \mathbb{R}: \lambda \cdot (\bar{a} + \bar{b}) = \lambda \cdot \bar{a} + \lambda \cdot \bar{b}.$$

Указанные свойства операций над арифметическими векторами, очевидно, выполняются, и называются *их аксиомами*.

Множество всех  $n$  – мерных арифметических векторов, в котором определены операции сложения векторов и умножение вектора на число, удовлетворяющие аксиомам B1-B8, называется  $n$  – мерным *арифметическим пространством* и обозначается через  $\mathbb{R}^n$ . Множество  $V$  элементов любой природы (например, направленных отрезков или функций), в котором определены две операции, удовлетворяющие аксиомам B1-B8, называется ***линейным пространством над полем действительных чисел  $\mathbb{R}$*** . Элементы множества  $V$  будем обозначать через  $v, w, \dots$  их принято также называть *векторами*, а для самого  $V$  часто применяют термин *векторное пространство*. Заметим, что линейное пространство над полем  $\mathbb{R}$  образует множество всех направленных отрезков на плоскости (в пространстве) с известными из геометрии операциями сложения (по правилу треугольника или параллелограмма) и умножения вектора на действительные числа, а также множество всех непрерывных на данном отрезке функций одной действительной переменной.

Дальнейшие рассуждения о линейных пространствах проведем для основного (модельного) случая – пространства  $\mathbb{R}^n$ . Выясним, каким образом информацию о всем пространстве векторов можно получать из некоторой (желательно, конечной) системы векторов. Это оказывается возможным, если задать базис линейного пространства.

Если  $\bar{a}_1, \bar{a}_2, \dots, \bar{a}_k$  – система  $k$  векторов, и  $\lambda_1, \lambda_2, \dots, \lambda_k$  – набор  $k$  числовых коэффициентов, то вектор  $\bar{a} = \lambda_1 \cdot \bar{a}_1 + \lambda_2 \cdot \bar{a}_2 + \dots + \lambda_k \cdot \bar{a}_k$  называется *линейной комбинацией векторов системы  $\bar{a}_1, \bar{a}_2, \dots, \bar{a}_k$* . Иначе говоря, вектор  $\bar{a}$  разлагается по векторам системы. Линейная комбинация



векторов системы с нулевыми коэффициентами всегда дает нулевой вектор и называется *тривиальной*:  $0 \cdot \bar{a}_1 + 0 \cdot \bar{a}_2 + \dots + 0 \cdot \bar{a}_k = \bar{0}$ .

Система векторов  $\bar{a}_1, \bar{a}_2, \dots, \bar{a}_k$  называется *линейно независимой*, если только ее тривиальная линейная комбинация равна нулевому вектору  $\bar{0}$ . В противном случае, т.е. для линейно зависимой системы векторов можно будет указать нетривиальную комбинацию (т.е. с отличным от нуля хотя бы одним коэффициентом), которая равна  $\bar{0}$ . Например, система двух векторов  $\bar{a}_1, \bar{a}_2$  линейно зависима в том и только в том случае, если они *коллинеарны*:  $\bar{a}_2 = \lambda \cdot \bar{a}_1, \lambda \in \mathbb{R}$ .

**Базисом линейного пространства  $V$**  называется такая упорядоченная система векторов  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k$ , которая является линейно независимой, и любой вектор  $\mathbf{v}$  пространства  $V$  разлагается по векторам этой системы:

$$\mathbf{v} = v_1 \cdot \mathbf{e}_1 + v_2 \cdot \mathbf{e}_2 + \dots + v_k \cdot \mathbf{e}_k.$$

Числовые коэффициенты  $v_1, v_2, \dots, v_k$  называют *координатами вектора  $\mathbf{v}$*  в базисе  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k$ . Они однозначно определены линейной комбинацией. Для арифметического пространства  $\mathbb{R}^n$  можно указать стандартный базис

$$\bar{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}; \bar{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \dots \\ 0 \end{pmatrix}; \dots; \bar{e}_n = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}, \text{ причем координаты вектора } \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix} \text{ в}$$

стандартном базисе: 
$$\begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix} = a_1 \cdot \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} + a_2 \cdot \begin{pmatrix} 0 \\ 1 \\ \dots \\ 0 \end{pmatrix} + \dots + a_n \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}$$

совпадают с его координатами  $a_1, a_2, \dots, a_n$  как числами заданного упорядоченного набора.

Таким образом, любой арифметический  $n$  – мерный вектор порождается (в виде некоторой линейной комбинации) конечным числом  $n$  векторов стандартного базиса. Говорят, что размерность арифметического пространства  $\mathbb{R}^n$  равна  $n$ :  $\dim \mathbb{R}^n = n$ . Для произвольного конечномерного линейного пространства (т.е. порожденного конечным числом базисных

векторов) *размерность линейного пространства* определяется как количество векторов произвольного его базиса (одинакового для всех базисов). Тем не менее, есть бесконечномерные линейные пространства: например, упомянутое выше пространство всех непрерывных на данном отрезке функций одной переменной.

*Подпространством линейного пространства  $V$*  называется (непустое) подмножество  $U$  векторов, удовлетворяющее двум условиям, называемым свойствами *замкнутости* относительно операций:

$$\text{П1. } \forall \mathbf{v}, \mathbf{w} \in U: \mathbf{v} + \mathbf{w} \in U;$$

$$\text{П2. } (\forall \mathbf{v} \in U) \forall \lambda \in \mathbb{R}: \lambda \cdot \mathbf{v} \in U.$$

Для подпространства будем использовать обозначение  $U \subset V$ .

Заметим, что подмножество  $U$ , которое наследует аксиомы B1-B8, само является линейным пространством. Тривиальными примерами подпространств служат само пространство  $V$  и один нулевой вектор.

Важным примером подпространства является совокупность всех линейных комбинаций заданной системы  $k$  векторов исходного линейного пространства. Для случая пространства  $\mathbb{R}^n$  ее записывают следующим образом:

$$\mathcal{L}(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_k) \subset \mathbb{R}^n,$$

и называют *линейной оболочкой* системы  $k$  векторов. Размерность линейной оболочки (как линейного пространства) называют *рангом системы векторов*

и обозначают следующим образом:  $\dim \mathcal{L}(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_k) = rk(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_k)$ .

Если записывать координаты векторов системы горизонтально, или в виде *векторов-строк*, то приведение такой матрицы элементарными преобразованиями к лестничному виду дает ответ на вопрос о ранге системы векторов:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kn} \end{pmatrix} \sim \dots \sim \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ \dots & \dots & \dots & \dots \\ 0 & b_{m2} & \dots & b_{mn} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Он равен числу ненулевых строк в лестничном виде:  $rk(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_k) = m$ .

## 5. Алгебра матриц. Линейные преобразования. Собственные значения и собственные векторы.

*Матрицей размера  $m \times n$*  называется прямоугольная таблица из чисел, содержащая  $m$  строк и  $n$  столбцов, которая обозначается через

$$A_{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \text{ или сокращенно, } A = (a_{ij}), \text{ где индексы}$$

$i = 1, \dots, m$  – номер строки элемента  $a_{ij}$  матрицы,  $j = 1, \dots, n$  – номер его столбца. Заметим, что каждая строка матрицы  $A$  – это вектор-строка из  $\mathbb{R}^n$ , а каждый столбец матрицы  $A$  – это вектор-столбец из  $\mathbb{R}^m$ .

Кроме двух операций над матрицами, напоминающими операции над арифметическими векторами:

(1) сложение матриц одинакового размера:

$$A_{m \times n} + B_{m \times n} = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij});$$

(2) умножение матрицы на число:  $\lambda \cdot A = \lambda \cdot (a_{ij}) = (\lambda \cdot a_{ij})$ , где  $\lambda \in \mathbb{R}$ , возможна третья операция над матрицами (но только в случае, если число  $n$  столбцов первой матрицы равно числу  $n$  строк второй матрицы). Эта операция называется *умножением матриц согласованных размеров*:

$$A_{m \times n} \cdot B_{n \times p} = (c_{ik}), \quad c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \dots + a_{in}b_{nk}; \quad k = 1, \dots, p.$$

Таким образом, произведение матриц состоит из элементов  $c_{ik}$ , которые являются стандартными скалярным произведениями каждой вектор-строки матрицы  $A = (a_{ij})$  на каждый вектор-столбец матрицы  $B = (b_{jk})$ .

Частный случай представляет класс *квадратных матриц порядка  $n$* :

$$A_{n \times n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}. \text{ Элементы такой матрицы с одинаковыми}$$

индексами  $a_{ii}, (i = 1, \dots, n)$  заполняют *главную диагональ*, вторую диагональ квадратной матрицы называют *побочной*. Среди квадратных

порядка  $n$  важна так называемая *единичная* матрица  $E_{n \times n} =$

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Основные свойства умножения матриц также напоминают аналогичные свойства операций с действительными числами  $\mathbb{R}$ :

$$M1. \forall A, B, C: (A \cdot B) \cdot C = A \cdot (B \cdot C);$$

$$M2. (\exists E)(\forall A): A \cdot E = A;$$

$$M3. \forall A, B, C: 1) (A + B) \cdot C = A \cdot C + B \cdot C;$$

$$2) A \cdot (B + C) = A \cdot B + A \cdot C;$$

$$M4. (\forall A, B) \forall \lambda \in \mathbb{R}: \lambda \cdot (A \cdot B) = (\lambda \cdot A) \cdot B = A \cdot (\lambda \cdot B).$$

Заметим, что при умножении матриц согласованных размеров в общем случае  $A \cdot B \neq B \cdot A$ .

Множество всех матриц размера  $m \times n$ , в котором определены операции сложения матриц и умножение матрицы на число (удовлетворяющие аксиомам B1-B8), а также операция умножения матриц согласованных размеров (удовлетворяющая аксиомам M1-M4), называется **алгеброй матриц размера  $m \times n$**  и обозначается через  $M_{m \times n}$ . Заметим, что относительно первых двух операций соответствующее множество матриц образует линейное пространство над полем  $\mathbb{R}$ , его часто обозначают через  $\mathbb{R}^{m \times n}$ . При этом  $\dim \mathbb{R}^{m \times n} = m \cdot n$ .

Пусть  $V$  — линейное пространство,  $\dim V = n$ . **Линейным преобразованием (оператором) пространства  $V$**  называется отображение  $f: V \rightarrow V$  пространства в себя, удовлетворяющее двум свойствам — *аксиомам линейности*:

$$Л1. \forall \mathbf{v}, \mathbf{w} \in V: f(\mathbf{v} + \mathbf{w}) = f(\mathbf{v}) + f(\mathbf{w});$$

$$Л2. (\forall \mathbf{v} \in V) \forall \lambda \in \mathbb{R}: f(\lambda \cdot \mathbf{v}) = \lambda \cdot f(\mathbf{v}).$$

Тривиальными примерами линейного преобразования служат линейный оператор с нулевым образом, т.е.  $\forall \mathbf{v} \in V: f(\mathbf{v}) = \mathbf{0}$ , и тождественный оператор:  $\forall \mathbf{v} \in V: f(\mathbf{v}) = \mathbf{v}$ . Важным примером линейного

преобразования является так называемое *проектирование* на  $k$  – мерное подпространство  $\mathcal{L}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k)$ , порожденное первыми  $k$  векторами базиса пространства  $(\mathbf{e}_1, \dots, \mathbf{e}_k, \mathbf{e}_{k+1}, \dots, \mathbf{e}_n)$ .

Фиксируем базис  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$  пространства  $V$ . Тогда линейное преобразование  $f: V \rightarrow V$  может быть задано квадратной матрицей  $A = (a_{ij})$ , состоящей из координат образов базисных векторов:

$$f(\mathbf{e}_j) = a_{1j}\mathbf{e}_1 + a_{2j}\mathbf{e}_2 + \dots + a_{nj}\mathbf{e}_n,$$

взятых по столбцам  $j = 1, 2, \dots, n$ . Матрица  $A = (a_{ij})$ , называемая *матрицей линейного преобразования  $f$  в базисе  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$* , позволяет выразить в этом базисе координаты вектора-столбца образа  $\mathbf{y} = f(\mathbf{x})$ :

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = f \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{n2} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{nn} \end{pmatrix},$$

$$\text{т.е.} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}.$$

Итак, имеем матричную форму действия линейного оператора:  $\mathbf{y} = A \cdot \mathbf{x}$ .

Рассмотрим важный случай *коллинеарности* вектора и его образа в линейном преобразовании. Ненулевой вектор  $\mathbf{x}$  называется **собственным вектором** линейного оператора  $f$ , если он преобразуется в коллинеарный вектор:

$$f(\mathbf{x}) = \lambda \cdot \mathbf{x}, \quad \lambda \text{ — число.}$$

При этом множитель пропорциональности  $\lambda$  называют **собственным значением** линейного оператора, *отвечающим* собственному вектору  $\mathbf{x}$ .

Поскольку  $f(\mathbf{x}) = A \cdot \mathbf{x}$ , то собственный вектор является решением матричного уравнения  $A \cdot \mathbf{x} = \lambda \cdot \mathbf{x}$ , которое с учетом согласования размеров матриц запишется как  $(A - \lambda \cdot E) \cdot \mathbf{x} = \mathbf{0}$ . Эта однородная система линейных уравнений с матрицей  $(A - \lambda \cdot E)_{n \times n}$  имеет ненулевое решение только, если

$$\det(A - \lambda \cdot E) = 0.$$

Раскрытие определителя приводит к многочлену  $n$ -й степени, который принято называть *характеристическим*, а его корни  $\lambda_1, \lambda_2, \dots, \lambda_n$  — искомые собственные значения.

Наконец, останется для каждого собственного значения  $\lambda_k$  решить отвечающую ему однородную систему  $(A - \lambda_k \cdot E) \cdot \mathbf{x} = \mathbf{0}$ . Искомые собственные подпространства будут определяться фундаментальным набором решений системы.

Если фиксирован базис  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ , в котором линейное преобразование  $f$  задается матрицей  $A = (a_{ij})$ , то для краткости говорят о собственных векторах (и собственных значениях) матрицы.

Линейный оператор в пространстве  $V$  над полем  $\mathbb{R}$  может не иметь собственных векторов. Однако комплексификация  $V(\mathbb{C})$  позволяет получить полезную информацию также о линейных операторах над полем  $\mathbb{R}$ . Рассмотрим оператор  $\varphi$  вращения плоскости на угол  $\frac{\pi}{2}$ . Если  $(\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2)$  — ортонормированный базис, то действие оператора задается правилом:  

$$\begin{cases} \varphi(\bar{\mathbf{e}}_1) = \bar{\mathbf{e}}_2, \\ \varphi(\bar{\mathbf{e}}_2) = -\bar{\mathbf{e}}_1. \end{cases}$$
 Значит, матрица оператора:  $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ .

Характеристический многочлен  $\begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1$  имеет два комплексно сопряженных корня  $\lambda_{1,2} = \mp i$ . Таким образом, собственные значения оператора вращения на угол  $\frac{\pi}{2}$  являются чисто мнимыми. Обе однородных системы для нахождения собственного вектора приводятся к одному уравнению с мнимыми коэффициентами:  $x_2 = \pm i \cdot x_1$ . Значит, собственные направления также задаются векторами с комплексными координатами  $\begin{pmatrix} 1 \\ \pm i \end{pmatrix}$ :  $\bar{\mathbf{x}}_{1,2} = \bar{\mathbf{e}}_1 \pm i \cdot \bar{\mathbf{e}}_2$ .

Итак, собственным значениям оператора вращения плоскости на угол  $\frac{\pi}{2}$   $\lambda_{1,2} = \mp i$  отвечают собственные векторы:  $\bar{\mathbf{x}}_{1,2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \pm i \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . Здесь они представлены как векторы-столбцы с действительной и мнимой частью.

## 6. Евклидово пространство. Выпуклые множества. Отделяющая гиперплоскость. Разделяющая гиперплоскость.

Скалярным произведением векторов линейного пространства  $V$  над полем  $\mathbb{R}$  называется функция  $(\cdot, \cdot): V \times V \rightarrow \mathbb{R}$ , удовлетворяющая четырем свойствам (аксиомам):

$$E1. \forall v, w \in V: (v, w) = (w, v);$$

$$E2. (\forall v, w \in V) \forall \lambda \in \mathbb{R}: (\lambda \cdot v, w) = (v, \lambda \cdot w);$$

$$E3. \forall u, v, w \in V: (u + v, w) = (u, w) + (v, w);$$

$$E4. \forall v \neq 0: (v, v) > 0.$$

Положительное число  $(v, v)$  называется *скалярным квадратом* вектора  $v$  и обозначается через  $v^2$ . Введение скалярного произведения обогащает линейное пространство, позволяя определить на нем понятия:

а) *длины вектора*  $v$ , которая обозначается через  $|v|$  и равна неотрицательному числу:  $|v| = \sqrt{v^2}$ ;

б) *угла между векторами*  $v$  и  $w$ , который обозначается через  $\widehat{vw}$  и определяется по формуле  $\cos \widehat{vw} = \frac{(v, w)}{|v| \cdot |w|}$ .

Важным является понятие *ортогональных векторов*  $v$  и  $w$ , если  $(v, w) = 0$ .

Линейное пространство над полем действительных чисел  $\mathbb{R}$  с фиксированной функцией скалярного произведения называется *евклидовым линейным пространством*.

В  $n$ -мерном арифметическом пространстве используется *стандартное* скалярное произведение  $(\bar{a}, \bar{b}) = a_1 \cdot b_1 + a_2 \cdot b_2 + \dots + a_n \cdot b_n$ , поэтому  $\mathbb{R}^n$  — евклидово. В нем действуют хорошо известные формулы длины вектора и угла между векторами:

$$|\bar{a}| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}; \quad \cos \widehat{\bar{a}, \bar{b}} = \frac{(\bar{a}, \bar{b})}{|\bar{a}| \cdot |\bar{b}|} = \frac{a_1 \cdot b_1 + a_2 \cdot b_2 + \dots + a_n \cdot b_n}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2} \cdot \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}}.$$

С другой стороны, в евклидовом линейном пространстве можно специализировать базис  $(e_1, e_2, \dots, e_n)$ , выбрав его, например, *ортонормированным*. Это означает, что любая пара векторов базиса

ортогональна, т.е.  $(\mathbf{e}_i, \mathbf{e}_j) = 0$  для  $\forall i \neq j$ , и каждый вектор нормирован:  $|\mathbf{e}_i| = 1$  для  $\forall i = 1, \dots, n$ . В ортонормированных базисах евклидова пространства большинство объектов «векторной природы» имеют достаточные простые координатные выражения.

Геометрия имеет дело с фигурами, т.е. множествами «точечной природы», поэтому с линейным пространством  $V$  необходимо ассоциировать точечное пространство  $T$ , элементы которого будем называть *точками*. В основе этого лежит операция *откладывания вектора от точки*, которая данной точки  $A \in T$  и вектору  $\mathbf{u} \in V$  ставит в соответствие новую точку  $B \in T$  такую, что  $A + \mathbf{u} = B$ .

*Аффинным (точечным) пространством*, ассоциированным с линейным пространством  $V$ , называется множество  $T$  с операцией  $T \times V \rightarrow T$  откладывания вектора от точки, удовлетворяющей трём условиям (аксиомам):

$$A1. (\forall \mathbf{v}, \mathbf{w} \in V) \forall A \in T: A + (\mathbf{v} + \mathbf{w}) = (A + \mathbf{v}) + \mathbf{w};$$

$$A2. \forall A \in T: A + \mathbf{0} = A;$$

$$A3. (\forall A, B \in T)(\exists! \mathbf{u} \in V): A + \mathbf{u} = B.$$

Вектор  $\mathbf{u}$  из аксиомы A3 называют *вектором, соединяющим точки A и B*, его принято обозначать  $\overline{AB} = \mathbf{u}$ . *Размерностью аффинного пространства* считают размерность линейного пространства, с которым оно ассоциируется.

Если в аффинном пространстве фиксировать начальную точку  $O$ , то можно отождествить каждую точку  $A \in T$  с ее радиус-вектором  $\overline{OA}$ . Поскольку вектор, соединяющий точки  $A$  и  $B$ , выражается разностью  $\overline{AB} = \overline{OB} - \overline{OA}$ , то он представляет собой разность точек  $B - A$ .

Начальная точка  $O$  вместе с базисом  $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$  пространства  $V$  образует *аффинную систему координат*, относительно которой дается координатное описание фигур.



Простейшие линейные фигуры в  $n$  – мерном аффинном пространстве – это прямые и  $k$  – плоскости ( $1 < k < n$ ). Прямой с начальной точкой  $A$  и направляющим вектором  $\mathbf{p}$  является фигура, состоящая из точек вида  $X = A + t \cdot \mathbf{p}$ , ( $t \in \mathbb{R}$ ), а  $k$  – плоскостью с начальной точкой  $A$  и направляющим подпространством  $\mathcal{L}(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k)$  является фигура – из точек вида

$$X = A + t_1 \cdot \mathbf{p}_1 + \dots + t_k \cdot \mathbf{p}_k \quad (t_1, \dots, t_k \in \mathbb{R}).$$

Если  $k = 2$ , то имеем собственно плоскость, а при  $k = n - 1$  используют название *гиперплоскость*. По аналогии со случаем  $k = 2$  координаты точек любой гиперплоскости  $\pi$  удовлетворяют одному линейному уравнению с  $n$  переменными и коэффициентами, не равными нулю одновременно:

$$A_1 \cdot x_1 + A_2 \cdot x_2 + \dots + A_n \cdot x_n + B = 0.$$

В свою очередь, *отрезок с концами в точках  $A$  и  $B$*  выделяется из точек прямой (с начальной точкой  $A$  и с направляющим вектором  $\overline{AB}$ ) условием на промежуток параметра  $t \in \mathbb{R}$ :

$$X = A + t \cdot \overline{AB} \quad (0 \leq t \leq 1).$$

Очевидно, текущая точка отрезка является линейной оболочкой его концов:  $X = t_1 \cdot A + t_2 \cdot B$  с коэффициентами  $0 \leq t_1, t_2 \leq 1$ , причем *выпуклой оболочкой*, т.е.  $t_1 + t_2 = 1$ . Отрезок – это простейший пример выпуклого множества.

Множество  $M \subset T$  аффинного пространства называется **выпуклым**, если вместе с любыми двумя точками оно полностью содержит отрезок с концами в этих точках.

Пересечение двух любых выпуклых множеств является выпуклым. Основа более сложной выпуклой фигуры точечного пространства состоит из непустого пересечения полупространств (с граничными гиперплоскостями). В результате получаются выпуклые многогранные области или выпуклые многогранники.

Аффинное пространство  $T$ , ассоциированное с евклидовым линейным пространством  $V$ , называется **евклидовым (точечным) пространством**.

Если базис  $(e_1, e_2, \dots, e_n)$  пространства  $V$  ортонормированный, то ему соответствует аффинная система координат, которая называется *прямоугольной декартовой*. Ряд евклидовых понятий удастся обобщить для  $n$  – мерного точечного пространства:

а) расстояние между точками  $A$  и  $B$  – это длина соединяющего их вектора  $\overline{AB}$ ;

б) проекцией точки  $M$  на гиперплоскость  $\pi$  является точка  $P$  такая, что вектор  $\overline{MP}$  ортогонален вектору, соединяющему любые две точки гиперплоскости  $\pi$ ;

в) расстояние от точки  $M$  до гиперплоскости  $\pi$  определяется как расстояние между точкой  $M$  и ее проекцией  $P$  на гиперплоскость.

Важное свойство *линейной отделимости* выпуклых множеств  $n$  – мерного евклидова (точечного) пространства основано на существовании гиперплоскостей двух видов: *отделяющих* и *разделяющих*.

**Теорема об отделяющей гиперплоскости.** Пусть выпуклое множество  $V \subset T$  аффинного пространства является компактом (т.е. ограниченным и замкнутым), а точка  $C \notin V$ . Тогда существует *отделяющая гиперплоскость*, проходящая через точку  $C$  так, что компакт  $V$  будет лежать от нее по одну сторону (рис.1).

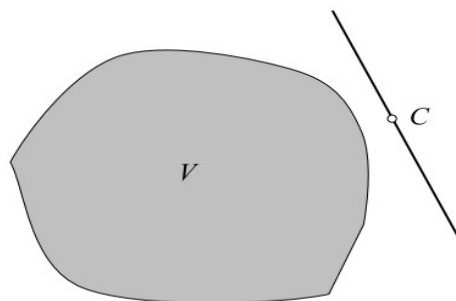


Рис.1.

**Теорема о разделяющей гиперплоскости.** Пусть два выпуклых множества  $V_1, V_2 \subset T$  аффинного пространства являются компактами без

общих точек. Тогда существует *разделяющая гиперплоскость*  $\pi \subset T$  и проходящая так, что компакты будут лежать от нее по разные стороны (рис.2).

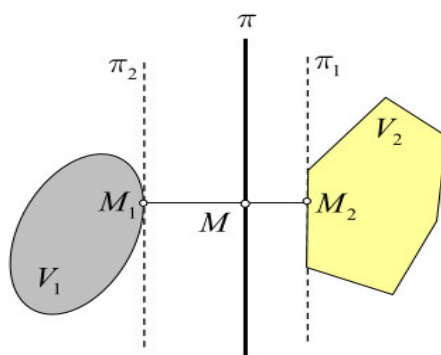


Рис.2.

## 7. Метрические пространства. Теорема Банаха о неподвижной точке сжимающего отображения.

Пусть  $X$  – множество. Функция  $d: X \times X \rightarrow [0, +\infty]$  называется **метрикой** на  $X$ , если она удовлетворяет следующим аксиомам:

1. отделимость:  $d(x, y) = 0 \Leftrightarrow x = y$ ;
2. симметричность:  $d(x, y) = d(y, x)$ ;
3. неравенство треугольника:  $d(x, y) \leq d(x, z) + d(z, y)$ .

Множество  $X$  с заданной на нем метрикой  $d$  называется **метрическим пространством**  $(X, d)$ .

Подмножество  $Y \subseteq X$ , рассматриваемое с той же самой метрикой  $d$ , называется **метрическим подпространством**  $(Y, d)$  пространства  $(X, d)$ .

Последовательность  $\{x_n\} \subseteq X$  называется **сходящейся в себе**, если для любого сколь угодно малого  $\varepsilon > 0$  при всех достаточно больших номерах  $n, m$  выполняется неравенство  $d(x_n, x_m) < \varepsilon$ .

Любая сходящаяся в себе последовательность ограничена, а любая сходящаяся (к некоторому пределу) последовательность является сходящейся в себе. Но, вообще говоря, не любая сходящаяся в себе последовательность сходится к некоторому пределу.

Отметим, что свойство последовательности быть сходящейся в себе не меняется при переходе от пространства к подпространству. Но последовательность, сходящаяся к пределу в пространстве, будучи рассмотренной в подпространстве, может уже не быть сходящейся к пределу, но в любом случае останется сходящейся в себе.

Метрическое пространство  $(X, d)$  называется **полным**, если в нем любая сходящаяся в себе последовательность сходится к некоторому пределу.

Любое метрическое пространство всегда можно рассматривать как подпространство некоторого полного пространства (используя процедуру пополнения).

Метрическое подпространство  $Y \subseteq X$  полного пространства  $(X, d)$  само полно ровно в том случае, когда оно замкнуто как подмножество в  $X$ .

Отображение  $f: X \rightarrow X$  метрического пространства  $X$  в себя называется **сжимающим**, если существует такое число  $C \in (0, 1)$ , что для любых  $x, y \in X$  справедливо неравенство  $d(f(x), f(y)) \leq C d(x, y)$ .

Сжимающее отображение является частным случаем липшицева отображения, в определении которого  $C \in (0, +\infty)$ .

Любое липшицево отображение непрерывно, т.е. переводит сходящуюся последовательность в сходящуюся последовательность. Но не любое непрерывное отображение является липшицевым.

Элемент  $a \in X$  называется **неподвижной точкой отображения**  $f: X \rightarrow X$ , если  $f(a) = a$ .

Если  $X$  – линейное пространство, то любое уравнение в  $X$  можно трактовать как задачу о поиске неподвижной точки некоторого отображения  $f: X \rightarrow X$ .

*Теорема Банаха.* Любое сжимающее отображение  $f: X \rightarrow X$  полного метрического пространства  $X$  в себя имеет единственную неподвижную точку  $a \in X$ , которая может быть получена как предел сходящейся

последовательности  $\{x_n\}$ , заданной рекуррентно  $x_{n+1} = f(x_n)$ , где  $x_0 \in X$  произвольно.

Если рассматриваемое отображение  $f: X \rightarrow X$  не является сжимающим во всем (полном) пространстве  $X$ , то можно перейти от  $X$  к некоторому его замкнутому подмножеству  $Y$ . При этом важна априорная информация о расположении неподвижной точки, поскольку в качестве  $Y$  имеет смысл брать ее малую замкнутую окрестность.

Теорема Банаха о сжимающем отображении является одной из важнейших теорем функционального анализа и имеет многочисленные приложения в других разделах математики. Отметим теорему о неявной функции, метод Ньютона приближенного решения уравнений и теорему о существовании и единственности решения задачи Коши для ОДУ.

Важным качеством теоремы Банаха как теоремы существования является ее **конструктивность**, т.е. свойство, состоящее в том, что эта теорема не только утверждает абстрактное существование некоторого объекта (корня уравнения), но и дает способ его построения (приближенного вычисления).

## **8. Логика высказываний. Логика предикатов. Логические законы.**

### *Высказывания и операции над ними*

В математике под высказыванием понимают утверждение, которому может быть приписано значение истинности. Обычно используют два значения: «истина» и «ложь», и говорят соответственно об истинности или ложности высказываний. Например, в арифметике высказывания « $1 \leq 2$ » и « $2 \cdot 2 = 4$ » считаются истинными, а высказывание « $0 = 1$ » — ложным.

Высказывания обозначаются, как правило, большими латинскими буквами; значения истинности — символами 0 (ложь) и 1 (истина). Значение истинности высказывания  $A$  обозначается через  $[A]$ . Запись

$[A] = 1$  означает, что высказывание  $A$  истинно, а запись  $[A] = 0$  — что высказывание  $A$  ложно.

**Отрицание.** Отрицание высказывания  $A$  обозначается через  $\bar{A}$  или через  $\neg A$ ; читается «не  $A$ ». Высказывание  $\neg A$  истинно, если  $A$  ложно, и ложно, если  $A$  истинно.

**Конъюнкция.** Конъюнкция высказываний  $A$  и  $B$  обозначается через  $A \wedge B$ ; читается « $A$  и  $B$ ». Высказывание  $A \wedge B$  истинно, если истинны оба высказывания  $A$  и  $B$ , и ложно, если ложно хотя бы одно из этих высказываний.

**Дизъюнкция.** Дизъюнкция высказываний  $A$  и  $B$  обозначается через  $A \vee B$ ; читается « $A$  или  $B$ ». Высказывание  $A \vee B$  ложно, если ложны оба высказывания  $A$  и  $B$ , и истинно, если истинно хотя бы одно из этих высказываний.

**Импликация.** Импликация высказываний  $A$  и  $B$  обозначается через  $A \rightarrow B$ ; читается «если  $A$ , то  $B$ », «из  $A$  следует  $B$ ». Высказывание  $A \rightarrow B$  ложно, если высказывание  $B$  ложно, а высказывание  $A$  истинно; во всех остальных случаях высказывание  $A \rightarrow B$  истинно. Высказывание  $A$  называют **посылкой** импликации  $A \rightarrow B$ , а высказывание  $B$  — **заключением**.

Определения операций можно свести в следующую таблицу:

$A$	$B$	$\bar{A}$	$A \wedge B$	$A \vee B$	$A \rightarrow B$
0	0	1	0	0	1
0	1	1	0	1	1
1	0	0	0	1	0
1	1	0	1	1	1

### *Формулы логики высказываний и логические законы*

При построении формул логики высказываний используются символы логических операций, скобки и символы  $X, Y, Z, \dots$  (быть может, с индексами) для обозначения *переменных высказываний* (*пропозициональных переменных*). Совокупность этих символов называется

**алфавитом логики высказываний.** Любая конечная последовательность символов алфавита называется **словом**. Среди всех слов выделяются те, которые построены по определенным правилам. Эти слова называются **формулами логики высказываний**.

Во-первых, формулой считается символ любой пропозициональной переменной. Во-вторых, если  $U$  и  $V$  — формулы, то  $(\neg U)$ ,  $(U \wedge V)$ ,  $(U \vee V)$ ,  $(U \rightarrow V)$  — также формулы. Слово в алфавите логики высказываний считается формулой, если оно получено в соответствии с этими правилами.

Формула называется **булевой**, если в ее записи используются только конъюнкция, дизъюнкция и отрицание (и не используется импликация).

Таблицы, подобные приведенной ниже, называются **таблицами истинности**.

$X$	$Y$	$\bar{X}$	$\bar{Y}$	$\bar{X}Y$	$X\bar{Y}$	$\bar{X}Y \vee X\bar{Y}$
0	0	1	1	0	0	0
0	1	1	0	1	0	1
1	0	0	1	0	1	1
1	1	0	0	0	0	0

*Определение.* Формулы  $U = U(X, Y, \dots, Z)$  и  $V = V(X, Y, \dots, Z)$  называются **равносильными**, если они принимают одинаковые значения для любой оценки переменных  $X, Y, \dots, Z$ .

*Определение.* Формула  $U$  называется **тождественно истинной** (тождественно ложной), если она истинна (ложна) для любой оценки входящих в нее переменных.

*Примеры.* Формулы

$$X \vee \bar{X}, \quad \overline{X \wedge \bar{X}}, \quad (X \wedge (X \rightarrow Y)) \rightarrow Y, \quad ((\bar{X} \rightarrow Y) \wedge (\bar{X} \rightarrow \bar{Y})) \rightarrow X$$

тождественно истинны. В этом несложно убедиться, построив таблицы истинности.

Тождественно истинные формулы остаются истинными независимо от того, какими высказываниями заменены входящие в них переменные. Они

соответствуют, в определенном смысле, некоторым универсальным логическим законам. Так, первая формула из предыдущего примера выражает так называемый **закон исключенного третьего**: из двух противоположных утверждений хотя бы одно истинно. Вторая формула — **закон противоречия**: два противоположных утверждения не могут быть истинными одновременно. Третья формула представляет собой **правило заключения**: из истинности посылки и импликации вытекает истинность заключения. Четвертая формула соответствует принципу **доказательства от противного**: утверждение верно, если из его отрицания следует одновременно некоторое заключение вместе со своим отрицанием.

### *Логика предикатов*

*Определение.* Будем говорить, что на множестве  $X$  задан **предикат**  $P(x_1, x_2, \dots, x_n)$  от переменных  $x_1, x_2, \dots, x_n$ , если каждому набору значений этих переменных из множества  $X$  поставлено в соответствие значение истинности: 1 (истина) или 0 (ложь).

Предикаты от одной переменной называют **одноместными**; от двух переменных — **двухместными** и т.д. Вообще,  $n$ -местный предикат можно рассматривать как отображение  $X^n$  в  $\{0,1\}$ .

Переменные в предикатах называют **предметными**, а множество, в котором они принимают значения, называют иногда **предметной областью**. Высказывания можно трактовать как нульместные предикаты, то есть постоянные предикаты, не зависящие от переменных.

**Одноместный** предикат  $P(x)$  на множестве  $X$  может трактоваться как **свойство**. Предмет  $x$  обладает свойством  $P$ , если  $P(x)$  истинно, и не обладает свойством  $P$ , если  $P(x)$  ложно.

**Двухместный** предикат  $P(x, y)$  на множестве  $X$  может трактоваться как **бинарное отношение**. Предметы  $x$  и  $y$  находятся в отношении  $P$ , если истинно  $P(x, y)$ .



*Определение.* Пусть  $P$  —  $n$ -местный предикат на множестве  $X$ . Множество всех тех наборов из  $X^n$ , для которых этот предикат истинен, называется **областью истинности** предиката  $P$ .

Область истинности предиката  $P$  будем обозначать через  $D_P$ .

*Кванторы существования и общности*

Пусть  $P(x)$  — предикат на множестве  $X$ . Определим высказывания  $\exists xP(x)$  (читается: «существует  $x$  такой, что  $P(x)$ ») и  $\forall xP(x)$  (читается: «для любого  $x$  верно  $P(x)$ »). Высказывание  $\exists xP(x)$  истинно, если область истинности предиката  $P(x)$  не пуста, и ложно в противном случае. Высказывание  $\forall xP(x)$  истинно, если область истинности предиката  $P$  совпадает с множеством  $X$ , и ложно в противном случае.

Равенства значений истинности

$$[\overline{\forall xP(x)}] = [\exists x\overline{P(x)}]; [\overline{\exists xP(x)}] = [\forall x\overline{P(x)}].$$

называют **законами де Моргана для кванторов**.

*Формулы логики предикатов и логические законы*

Некоторые формулы логики предикатов являются равносильными (или связаны отношением следования) в силу самой их формы, независимо от того, как они будут **проинтерпретированы**, то есть какова выбранная предметная область, какой «смысл» придан предикатным символам (то есть, какие предикаты подставлены в формулы вместо предикатных символов), какие предметы будут подставлены вместо предметных переменных. Подобного рода равносильности и следования выражают логические законы. В качестве примера можно привести рассмотренные в предыдущем пункте законы де Моргана для кванторов. Особый интерес представляют формулы, которые остаются истинными при любой интерпретации (такие формулы называют **общезначимыми**).

Важной проблемой логики вообще, и, в частности, логики предикатов, является поиск универсального алгоритма, позволяющего отличать

истинные суждения от ложных. Эту проблему в различных ее формулировках называют **проблемой разрешения**. Для логики предикатов проблема разрешения решается отрицательно. Американским математиком А. Черчем установлено, что не существует общего алгоритма для определения того, что произвольная формула логики предикатов общезначима.

Примеры логических законов:

$$\forall x(P(x) \wedge Q(x)) \Leftrightarrow \forall xP(x) \wedge \forall xQ(x);$$

$$\forall x(Q \rightarrow P(x)) \Leftrightarrow Q \rightarrow \forall xP(x);$$

$$\forall x\forall yP(x, y) \Leftrightarrow \forall y\forall xP(x, y);$$

$$\exists y\forall xP(x, y) \Rightarrow \forall x\exists yP(x, y).$$

*Формулы  $\forall x(P(x) \vee Q(x))$  и  $\forall xP(x) \vee \forall xQ(x)$  не равносильны.*

Для доказательства достаточно привести контрпример. На множестве целых чисел рассмотрим предикаты « $x$  — четное число» и « $x$  — нечетное число» и подставим их вместо соответственно  $P(x)$  и  $Q(x)$ . Имеем:

$$[\forall x((x - \text{четное число}) \vee (x - \text{нечетное число}))] = 1;$$

$$[\forall x(x - \text{четное число}) \vee \forall x(x - \text{нечетное число})] =$$

$$= [\forall x(x - \text{четное число})] \vee [\forall x(x - \text{нечетное число})] = 0 \vee 0 = 0.$$

В рассмотренной интерпретации формулы  $\forall x(P(x) \vee Q(x))$  и  $\forall xP(x) \vee \forall xQ(x)$  имеют разные значения истинности. Следовательно, эти формулы не равносильны.

## 9. Булевы функции. Важнейшие классы булевых функций. Теорема о полноте (теорема Поста).

*Определение.* Переменные, пробегающие множество  $\{0,1\}$ , мы будем называть булевыми. Функция от одной или нескольких булевых переменных, принимающая значение в множестве  $\{0,1\}$ , называется **булевой**.

Любую булеву функцию можно задать таблицей, подобной таблице истинности. Например, следующая таблица задает булеву функцию трех аргументов:

Булевы функции от одной переменной — это отображения множества  $\{0,1\}$  в себя. Булевы функции от одной переменной можно рассматривать как унарные операции на множестве  $\{0,1\}$ . В следующей таблице приведены все четыре булевы функции от одной переменной:

$x$	$g_0(x)$	$g_1(x)$	$g_2(x)$	$g_3(x)$
0	0	0	1	1
1	0	1	0	1

Булевы функции от двух переменных можно рассматривать как бинарные операции на множестве  $\{0,1\}$ . Имеется шестнадцать булевых функций от двух переменных (значения аргументов и функций выписаны в строках, функции перечисляются в столбце). Примеры (номера в двоичной записи соответствуют набору аргументов):

$x_1$	0	0	1	1	
$x_2$	0	1	0	1	
$f_0$	0	0	0	0	0 — тождественный ноль
$f_1$	0	0	0	1	·, $\wedge$ — умножение, конъюнкция
$f_6$	0	1	1	0	$\oplus$ — сложение по модулю 2
$f_7$	0	1	1	1	$\vee$ — дизъюнкция
$f_{14}$	1	1	1	0	$ $ — штрих Шеффера
$f_{15}$	1	1	1	1	1 — тождественная единица

Приведем некоторые уравнения, в которых одни булевы функции выражены через другие:

$$\bar{x} = x \rightarrow 0 = x|x = 1 \oplus x; x \vee y = x \oplus y \oplus xy; x \rightarrow y = \bar{x} \vee y; x \oplus y = x\bar{y} \vee \bar{x}y.$$

Все эти равенства легко проверяются с помощью таблиц.

*Определение.* **Двойственной** к булевой функции  $f(x, y, \dots, z)$  называется функция

$$f^*(x, y, \dots, z) = \overline{f(\bar{x}, \bar{y}, \dots, \bar{z})}.$$

Существует простой способ дать аналитическое представление функции в виде формулы, содержащей конъюнкции, дизъюнкции и отрицания, используя ее табличное задание. Например, пусть функция от трех переменных задана следующей таблицей:

$x$	$y$	$z$	$f(x, y, z)$
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

Тогда

$$f(x, y, z) = \bar{x} \bar{y} \bar{z} \vee x \bar{y} \bar{z} \vee x y z.$$

Получаемые при этом формулы называются **совершенными дизъюнктивными нормальными формами**, сокращенно **СДНФ**. Считается, что СДНФ тождественного нуля — это «пустая» дизъюнкция, не содержащая ни одного дизъюнктивного слагаемого.

Двойственная конструкция приводит к представлению функции в так называемой **совершенной конъюнктивной нормальной форме**, сокращенно **СКНФ**. СКНФ рассмотренной ранее функции имеет следующий вид:

$$f(x, y, z) = (x \vee y \vee \bar{z})(x \vee y \vee \bar{z})(x \vee \bar{y} \vee \bar{z})(\bar{x} \vee y \vee \bar{z})(\bar{x} \vee \bar{y} \vee z).$$

*Теорема.* Пусть  $f(x, y, \dots, z)$  — произвольная булева функция. Тогда

$$f(x, y, \dots, z) = x f(1, y, \dots, z) \vee \bar{x} f(0, y, \dots, z);$$

$$f(x, y, \dots, z) = (x \vee f(0, y, \dots, z))(\bar{x} \vee f(1, y, \dots, z)).$$

Последовательно применяя несколько раз (по числу переменных) разложение из предыдущей теоремы, можно получить СДНФ булевой функции. Например,

$$\begin{aligned} f(x, y) &= x \cdot f(1, y) \vee \bar{x} \cdot f(0, y) = \\ &= x \cdot (y \cdot f(1, 1) \vee \bar{y} \cdot f(1, 0)) \vee \bar{x} \cdot (y \cdot f(1, 1) \vee \bar{y} \cdot f(1, 0)) = \\ &= x \cdot y \cdot f(1, 1) \vee x \cdot \bar{y} \cdot f(1, 0) \vee \bar{x} \cdot y \cdot f(1, 1) \vee \bar{x} \cdot \bar{y} \cdot f(1, 0). \end{aligned}$$

*Определение.* **Элементарной конъюнкцией** (конъюнктом) называют конъюнкцию переменных и/или их отрицаний, в которой каждая переменная встречается не более одного раза. **Дизъюнктивной нормальной формой** (сокращенно ДНФ) называется дизъюнкция конечного числа конъюнктов.

*Определение.* **Элементарной дизъюнкцией** (дизъюнктом) называют дизъюнкцию переменных и/или их отрицаний, в которой каждая переменная встречается не более одного раза. **Конъюнктивной нормальной формой** называется конъюнкция конечного числа дизъюнктов.

Ранее было показано, что произвольная булева функция может быть записана с помощью конъюнкций, дизъюнкций и отрицаний. Это означает, что конъюнкция, дизъюнкция и отрицание образуют **полную систему функций**.

*Определение.* Система булевых функций называется **полной**, если любая булева функция может быть выражена через функции этой системы с помощью суперпозиций.

*Теорема.* Пусть  $K$  и  $K'$  — системы булевых функций, причем система функций  $K$  полна. Если каждая функция из  $K$  может быть представлена как суперпозиция функций из  $K'$ , то система функций  $K'$  также полна.

Через  $\oplus$  и 1 можно выразить отрицание, так что система функций  $\{1, \oplus\}$  также является полной. Последнее означает, что любая булева

функция может быть представлена в виде многочлена. При этом ненулевыми коэффициентами при одночленах служат единицы, а одночлены не содержат степеней переменных, поскольку умножение (конъюнкция) идемпотентна. Такие многочлены называют **полиномами Жегалкина**. Отрицание и дизъюнкция выражаются через штрих Шеффера, так что  $\{\bar{\phantom{x}}\}$  — тоже полная система функций.

*Определение.* Пусть  $K$  — некоторый класс булевых функций. Замыканием класса  $K$  называется множество всех тех функций, которые могут быть выражены через функции класса  $K$  с помощью суперпозиции. Класс функций называется **замкнутым**, если он совпадает со своим замыканием.

Важнейшие замкнутые классы булевых функций:

$P_0$  — класс всех функций, сохраняющих 0, то есть таких функций  $f(x_1, x_2, \dots, x_n)$ , для которых  $f(0, 0, \dots, 0) = 0$ ;

$P_1$  — класс всех функций, сохраняющих 1, то есть таких функций  $f(x_1, x_2, \dots, x_n)$ , для которых  $f(1, 1, \dots, 1) = 1$  ;

$S$  — класс всех самодвойственных функций, то есть таких функций  $f$ , которые совпадают со своей двойственной функцией,  $f^* = f$ ;

$L$  — класс всех линейных функций, то есть функций  $f(x_1, x_2, \dots, x_n)$ , представимых в виде  $f(x_1, x_2, \dots, x_n) = \alpha_0 \oplus \alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \dots \oplus \alpha_n x_n$ , где  $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n \in \{0, 1\}$ ;

$M$  — класс всех монотонных функций. Функция  $f(x_1, x_2, \dots, x_n)$  монотонна, если  $f(\alpha_1, \alpha_2, \dots, \alpha_n) \leq f(\beta_1, \beta_2, \dots, \beta_n)$  при  $(\alpha_1, \alpha_2, \dots, \alpha_n) \leq (\beta_1, \beta_2, \dots, \beta_n)$ .

*Теорема Поста о полноте.* Класс функций  $K$  полон тогда и только тогда, когда он не содержится целиком ни в одном из перечисленных выше пяти классов  $P_0, P_1, S, L, M$ .

Ни один из перечисленных пяти замкнутых классов не является полным (имеются не входящие в него булевы функции). Поэтому не может

быть полным ни один класс функций, целиком содержащийся в одном из них. Имеются булевы функции, не входящие ни в один из классов  $P_0$ ,  $P_1$ ,  $S$ ,  $L$ ,  $M$ . Любая такая функция в соответствии с теоремой Поста составляет полную систему функций, то есть через эту функцию может быть выражена любая булева функция. Среди функций от двух переменных такими функциями являются стрелка Пирса и штрих Шеффера.

С помощью теоремы Поста можно установить полноту системы функций, не выписывая непосредственно выражения для булевых функций.

*Пример.* Покажем, что система функций  $\{\rightarrow, \neg\}$  является полной. Составим таблицу, в которой символ 1 означает, что функция входит в соответствующий класс, а символ 0 – что не входит:

	$P_0$	$P_1$	$S$	$L$	$M$
$\rightarrow$	0	1	0	0	0
$\neg$	0	0	1	1	0

В каждом столбце таблицы имеется 0, значит, нет ни одного класса из пяти, который содержал бы обе функции. Следовательно, система функций  $\{\rightarrow, \neg\}$  полна.

## 10. Устойчивость численных методов интегрирования систем дифференциальных уравнений. Проявления неустойчивости. Способы решения неустойчивых задач.

Математическое моделирование любого динамического процесса приводит к необходимости решения и исследования дифференциальных уравнений.

Рассмотрим уравнение, разрешенное относительно старшей производной.

$$x^{(n)}(t) = f(t, x'(t), x''(t), \dots, x^{(n-1)}(t)),$$

Решить обыкновенное дифференциальное уравнение - означает найти функцию  $x(t)$ , которая непрерывна и непрерывно дифференцируема требуемое число раз на некотором промежутке изменения независимой переменной и при подстановке в уравнение обращает его в тождество.

Чаще всего в практических задачах независимая переменная  $t$  имеет смысл времени.

Пусть требуется найти такое решение уравнения, которое бы в некоторой фиксированной точке  $t_0$  (называемой начальной точкой) удовлетворяло бы условиям (называемым начальными условиями):

$$x(t_0) = x_0; x'(t_0) = x'_0; \dots; x^{(n-1)}(t_0) = x_0^{(n-1)}.$$

Такая задача называется **задачей Коши**.

Большинство практических задач не могут быть решены аналитически. Например, в случае, когда функции, стоящие в правой части системы, являются вычислимыми или заданы на сетке значений. Они также могут быть разрывны, недифференцируемы, содержать запаздывание и т.д.

Численные методы позволяют найти решение лишь приближенно. Однако они универсальны с точки зрения области применения и позволяют решать практически любые реальные задачи.

*Замечание.* Задача вычисления производных искомой функции обычно не ставится, поэтому проверка того факта, что полученная функция действительно является решением, подстановкой в уравнение, как, например, при решении алгебраических уравнений, невозможна.

Наиболее распространенными численными методами решения задачи Коши являются методы, которые позволяют найти приближенное решение лишь в отдельных точках. Отрезок интегрирования  $[t_0; T]$  разбивают на  $n$  частей (не обязательно равной длины):

$$t_0 < t_1 < \dots < t_n = T.$$



Точки деления являются *узлами сетки*. Задача состоит в вычислении значений искомой функции в узлах, то есть  $x_i = x(t_i)$ .

**Определение.** Величина  $h_i = t_{i+1} - t_i$  называется **шагом интегрирования**.

*Одношаговые* методы позволяют, зная  $t_k$  и значение  $x(t_k)$  определить значение  $x(t_{k+1})$ , где  $t_{k+1} = t_k + h_k$ . То есть сделать один шаг. В общем виде:

$$\tilde{x}_k = F(t_k, x_k, h_k, f),$$

где  $\tilde{x}_k$  - значение искомой функции в точке  $t_k$ , найденное с помощью численного метода. Пусть  $x(t_k)$  – точное значение.

**Определение.** Значение  $|\tilde{x}_k - x(t_k)| = \delta_n$  называется **локальной ошибкой** метода. Это ошибка, которая определяется только выбором метода (вычислительной схемы) на конкретном шаге, при условии, что все вычисления выполнены точно.

Каждый численный метод имеет свой *порядок*.

Численный метод характеризуется разбиением отрезка интегрирования (сеткой), величиной шага и порядком. Эти величины определяют скорость и точность решения и являются **параметрами метода**. Теоретически ошибка метода должна уменьшаться при уменьшении шага, либо при увеличении порядка.

Как показывает практика, далеко не все численные методы даже высокого порядка позволяют получить приемлемые с точки зрения точности результаты.

Метод называют **вычислительно устойчивым (устойчивым)**, если существует такая величина шага  $h_{кр} > 0$ , что при решении задачи (7.19) ( $Re \lambda < 0$ ) с шагом  $h \in (0; h_{кр})$  глобальная ошибка интегрирования ограничена. Величина  $h_{кр}$  называется *критическим шагом*.

Интервал  $(0; h_{кр})$  характеризует область устойчивости метода, то есть задает возможный диапазон шага интегрирования. Выход за этот диапазон приводит к нарастанию ошибки. Тогда говорят, что **метод расходится**.

Для сложных систем уравнений нахождение критического шага является самостоятельной и трудной задачей.

Рассмотрим простейшую задачу Коши для одного уравнения:

$$\frac{dx}{dt} = \lambda x, \quad x(t_0) = x_0.$$

Известно, что решением является функция  $x(t) = x_0 e^{\lambda(t-t_0)}$ . Пусть  $Re \lambda < 0$ , тогда решение асимптотически стремится к нулю.

Для сложных систем уравнений нахождение критического шага является самостоятельной и трудной задачей. Проиллюстрируем понятие области устойчивости для приведенной задачи.

Применим методы Рунге–Кутты 1-го, второго и т.д. порядка. Получим:

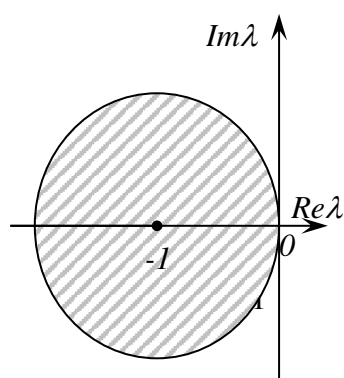
$$x_{r+1} = \left(1 + h \cdot \lambda + \frac{h^2}{2} \lambda^2 + \frac{h^3}{6} \lambda^3 + \frac{h^4}{24} \lambda^4\right) \cdot x_r.$$

Продолжая повышать порядок метода, можно убедиться, что для метода  $p$ -го порядка имеет место соотношение

$$x_{r+1} = \left(1 + h \cdot \lambda + \dots + \frac{h^p}{p!} \lambda^p\right) \cdot x_r,$$

то есть применяя метод Рунге-Кутты  $p$ -го порядка, мы получаем первые  $p$  членов разложения точного решения в ряд по шагу.

Теперь будем последовательно применять последнюю формулу, взяв в качестве  $x_0$  начальное условие. Получим:



$$x_1 = (1 + h \cdot \lambda + \dots + \frac{h^p}{p!} \lambda^p) \cdot x_0,$$

$$x_2 = (1 + h \cdot \lambda + \dots + \frac{h^p}{p!} \lambda^p) \cdot x_1 = (1 + h \cdot \lambda + \dots + \frac{h^p}{p!} \lambda^p)^2 \cdot x_0,$$

...,

$$x_{r+1} = (1 + h \cdot \lambda + \dots + \frac{h^p}{p!} \lambda^p)^r \cdot x_0.$$

Чтобы полученное решение было устойчивым, то есть чтобы  $x_r \rightarrow 0$  при  $r \rightarrow \infty$ , необходимо, чтобы выполнялось соотношение:

$$\left| 1 + h \cdot \lambda + \dots + \frac{h^p}{p!} \lambda^p \right| < 1.$$

Например, для метода Эйлера это означает, что  $|1 + \bar{\lambda}| < 1$ ,  $\bar{\lambda} = \lambda \cdot h$ .

Очевидно, что для систем с  $\bar{\lambda} > 0$  метод расходится при любом выборе шага. В остальных случаях выбор шага подчиняется не требованиям точности, а условию обеспечения устойчивости.

Если рассматривать не одно уравнение, а систему

$$\frac{dx}{dt} = Ax, \quad x \in R^n, \quad x(t_0) = x_0.$$

то ее решением, как известно, будет  $x = x_0 e^{At}$ . По определению

матричной экспоненты  $e^{At} = E + At + \frac{A^2 t^2}{2!} + \dots + \frac{A^p t^p}{p!} + \dots$  Здесь выбор

шага интегрирования будет определяться самым «плохим» собственным значением. Шаг должен быть таким, чтобы все собственные значения, умноженные на этот шаг, оказались в области устойчивости. Поэтому если с точки зрения точности расчета можно было бы менять шаг в сторону увеличения, то с точки зрения устойчивости этого делать нельзя. Известно, при увеличении порядка расширение области устойчивости происходит очень медленно.

Для нелинейных систем уравнений можно оценить вычислительную устойчивость на основании ее линеаризации. Каждую входящую в уравнения нелинейную функцию можно разложить в ряд Тейлора в окрестности начальных условий. Отбросив все члены степени выше первой, получим линейную систему, поведение которой будет близким к поведению исходной системы.

Систему дифференциальных уравнений, которая не допускает увеличения шага интегрирования в области плавного изменения параметров, называют **жесткой**. Такие системы требуют больших затрат для исследования. Мерой жесткости системы является число обусловленности матрицы линейной или линеаризованной системы. Число обусловленности растет с увеличением размерности математической модели, следовательно, при усложнении модели можно ожидать возникновения дополнительных вычислительных трудностей.

Один из способов борьбы с малой областью вычислительной устойчивости заключается в переходе к неявным схемам интегрирования. Общий вид процесса вычислений в этом случае имеет вид:

$$x_{r+1} = x_r + F(h, x_r, x_{r+1}),$$

то есть искомое значение функции в следующей точке входит как в левую, так и в правую часть соотношения. Поэтому каждый шаг интегрирования связан с решением системы уравнений, в общем случае нелинейных. Но выбор шага интегрирования определяется только требуемой точностью расчета.

В общем случае трудоемкость неявного метода существенно выше.

## **11. Вероятностные пространства. Независимость событий. Формула полной вероятности. Формула Байеса.**

*Определение.* События  $A_1, A_2, \dots$  образуют **полную группу** попарно несовместных событий (или просто полную группу), если при любом исходе опыта из них наступает ровно одно событие.

Пусть, как обычно,  $\Omega$  — пространство элементарных исходов опыта. Тот факт, что  $A_1, A_2, \dots$  образуют полную группу, можно записать в виде двух условий:

- $A_1 + A_2 + \dots = \Omega$ ;
- $A_i A_j = \emptyset, i \neq j$ .

Поскольку  $A_1, A_2, \dots$  — подмножества в  $\Omega$ , в теоретико-множественных обозначениях эти же условия выглядят таким образом:

- $\bigcup_{i=1,2,\dots} A_i = \Omega$ ;
- $A_i \cap A_j = \emptyset, i \neq j$ .

Следовательно, полная группа попарно несовместных событий — это то же самое, что разбиение  $\Omega$ .

*Примеры полных групп*

1)  $A$  и  $\bar{A}$ .

2)  $AB, \bar{A}B, A\bar{B}$  и  $\bar{A}\bar{B}$ .

3) Монета подбрасывается до тех пор, пока не выпадет орел.  $A_n$  — событие, состоящее в том, что будет произведено  $n$  бросков.  $A_1, A_2, \dots$  — полная группа из бесконечного числа событий.

*Теорема (формула полной вероятности).* Пусть  $H_1, H_2, \dots$  — полная группа событий, причем все вероятности  $P(H_i) > 0$ . Тогда для любого события  $A$  выполняется соотношение

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \dots$$

*Доказательство.* Имеем

$$A = A\Omega = A(H_1 + H_2 + \dots) = AH_1 + AH_2 + \dots$$

Так как  $H_1, H_2, \dots$  попарно несовместны, то попарно несовместны и  $AH_1, AH_2, \dots$  Следовательно,

$$\begin{aligned} P(A) &= P(AH_1) + P(AH_2) + \dots = \\ &= P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \dots \end{aligned}$$

В контексте формулы полной вероятности события  $H_1, H_2, \dots$  часто называются **гипотезами**. Таким образом, полнота группы гипотез, означает, что из данных гипотез выполняется ровно одна гипотеза.

*Теорема (формула Байеса).* Пусть выполнены условия предыдущей теоремы и  $P(A) > 0$ . Тогда для любого  $i = 1, 2, \dots$  имеем

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \dots}.$$

*Доказательство.* По определению условной вероятности

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

Применяя правило произведения для преобразования числителя и формулу полной вероятности для преобразования знаменателя, получим

$$P(H_i|A) = \frac{P(H_i A)}{P(A)} = \frac{P(H_i)P(A|H_i)}{P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \dots}.$$

*Пример.* Заемщики делятся на два класса  $H_1$  и  $H_2$ . Заемщик класса  $H_1$  возвращает долг с вероятностью 0,9. Заемщик класса  $H_2$  возвращает долг с вероятностью 0,2. Класс  $H_1$  составляет 70% всех заемщиков, а класс  $H_2$  — 30%. Какова вероятность, что случайный заемщик вернет долг?

*Решение.* Определяем события:

$$A = \{\text{Вернет долг}\}, H_i = \{\text{Заемщик} \in H_i\}.$$

Задача состоит в вычислении  $P(A)$ . Применяем формулу полной вероятности

$$\begin{aligned} P(A) &= P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \\ &= 0,7 \cdot 0,9 + 0,3 \cdot 0,2 = 0,63 + 0,06 = 0,69. \end{aligned}$$

*Пример.* В условиях предыдущего примера заемщик вернул долг. Какова вероятность, что он принадлежит классу  $H_1$ ?

*Решение.* Применяем формулу Байеса

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(H_1)P(A|H_1) + P(H_2)P(A|H_2)}.$$

Знаменатель и числитель мы уже вычислили в предыдущем примере, поэтому

$$P(H_1|A) = \frac{0,63}{0,69} = \frac{21}{23}.$$

## 12. Центральная предельная теорема. Закон больших чисел. Применение центральной предельной теоремы.

Пусть  $X_1, \dots, X_n$  — независимые случайные величины,  $S_n = X_1 + \dots + X_n$  — их сумма. **Стандартизованной суммой** называется случайная величина вида

$$S'_n = \frac{S_n - E(S_n)}{\sigma(S_n)}.$$

*Центральными предельными теоремами* в теории вероятностей принято называть теоремы, утверждающие при определенных условиях существование предельного распределения для стандартизованных сумм  $S'_n$ , когда  $n \rightarrow \infty$ .

*Центральная предельная теорема для одинаково распределенных случайных величин*

*Теорема.* Пусть  $X_1, X_2, \dots$  — последовательность независимых одинаково распределенных случайных величин, с математическим ожиданием  $E(X_i) = \mu$  и дисперсией  $D(X_i) = \sigma^2$ . Тогда для стандартизованных сумм

$$S'_n = \frac{S_n - E(S_n)}{\sigma(S_n)}$$

выполняется соотношение

$$\lim_{n \rightarrow \infty} P(S'_n < x) = N(x),$$

где  $N(x)$  — функция нормального распределения с параметрами 0 и 1,

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

ЦПТ позволяет достаточно просто обосновать интегральную приближенную формулу Лапласа.

Пусть производится неограниченная серия независимых испытаний с вероятностью успеха  $p$  в каждом испытании,  $S_n$  — число успехов в первых  $n$  испытаниях.

Случайную величину  $S_n$  можно представить как сумму

$$S_n = X_1 + \dots + X_n,$$

где  $X_i$  — индикатор успеха в  $i$ -м испытании.

В силу независимости испытаний, индикаторы  $X_1, X_2, \dots$  — независимые случайные величины, одинаково распределенные по закону  $Bin(1, p)$ .

Таким образом, условия ЦПТ выполняются, и остается только найти стандартизованные суммы  $S'_n$ .

Поскольку

$$S_n \sim Bin(n, p),$$

имеем

$$S'_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - np}{\sqrt{npq}},$$

где, как обычно,  $q = 1 - p$ .

При  $n \gg 1$  получаем

$$\begin{aligned} P(a < S_n < b) &= P\left(\frac{a - np}{\sqrt{npq}} < \frac{S_n - np}{\sqrt{npq}} < \frac{b - np}{\sqrt{npq}}\right) = \\ &= P\left(\frac{a - np}{\sqrt{npq}} < S'_n < \frac{b - np}{\sqrt{npq}}\right) \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right). \end{aligned}$$

*Центральная предельная теорема Ляпунова*

*Теорема.* Если  $X_1, X_2, \dots$  — последовательность независимых случайных величин, удовлетворяющая условию Ляпунова:

$$\lim_{n \rightarrow \infty} \frac{E(|X_1 - E(X_1)|^3) + \dots + E(|X_n - E(X_n)|^3)}{(D(X_1) + \dots + D(X_n))^{3/2}} = 0,$$

то для



$$S'_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - (E(X_1) + \dots + E(X_n))}{(D(X_1) + \dots + D(X_n))^{1/2}}$$

функции распределения стандартизованных сумм  $S'_n$  сходятся к  $N(x)$ ,

$$\lim_{n \rightarrow \infty} P(S'_n < x) = N(x).$$

*Замечание.* ЦПТ Ляпунова не предполагает, что  $X_1, X_2, \dots$  имеют одинаковое распределение.

### 13. Линейные дифференциальные уравнения первого порядка и методы их решения. Задача Коши. Существование и единственность решения задачи Коши.

Уравнение вида

$$y' + p(x)y = q(x) \quad (13.1)$$

где  $p(x)$  и  $q(x)$  – непрерывные функции, называется линейным. Соответствующее ему линейное однородное уравнение получается заменой правой части уравнения нулем:

$$y' + p(x)y = 0, \quad (13.2)$$

Для интегрирования линейных уравнений первого порядка используют метод вариации постоянной (другое название этого метода – метод Лагранжа). Применяется также метод подстановки (метод Бернулли).

Уравнение вида

$$y' + p(x)y = f(x)y^n, \quad n \neq 1, \quad (13.3)$$

называемое уравнением Бернулли, сводится к линейному заменой  $z = y^{1-n}$ . Решение уравнения Бернулли можно также найти, применяя метод вариации постоянной или метод подстановки.

*Метод вариации произвольной постоянной.* Метод вариации произвольной постоянной (другое название этого метода – метод Лагранжа) заключается в следующем. Сначала найдем решение уравнения

(13.2), которое представляет собой уравнение с разделяющимися переменными. Очевидно, что  $y = 0$  является решением уравнения (13.2).

$$\text{При } y \neq 0 \text{ имеем } \frac{dy}{y} = -p(x)dx.$$

Интегрируя это уравнение, получим,

$$\ln|y| = -P(x) + \ln|C|$$

где  $P(x) = \int p(x)dx$ , а  $C$  – отличная от нуля постоянная. Из последнего уравнения находим общее решение уравнения (2):

$$y = Ce^{-P(x)} \quad (13.4)$$

здесь  $C$  – произвольная постоянная, так как решение  $y = 0$  входит в (13.4) при  $C = 0$ .

Теперь варьируем постоянную, т.е. заменим в формуле (4) постоянную  $C$  на некоторую неизвестную функцию  $C(x)$ , т.е. общее решение уравнения (1) будем искать в виде

$$y = C(x)e^{-P(x)} \quad (13.5)$$

Подставим данное решение в уравнение (13.1):

$$C'(x)e^{-P(x)} + C(x)e^{-P(x)}(-p(x)) + p(x)C(x)e^{-P(x)} = q(x)$$

Откуда получаем, что

$$C'(x) = q(x)e^{P(x)}$$

Следовательно,

$$C(x) = \int q(x)e^{P(x)}dx \quad (13.6)$$

Подставив в (13.4) выражение (13.6) для  $C(x)$ , получим общее решение уравнения (13.1).

*Замечание.* В некоторых случаях дифференциальное уравнение может быть приведено к линейному, если поменять ролями переменные  $y$  и  $x$  – искомую функцию и ее аргумент.

*Метод подстановки (метод Бернулли)*

Решение уравнения (13.1) данным методом надо искать в виде

$$y(x) = u(x)v(x) \quad (13.7)$$

где  $u(x)$  и  $v(x)$  – некоторые функции, которые необходимо определить.

Подставляя данное выражение (13.7) в исходное уравнение (13.1), имеем

$$u'(x)v(x) + u(x)v'(x) + p(x)u(x)v(x) = q(x).$$

После группировки слагаемых оно принимает вид

$$u'(x)v(x) + u(x)(v'(x) + p(x)v(x)) = q(x) \quad (13.8)$$

Функция  $v(x)$  находится из уравнения

$$v'(x) + p(x)v(x) = 0.$$

Данное уравнение является уравнением с разделяющимися переменными. После подстановки найденной функции  $v(x)$  в уравнение (13.8), получается следующее уравнение для определения функции  $u(x)$

$$u'(x) = \frac{q(x)}{v(x)}.$$

После того, как функции  $u(x)$  и  $v(x)$  будут найдены, останется записать общее решение линейного уравнения первого порядка в виде (13.7).

Одной из важнейших задач теории дифференциальных уравнений является задача, в которой требуется найти решение, удовлетворяющее некоторому дополнительному условию. Чаще всего такое условие задается в виде начального условия  $y(x_0) = y_0$ .

Задача о нахождении решений дифференциального уравнения  $y' = f(x, y)$ , удовлетворяющих начальному условию  $y(x_0) = y_0$ , называется *задачей Коши*.

При решении конкретной задачи, связанной с моделированием реального процесса, вопрос о существовании решения задачи Коши, а также о его единственности, является исключительно важным. В общем случае задача Коши может иметь единственное решение, бесконечно много решений, либо вообще не иметь решений. Условия, при которых решение задачи Коши существует и единственно, формулируются в следующей теореме Коши.

*Теорема (о существовании и единственности решения задачи Коши).* Если в некоторой окрестности точки  $(x_0, y_0)$  функция  $f(x, y)$  определена, непрерывна и имеет непрерывную частную производную  $f'_y$ , то существует такая окрестность точки  $(x_0, y_0)$ , в которой задача Коши

$$y' = f(x, y)$$

$$y(x_0) = y_0$$

имеет решение, притом единственное.

#### **14. Линейные уравнения $n$ -го порядка. Пространство решений линейного однородного уравнения.**

Линейным дифференциальное уравнение  $n$ -го порядка называется уравнение вида

$$y^{(n)} + a_1(x)y^{(n-1)} + a_2(x)y^{(n-2)} + \dots + a_n(x)y = f(x), \quad (14.1)$$

где функции  $a_1(x), a_2(x), \dots, a_n(x), f(x)$  определены и непрерывны на некотором числовом множестве  $[a, b]$  (конечным или бесконечным, открытым или замкнутым).

Если  $f(x) \equiv 0$ , то уравнение (14.1) называется *линейным однородным уравнением*, в противном случае - *линейным неоднородным уравнением*.

Назовем линейным дифференциальным оператором  $n$ -го порядка следующее выражение:

$$L(x) = y^{(n)} + a_1(x)y^{(n-1)} + \dots a_n(x)y$$

*Лемма 1.* Пусть  $y_1(x)$  и  $y_2(x)$  – произвольные функции, имеющие производные до  $n$ -го порядка включительно,  $C_1$  и  $C_2$  – произвольные постоянные, тогда

$$L(C_1y_1 + C_2y_2) = C_1L(y_1) + C_2L(y_2)$$

Справедливость этого утверждения легко установить непосредственной проверкой.

Используя принятые выше обозначения, уравнение (14.1) можно записать в следующем виде

$$L(y) = f(x) \tag{14.2}$$

Уравнение

$$L(y) = 0 \tag{14.3}$$

называется *линейным однородным уравнением*, соответствующим уравнению (14.1). Следующее утверждение связывает решения уравнений (14.2) и (14.3).

*Теорема 1.* Общее решение неоднородного уравнения (14.2) есть сумма любого частного решения  $\tilde{y}(x)$  этого уравнения и общего решения  $y_0(x)$  соответствующего ему однородного уравнения (14.3).

*Теорема 2.* Пусть  $y_1(x)$  и  $y_2(x)$  – соответственно решения уравнений  $L(y) = f_1(x)$  и  $L(y) = f_2(x)$ , тогда  $y_1(x) + y_2(x)$  есть решение уравнения  $L(y) = f_1(x) + f_2(x)$ .

*Лемма 2.* Пусть  $y_1(x), y_2(x), \dots, y_k(x)$  – произвольные решения линейного однородного дифференциального уравнения и  $C_1, C_2, \dots, C_k$  – произвольные постоянные, тогда линейная комбинация  $C_1 y_1(x) + C_2 y_2(x) + \dots + C_k y_k(x)$  также является решением этого уравнения.

Действительно, на основании леммы 1 имеем:

$$L(C_1 y_1 + C_2 y_2 + \dots + C_k y_k) = C_1 L(y_1) + C_2 L(y_2) + \dots + C_k L(y_k) = 0,$$

что и требовалось доказать.

*Определение.* Система функций  $y_1(x), y_2(x), \dots, y_k(x)$  называется *линейно зависимой на некотором множестве  $D \subset R$* , если существуют такие числа  $a_1, a_2, \dots, a_k$ , одновременно не равные нулю, что линейная комбинация

$$a_1 y_1(x) + a_2 y_2(x) + \dots + a_k y_k(x) = 0 \text{ для всех } x \in D.$$

В противном случае система функций называется *линейно независимой на  $D$* .

*Определителем Вронского* для системы функций  $y_1(x), y_2(x), \dots, y_k(x)$  называется определитель вида

$$W = \begin{vmatrix} y_1 & y_2 & \dots & y_k \\ y_1' & y_2' & \dots & y_k' \\ \dots & \dots & \dots & \dots \\ y_1^{(k)} & y_2^{(k)} & \dots & y_k^{(k)} \end{vmatrix}$$

Для определения линейно зависима ли система функций применимы следующие теоремы.

*Теорема 3.* Если функции  $y_1(x), y_2(x), \dots, y_k(x)$  линейно зависимы на  $D$ , то их определитель Вронского тождественно равен нулю на этом множестве.

*Теорема 4.* Если  $y_1(x), y_2(x), \dots, y_k(x)$  – линейно независимые решения линейного однородного уравнения, то их определитель Вронского ни при одном значении  $x$  не обращается в нуль.

Система  $n$  линейно независимых решений уравнения (14.3)  $y_1(x), y_2(x), \dots, y_n(x)$  называется *фундаментальным набором решений* этого уравнения.

Итак, если  $y_1(x), y_2(x), \dots, y_n(x)$  – решения уравнения (14.3), то любая их линейная комбинация также является решением этого уравнения.

Для получения общего решения линейного однородного дифференциального уравнения этого уравнения существует следующая теорема.

*Теорема 5 (об общем решении линейного однородного уравнения).* Общим решением линейного однородного уравнения (3) является линейная комбинация

$$y(x) = C_1 y_1(x) + C_2 y_2(x) + \dots + C_n y_n(x) \quad (14.4)$$

$n$  линейно независимых частных решений этого уравнения  $y_1(x), y_2(x), \dots, y_n(x)$ .

*Доказательство.* То, что функция  $y(x)$ , определяемая формулой (14.4), является решением уравнения (14.3), следует из леммы 2. Покажем теперь, что любое решение  $g(x)$  уравнения (14.3) представимо в виде линейной комбинации функций  $y_1(x), y_2(x), \dots, y_n(x)$ . Для некоторой фиксированной точки  $x_0$  рассмотрим систему линейных алгебраических уравнений:

$$\begin{cases} C_1 y_1(x_0) + C_2 y_2(x_0) + \dots + C_n y_n(x_0) = g_0 \\ C_1 y_1'(x_0) + C_2 y_2'(x_0) + \dots + C_n y_n'(x_0) = g_0' \\ \dots \\ C_1 y_1^{(n)}(x_0) + C_2 y_2^{(n)}(x_0) + \dots + C_n y_n^{(n)}(x_0) = g_0^{(n-1)} \end{cases} \quad (14.5)$$

Определителем этой системы является определитель Вронского для функций  $y_1(x), y_2(x), \dots, y_n(x)$  в точке  $x_0$ , который в силу их линейной независимости не равен нулю. Следовательно, система (14.5) имеет единственное решение  $(\tilde{C}_1, \tilde{C}_2, \dots, \tilde{C}_n)$ .

Тогда функция  $g(x) = \tilde{C}_1 y_1(x) + \tilde{C}_2 y_2(x) + \dots + \tilde{C}_n y_n(x)$

удовлетворяет тем же начальным условиям, что и функция (4). В силу единственности решения задачи Коши имеем

$g(x) = y(x)$ , т.е.  $g(x)$  есть линейная комбинация функций  $y_1(x), y_2(x), \dots, y_n(x)$ .

## 15. Модель множественной линейной регрессии: спецификация, предпосылки и их тестирование.

Спецификация модели множественной линейной регрессии имеет вид:

$$Y_t = \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_j X_{tj} + \dots + \beta_k X_{tk} + \varepsilon_t = \sum_{j=1}^k \beta_j X_{tj} + \varepsilon_t, \quad (15.1)$$

где  $Y_t$  — эндогенная (зависимая) переменная,  $X_{tj}$  —  $j$ -ый регрессор в наблюдении  $t$ ,  $\beta_j$  —  $j$ -ый параметр модели,  $\varepsilon_t$  — случайное возмущение,  $k$  — число параметров,  $n$  — число наблюдений. В модели со свободным членом  $X_{t1}$  — единичный регрессор. В соответствии со спецификацией (15.1) выборочные данные  $Y_t, X_{tj}, t = 1, \dots, n, j = 1, \dots, k$  удовлетворяют системе уравнений наблюдений — схеме Гаусса-Маркова:

$$\begin{array}{rcll} Y_1 & = & \beta_1 X_{11} + \beta_2 X_{12} + \dots + \beta_j X_{1j} + \dots + \beta_k X_{1k} & + \varepsilon_1 \\ Y_2 & = & \beta_1 X_{21} + \beta_2 X_{22} + \dots + \beta_j X_{2j} + \dots + \beta_k X_{2k} & + \varepsilon_2 \\ \vdots & \vdots & \vdots & \vdots \\ Y_n & = & \beta_1 X_{n1} + \beta_2 X_{n2} + \dots + \beta_j X_{nj} + \dots + \beta_k X_{nk} & + \varepsilon_n \end{array}$$

или в матричной форме<sup>1</sup>

$$Y = X\beta + \varepsilon, \quad (15.2)$$

где  $Y = (Y_1, Y_2, \dots, Y_n)^T$  —  $(n \times 1)$ -вектор-столбец значений эндогенной переменной,  $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$  —  $(k \times 1)$ -вектор-столбец параметров

<sup>1</sup> Значком « $T$ » обозначена операция транспонирования.



модели,  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$  —  $(n \times 1)$ -вектор-столбец случайных возмущений.

В классической нормальной линейной регрессионной модели принимаются следующие предпосылки:

- $X_{n \times k} = \begin{bmatrix} X_{11} & \dots & X_{1k} \\ \vdots & & \vdots \\ X_{n1} & \dots & X_{nk} \end{bmatrix}$  — детерминированная матрица регрессоров

полного ранга —  $\text{rank}(X) = k$ ;  $\varepsilon$  — вектор случайных возмущений с основными числовыми характеристиками:

- математическое ожидание:

$$E(\varepsilon) = 0, \quad (15.3)$$

- автоковариационная матрица:

$$C_{\varepsilon\varepsilon} = E(\varepsilon\varepsilon^T) = \sigma^2 \cdot I_n = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}, \quad (15.4)$$

где  $I_n$  — единичная матрица размером  $n \times n$ ,

$$\sigma^2 = \text{Var}(\varepsilon_t) = \text{const}, \quad t = 1, \dots, n \quad (15.5)$$

— дисперсия возмущений не зависит от номера наблюдения  $t$  (возмущение *гомоскедастично*); возмущения с различными номерами наблюдений статистически независимы (некоррелированы)

$$\text{Cov}(\varepsilon_t, \varepsilon_s) = 0, \quad \text{при } t \neq s, \quad (15.6)$$

- вектор случайных возмущений имеет нормальное распределение:

$$\varepsilon \sim N(0, C_{\varepsilon\varepsilon}). \quad (15.7)$$

Для проверки предпосылок (15.3)-(15.7) разработано множество тестов, основанных на формальных статистических критериях. Широкое практическое применение получили те из них, которые позволяют выполнить диагностику при малом объеме выборки.

**Тест Голдфельда-Квандта (Goldfeld-Quandt).** Нулевая гипотеза теста Голдфельда-Квандта  $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$  против альтернативной  $H_1$ : не  $H_0$ . *Предпосылки теста:* пропорциональность дисперсии

случайного возмущения величине регрессора  $X_j$ ; случайное возмущение имеет нормальное распределение, автокорреляция возмущений отсутствует. Алгоритм теста:

1. Данные упорядочиваются по возрастанию модуля регрессора  $|X_j|$ , 5 предположительно вызывающего гетероскедастичность.

2. По первым и последним  $m$  выборочным данным оцениваются две вспомогательные регрессии, и вычисляются суммы квадратов остатков:

$$ESS_1 = \sum_{t=1}^m e_{1t}^2, \quad ESS_2 = \sum_{t=1}^m e_{2t}^2,$$

где  $k < m \leq n/2$ ,  $k$  — число параметров модели,  $n$  — объём исходной выборки,  $e_{it}$  — значение остатка  $i$ -ой вспомогательной регрессии в момент  $t$ ,  $i = 1, 2$ . Для разделения групп с малыми и большими дисперсиями возмущений, данные (объёмом  $n - 2m$ ) из центральной части выборки отбрасываются.

3. Вычисляется статистика теста, имеющая  $F$  – распределение:

$$GQ = ESS_2 / ESS_1. \quad (15.8)$$

4. Определяется критическое значение  $F_\alpha(m - k, m - k)$  для заданного уровня значимости  $\alpha$ .

5. Нулевая гипотеза не отклоняется, если справедливо неравенство:  $GQ < F_\alpha$ , в противном случае делается вывод о гетероскедастичности случайных возмущений.

**Тест Дарбина-Уотсона (Durbin, Watson).** Предпосылки теста: случайное возмущение  $\varepsilon$  имеет нормальное распределение, не подвержено гетероскедастичности, регрессоры детерминированы, модель включает свободный член. Статистика теста вычисляется по формуле:

$$DW = \sum_{t=2}^n (e_t - e_{t-1})^2 / \sum_{t=1}^n e_t^2 = \sum_{t=2}^n (e_t - e_{t-1})^2 / ESS \approx 2(1 - \hat{\rho}), \quad (15.9)$$

где  $e_t$  — остатки регрессии,  $\hat{\rho}$  — выборочный коэффициент корреляции между остатками регрессии, разделёнными одним лагом. Коэффициент корреляции принимает значения в диапазоне  $-1 \leq \rho \leq 1$ , поэтому для значений статистики  $DW$  выполняется неравенство  $0 \leq DW \leq 4$ . Для

проверки нулевой гипотезы<sup>2</sup>  $H_0: \rho = 0$ , против альтернативной, например,  $H_1$ : не  $H_0$ , авторы теста оценили верхнюю и нижнюю границы критического значения статистики  $d_L < DW_{kp} < d_u$ , которые делят интервал её возможных значений на пять частей, представленных на рис.3.

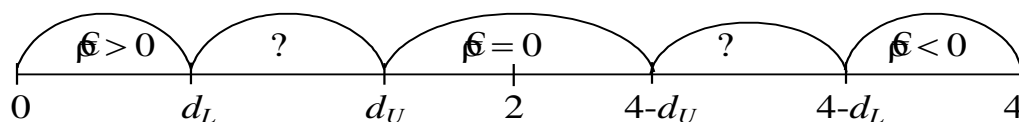


Рис.3. Множество возможных значений статистики  $DW$ .

Алгоритм теста имеет следующую последовательность: 1) оценка параметров модели и вычисление остатков:  $e = Y - \hat{Y}$ ; 2) вычисление статистики  $DW$  (по формуле (15.9)); 3) выбор табличных значений границ критического значения статистики:  $d_U$  и  $d_L$  (по параметрам:  $n$  — число наблюдений,  $K$  — число регрессоров,  $\alpha$  — уровень значимости), 4) определение промежутка, в который попадает вычисленное значение статистики  $DW$ .

Если:

$d_U < DW < 4 - d_U$  — нулевая гипотеза  $H_0: \rho = 0$  не отклоняется;

$0 < DW < d_L$  — принимается альтернативная гипотеза  $H_1: \rho > 0$ ;

$4 - d_L < DW < 4$  — принимается альтернативная гипотеза  $H_1: \rho < 0$

. На участках  $d_L < DW < d_U$  и  $4 - d_U < DW < 4 - d_L$  тест не работает (зона неопределённости).

### Тест *RESET* (*Ramsey's Regression Specification Error Test*)

Для тестирования правильности выбора спецификации в эконометрических пакетах применяется тест *RESET*. Алгоритм теста состоит из следующих шагов:

1. Оценивается спецификация исследуемой модели:

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_j X_{jt} + \dots + \beta_k X_{kt} + \varepsilon_t, t = 1, \dots, n, \quad (15.10)$$

<sup>2</sup>  $\rho$  — теоретический коэффициент корреляции между остатками регрессии, разделёнными одним лагом.

$$\hat{Y}_t = \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_j X_{jt} + \dots + \hat{\beta}_k X_{kt}.$$

2. Оценивается вспомогательная регрессия:

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_j X_{jt} + \dots + \beta_k X_{kt} + b_1 \hat{Y}_t^2 + \dots + b_p \hat{Y}_t^{p+1} + v_t. \quad (15.11)$$

Нулевая гипотеза формулируется в рамках вспомогательной модели (15.11)

$$H_0: b_1 = b_2 = \dots = b_p = 0. \quad (15.12)$$

Нулевую гипотезу (15.12) можно проверить при помощи  $F$ -теста:

$$F = \frac{(ESS_R - ESS_{UR})/p}{ESS_{UR}/(n - k - p)}, \quad (15.13)$$

где  $ESS_R$  — сумма квадратов остатков усечённой (исследуемой) регрессии (*Restricted Regression*),  $ESS_{UR}$  — сумма квадратов остатков неусечённой (вспомогательной) регрессии (*Unrestricted Regression*). Если вычисленное значение статистики (15.13) окажется больше критического значения

$$F > F_\alpha(p, n - k - p),$$

нулевая гипотеза не принимается и спецификация модели (15.10) признаётся неверной.  $F$ -статистики типа (15.13) имеют распределение Фишера только в случае, если случайные возмущения регрессионной модели являются независимыми и нормально распределёнными.

## 16. Точечные и интервальные оценки параметров и эндогенной переменной, их применение в статистическом анализе эконометрических моделей.

Удобной формализацией модели множественной регрессии является матричная форма:

$$Y = X\beta + \varepsilon, \quad (16.1)$$

где  $Y = (Y_1, \dots, Y_t, \dots, Y_n)^T$  — вектор-столбец значений эндогенной переменной,  $X$  — детерминированная  $(n \times k)$ -матрица регрессоров,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_t, \dots, \varepsilon_n)^T$  — вектор-столбец возмущений,  $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$  — вектор-столбец параметров модели,  $k$  — число параметров,  $t$  — номер

наблюдения,  $n$  — объем выборки. Вектор возмущений  $\varepsilon$  является случайным вектором с основными числовыми характеристиками:

$$E\{\varepsilon\} = 0, C_{\varepsilon\varepsilon} = \sigma^2 \cdot I_n, \quad (16.2)$$

где  $C_{\varepsilon\varepsilon}$  — автоковариационная матрица вектора  $\varepsilon$ ,  $\sigma^2$  — дисперсия возмущений,  $I_n$  —  $(n \times n)$ -единичная матрица.

Числовые характеристики вектора значений эндогенной переменной модели (16.1):

$$E\{Y\} = X\beta, C_{YY} = \sigma^2 \cdot I_n, \quad (16.3)$$

определяются с учетом свойств операции математического ожидания и автоковариационной матрицы вектора возмущений (16.2):

$$\begin{aligned} E\{Y\} &= E\{X\beta\} + E\{\varepsilon\} = X\beta, \\ C_{YY} &= Cov(Y, Y) = Cov(X\beta + \varepsilon, X\beta + \varepsilon) = \\ &= Cov(X\beta, X\beta) + 2Cov(X\beta, \varepsilon) + Cov(\varepsilon, \varepsilon) = \\ &= Cov(\varepsilon, \varepsilon) = C_{\varepsilon\varepsilon} = \sigma^2 I_n. \end{aligned}$$

Вектор точечных МНК-оценок параметров модели множественной линейной регрессии (16.1) является случайным вектором, оцениваемым по выборочным данным:

$$\hat{\beta} = (X^T X)^{-1} X^T Y = AY, \quad (16.4)$$

где  $A = (X^T X)^{-1} X^T$  — детерминированная матрица. Основными числовыми характеристиками вектора оценок (16.4) являются:

$$E\{\hat{\beta}\} = \beta, C_{\hat{\beta}\hat{\beta}} = \sigma^2 (X^T X)^{-1}. \quad (16.5)$$

Характеристики (16.5) получены с учетом предпосылок (16.2):

$$\begin{aligned} E\{\hat{\beta}\} &= E\{AY\} = AE\{Y\} = AE\{X\beta + \varepsilon\} = A \cdot X \cdot \beta = I_k \cdot \beta = \beta, \\ Cov\{\hat{\beta}, \hat{\beta}\} &= Cov\{AY, AY\} = ACov\{Y, Y\}A^T = A\sigma^2 I_n A^T = \\ &= \sigma^2 AA^T = \sigma^2 (X^T X)^{-1}. \end{aligned}$$

МНК-оценки параметров (16.4) используются для оценки и прогнозирования вектора значений эндогенной переменной:

$$\hat{Y} = \hat{E}\{Y\} = X\hat{\beta} = XAY = NY, \quad (16.6)$$

с основными числовыми характеристиками:

$$E\{\hat{Y}\} = X\beta, C_{\hat{Y}\hat{Y}} = \sigma^2 \cdot N, \quad (16.7)$$

$$E\{X\hat{\beta}\} = XE\{\hat{\beta}\} = X\beta,$$

$$Cov\{\hat{Y}, \hat{Y}\} = Cov\{NY, NY\} = NCov\{Y, Y\}N^T = N\sigma^2 I_n N^T = \sigma^2 N,$$

где  $N = XA = X(X^T X)^{-1} X^T$  —  $(n \times n)$  - детерминированная идемпотентная матрица. Оценки и прогнозы эндогенных переменных ( $\hat{Y}$ ) не совпадают с их истинными значениями ( $Y$ ) в силу разных причин: ограниченность выборочных данных, ошибки спецификации (пропуск существенных регрессоров, неправильный выбор уравнения регрессии), ошибки измерений.

Обозначим: вектор остатков регрессии (отклонений на интервале оценивания,  $t \leq n$ )

$$e_e = Y_e - \hat{Y}_e = (I_n - N)Y_e = MY_e = M\varepsilon, \quad (16.8)$$

где  $Y_e = (Y_1, Y_2, \dots, Y_n)^T$ ,  $\hat{Y}_e = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)^T$ ,  $M = (I_n - N)$ -идемпотентная матрица,  $MX = 0$ ; вектор ошибок прогнозов (отклонений на интервале прогнозирования, например, для  $t > n$ )

$$e_p = Y_p - \hat{Y}_p, \quad (16.9)$$

где  $Y_p = (Y_{n+1}, Y_{n+2}, \dots, Y_{n+m})^T$ ,  $\hat{Y}_p = (\hat{Y}_{n+1}, \hat{Y}_{n+2}, \dots, \hat{Y}_{n+m})^T$ ,  $m$  — период упреждения. Определим автоковариационную матрицу вектора отклонений  $Y - \hat{Y}$  для любого интервала:

$$Cov(Y - \hat{Y}, Y - \hat{Y}) = Cov(Y, Y) - 2Cov(\hat{Y}, Y) + Cov(\hat{Y}, \hat{Y}). \quad (16.10)$$

На интервале оценивания:

$$Cov(\hat{Y}, Y) = Cov(\hat{Y}_e, Y_e) = Cov(NY_e, Y_e) = NCov(Y_e, Y_e) = NC_{YY} = \sigma^2 N, \quad (16.11)$$

поэтому автоковариационная матрица вектора остатков (16.8), с учетом (16.10), (16.3), (16.7) и (16.11), принимает вид:

$$C_{ee} = \sigma^2 \cdot I_n - 2\sigma^2 \cdot N + \sigma^2 \cdot N = \sigma^2(I_n - N) = \sigma^2 M. \quad (16.12)$$

На интервале прогнозирования:

$$Cov(\hat{Y}, Y) = Cov(\hat{Y}_p, Y_p) = Cov(NY_n, Y_p) = NCov(Y_n, Y_p) = 0,$$

в силу некоррелированности возмущений в различных наблюдениях, поэтому автоковариационная матрица вектора ошибок прогнозов (16.9), с учетом (16.7), (16.10) равна:

$$C_{e_p e_p} = \sigma^2 \cdot I_p + \sigma^2 \cdot N_p = \sigma^2 (I_p + N_p) = \sigma^2 (I_p + X_p (X^T X)^{-1} X_p^T), \quad (16.13)$$

где  $X_p$  — матрица регрессоров на интервале прогнозирования. Несмещенная оценка дисперсии возмущений, вычисляемая через вектор остатков регрессионной модели  $s^2 = \hat{\sigma}^2 = e^T e / (n - k)$ , позволяет оценить автоковариационные матрицы всех случайных векторов эконометрической модели.

### *Интервальная оценка параметров регрессионной модели*

Одной из задач статистического анализа эконометрической модели является сравнение оценок параметров с их теоретическими значениями. Основным инструментом такого сравнения — статистическая *проверка гипотез*. Для проверки нулевой гипотезы

$$H_0: \beta_i = \hat{\beta}_i \quad (16.14)$$

используется статистика, имеющая распределение Стьюдента (с параметром, равным числу степеней свободы):

$$t_{\hat{\beta}_i} = \frac{\beta_i - \hat{\beta}_i}{s_{\hat{\beta}_i}} \sim t(n - k), \quad (16.15)$$

где  $\beta_i$  —  $i$ -тый элемент вектора параметров  $\beta$ . Числитель дроби (16.15) — ошибка оценки, знаменатель —

$$s_{\hat{\beta}_i} = \sqrt{\text{var}\{\beta_i - \hat{\beta}_i\}} = \sqrt{\text{var}\{\hat{\beta}_i\}} \quad (16.16)$$

— стандартная ошибка коэффициента. В формуле (16.16) через  $\text{var}\{\hat{\beta}_i\}$  обозначена оценка дисперсии элемента  $\hat{\beta}_i$ , которая является  $i$ -м элементом главной диагонали матрицы  $\hat{C}_{\hat{\beta}\hat{\beta}}$ . Гипотеза (16.14) не отклоняется, если

$$|t_{\hat{\beta}_i}| = |(\beta_i - \hat{\beta}_i) / s_{\hat{\beta}_i}| < t_\alpha, \quad (16.17)$$

где  $\alpha$  — уровень значимости,  $t_\alpha$  — квантиль уровня значимости  $\alpha$  (аргумент функции распределения Стьюдента для уровня значимости  $\alpha$ ). Неравенство (16.17) можно записать в виде двойного:

$$-t_\alpha < t_{\hat{\beta}_i} < t_\alpha, \quad -t_\alpha < (\beta_i - \hat{\beta}_i)/s_{\hat{\beta}_i} < t_\alpha,$$

или относительно неизвестного параметра:  $\hat{\beta}_i - t_\alpha \cdot s_{\hat{\beta}_i} < \beta_i < \hat{\beta}_i + t_\alpha \cdot s_{\hat{\beta}_i}$ .

Таким образом, границы доверительного интервала, который с доверительной вероятностью  $1 - \alpha$  накрывает истинное значение параметра  $\beta_i$ , равны:

$$\beta^- = \hat{\beta}_i - t_{kp} \cdot s_{\hat{\beta}_i}, \quad \beta^+ = \hat{\beta}_i + t_{kp} \cdot s_{\hat{\beta}_i}. \quad (16.18)$$

Частным случаем нулевой гипотезы (16.14) является

$$H_0: \beta_i = 0, \quad (16.19)$$

которая позволяет проверить статистическую значимость оценки параметра (значимость влияния  $i$ -го регрессора на эндогенную переменную). Статистика (16.15), с учетом (16.19), принимает вид:

$$|t_{\hat{\beta}_i}| = |\hat{\beta}_i/s_{\hat{\beta}_i}|.$$

Если неравенство

$$|t_{\hat{\beta}_i}| = |\hat{\beta}_i/s_{\hat{\beta}_i}| \leq t_\alpha, \quad (16.20)$$

выполняется, то влияние регрессора  $X_i$  на эндогенную переменную признается статистически *незначимым* при заданном уровне  $\alpha$ , если не выполняется, то гипотеза  $H_0$  не принимается, и регрессор  $X_i$  признается значимым. Проверка статистической значимости регрессоров используется при построении спецификации регрессионной модели.

## **17. Динамические регрессионные модели: модели с распределенными лагами, авторегрессионные модели (спецификация, интерпретация параметров, методы оценки параметров).**

При моделировании экономических процессов часто возникают ситуации, когда влияние одной экономической переменной на другую распределено во времени, т.е. эффект воздействия проявляется не сразу, а с



некоторой задержкой. Например, инвестиции только с определенной задержкой во времени переходят в приращение основного капитала. Для отражения фактора времени в спецификацию модели вводятся *лаговые* (запаздывающие) *переменные*, датированные предыдущими моментами времени и находящиеся в уравнении с текущими переменными. Модели с лаговыми переменными имеют два важных частных случая: *модели с распределенными лагами* — модели, включающие в качестве лаговых объясняющих переменных *независимые переменные*

$$Y_t = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + \varepsilon_t, \quad (17.1)$$

$DL(k)$  — краткое обозначение спецификации (*distributed lags*),  $k$  — максимальная величина лага; *авторегрессионные модели* — модели, включающие в качестве лаговых объясняющих переменных *зависимые и независимые переменные*

$$Y_t = \alpha + \beta X_{t-1} + \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + \dots + \lambda_k Y_{t-k} + \varepsilon_t, \quad (17.2)$$

$ADL(k_Y, k_X)$  — краткое обозначение спецификации (*autoregressive distributed lags*),  $k_Y$  — максимальная величина лага эндогенной переменной,  $k_X$  — максимальная величина лага экзогенной переменной.

#### *Характеристики лаговой структуры моделей с распределенными лагами*

В моделях с распределенными лагами (17.1) параметр  $\beta_0$  называется *краткосрочным мультипликатором*. Он характеризует среднее абсолютное изменение  $Y_t$  при изменении  $X_t$  на единицу своего измерения (без учёта лаговых значений):

$$\frac{\partial Y_t}{\partial X_t} = \beta_0 \approx \frac{\Delta Y}{\Delta X}, \quad \Delta Y \approx \beta_0 \cdot \Delta X, \quad \text{при } \Delta X = 1, \Delta Y \approx \beta_0.$$

#### *Долгосрочный мультипликатор*

$$\beta = \sum_{i=0}^k \beta_i \quad (17.3)$$

— характеризует среднее абсолютное изменение  $Y_t$  при изменении  $X$  на единицу своего измерения в каждом из рассматриваемых временных периодов (измеритель общего влияния  $X$  на  $Y_t$ ). Для характеристики вклада отдельного лага используются *относительные параметры*:

$$b_i = \beta_i / \sum_{i=0}^k \beta_i, \quad \sum_{i=0}^k b_i = 1, \quad (17.4)$$

по которым вычисляется *средний лаг*:

$$\bar{l} = \sum_{i=1}^k i \cdot b_i \quad (17.5)$$

— величина, показывающая на сколько периодов (в среднем) запаздывает влияние  $X$  на  $Y$  (скорость реакции  $Y$  на изменение  $X$ ).

#### *Методы оценки параметров моделей с распределенными лагами*

Для оценки параметров модели (17.1) с *конечным числом лагов* применяется *метод замены переменных*:

$$X_{0t}^* = X_t, \quad X_{1t}^* = X_{t-1}, \dots, \quad X_{kt}^* = X_{t-k}, \quad t = k+1, \dots, n. \quad (17.6)$$

С учетом обозначений (17.6), спецификация (17.1) представляет собой модель множественной линейной регрессии:

$$Y_t = \alpha + \beta_0 \cdot X_{0t}^* + \beta_1 \cdot X_{1t}^* + \dots + \beta_k \cdot X_{kt}^* + \varepsilon_t, \quad t = k+1, \dots, n,$$

параметры которой оцениваются при помощи обычного МНК.

Для оценки моделей с *бесконечным числом лагов*

$$\begin{aligned} Y_t &= \alpha + \beta_0 \cdot X_t + \beta_1 \cdot X_{t-1} + \beta_2 \cdot X_{t-2} + \dots + \varepsilon_t = \\ &= \alpha + \sum_{k=0}^{\infty} \beta_k \cdot X_{t-k} + \varepsilon_t, \end{aligned} \quad (17.7)$$

на параметры накладываются ограничения. В методе *геометрической прогрессии* предполагается, что параметры  $\beta_k$ , при лаговых значениях регрессоров, убывают в геометрической прогрессии:

$$\beta_k = \beta_0 \lambda^k, \quad k = 0, 1, 2, \dots, \quad 0 < \lambda < 1. \quad (17.8)$$

Параметр  $\lambda$  характеризует скорость убывания параметров с увеличением лага. Действительно, чем дальше по времени удалены значения регрессоров, тем меньше их влияние на значение эндогенной переменной в текущий момент времени  $t$ . Спецификация модели (17.7), с учетом зависимости (17.8), принимает вид:

$$Y_t = \alpha + \beta_0 \cdot X_t + \beta_0 \cdot \lambda \cdot X_{t-1} + \beta_0 \cdot \lambda^2 \cdot X_{t-2} + \dots + \varepsilon_t. \quad (17.9)$$

Модель с бесконечным числом членов можно свести к модели с конечным числом членов при помощи *преобразования Койка* (Koyck).

Запишем спецификацию (17.9) для момента  $t - 1$  и умножим на  $\lambda$  левую и правую часть:

$$\lambda Y_{t-1} = \lambda \alpha + \beta_0 \lambda \cdot X_{t-1} + \beta_0 \cdot \lambda^2 \cdot X_{t-2} + \beta_0 \cdot \lambda^3 \cdot X_{t-3} + \dots + \lambda \varepsilon_{t-1}. \quad (17.10)$$

Вычтем (17.10) из уравнения (17.9)

$$Y_t - \lambda Y_{t-1} = (1 - \lambda) \alpha + \beta_0 X_t + (\varepsilon_t - \lambda \varepsilon_{t-1}), \text{ или,} \\ Y_t = (1 - \lambda) \alpha + \beta_0 X_t + \lambda Y_{t-1} + v_t = \alpha^* + \beta_0 X_t + \lambda Y_{t-1} + v_t, \quad (17.11)$$

где  $v_t = \varepsilon_t - \lambda \varepsilon_{t-1}$  — скользящая средняя между  $\varepsilon_t$ ,  $\varepsilon_{t-1}$ . Таким образом, при помощи преобразования Койка модель с бесконечным числом лагов (с убывающими по степенному закону параметрами) сводится к авторегрессионной модели типа (17.2), для которой требуется оценить только три параметра  $\alpha, \beta_0, \lambda$ . Такой подход позволяет устранить мультиколлинеарность, часто возникающую при оценке параметров с лаговыми переменными. Однако использование преобразования Койка приводит к следующим проблемам: 1) среди регрессоров появляется лаговая переменная  $Y_{t-1}$ , которая коррелирует со случайным возмущением  $v_t$ , что приводит к проблеме эндогенности; 2) для случайных возмущений исходной модели справедлива предпосылка о некоррелированности, а для случайного возмущения преобразованной модели имеет место автокорреляция. Спецификация авторегрессионной модели (17.11) нуждается в применении соответствующих способов корректировки, в противном случае МНК-оценки параметров будут смещёнными и несостоятельными.

*Распределённые лаги Алмон (Almon)* обладают большей гибкостью по сравнению с геометрически распределёнными лагами. В основе полиномиально-распределённых лагов Алмон лежит предположение о том, что если эндогенная переменная зависит от текущих и лаговых значений регрессора

$$Y_t = \alpha + \beta_0 \cdot X_t + \beta_1 \cdot X_{t-1} + \dots + \beta_i \cdot X_{t-1} + \dots + \beta_k \cdot X_{t-k} + \varepsilon_t, \quad (17.12)$$

то веса в этой зависимости подчиняются полиномиальному распределению

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \dots + a_m i^m, \quad (17.13)$$

где  $m$  — степень полинома,  $k$  — максимальная величина лага. Вспомогательная регрессионная модель в методе Алмон получается подстановкой (17.13) в (17.12), например, для  $m = 2$ :

$$Y_t = \alpha + \sum_{i=0}^k (a_0 + a_1 i + a_2 i^2) X_{t-i} + \varepsilon_t = \alpha + a_0 \sum_{i=0}^k X_{t-i} + a_1 \sum_{i=0}^k i X_{t-i} + a_2 \sum_{i=0}^k i^2 X_{t-i} + \varepsilon_t.$$

Оцениваются параметры вспомогательной регрессии:

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + \varepsilon_t, \quad (17.14)$$

где

$$Z_{0t} = \sum_{i=0}^k X_{t-i}, \quad Z_{1t} = \sum_{i=0}^k i X_{t-i}, \quad Z_{2t} = \sum_{i=0}^k i^2 X_{t-i}$$

— вспомогательные регрессоры, количество вспомогательных регрессоров —  $m + 1$ , число слагаемых при их формировании —  $k + 1$ . Матричная форма спецификации модели (17.14) записывается в виде

$$Y = ZA + \varepsilon, \quad (17.15)$$

где  $A = (\alpha, a_0, a_1, \dots, a_m)^T$  — вектор параметров вспомогательной модели,  $Z$  —  $((n - k) \times (m + 2))$ -матрица регрессоров, включающая единичный вектор столбец (при наличии свободного члена в спецификации) и значения вспомогательных регрессоров в наблюдениях  $t = k + 1, \dots, n$ , где  $n$  — объём выборки. Оценки параметров модели (17.15) получают обычным методом наименьших квадратов (МНК),

$$\hat{A} = (Z^T Z)^{-1} Z^T Y,$$

который, как известно, обеспечивает им несмещенность и эффективность при выполнении предпосылок Гаусса-Маркова относительно вектора случайных возмущений, при условии, что  $\text{rank}(Z) = m + 2$ .

Если представить зависимость (17.13) между оценками параметров исходной модели (17.12) и вспомогательной модели (17.14) в матричной форме

$$\beta = HA,$$

где  $\beta = (\alpha, \beta_0, \beta_1, \dots, \beta_i, \dots, \beta_k)^T$  — вектор параметров исходной модели,  $H$  — детерминированная матрица преобразований, то можно оценить автоковариационную матрицу вектора оценок параметров исходной модели с распределенными лагами

$$C_{\hat{\beta}\hat{\beta}} = Cov(H\hat{A}, H\hat{A}) = HC_{\hat{A}\hat{A}}H^T,$$

где  $C_{\hat{A}\hat{A}} = \sigma^2(Z^T Z)^{-1}$  — автоковариационная матрица оценок параметров вспомогательной регрессии.

### 18. Системы одновременных уравнений: структурная и приведенная формы модели, проблема оценки структурных параметров, проблема идентификации, методы оценки структурных параметров.

Для описания сложных экономических объектов в эконометрике используются системы уравнений: *системы независимых уравнений* — переменные в отдельных уравнениях независимы, *системы внешне не связанных уравнений* — взаимосвязь уравнений объясняется только корреляцией их случайных возмущений (*Seemingly Unrelated Regression, SUR*), *системы одновременных уравнений* — набор взаимосвязанных уравнений, в которых одни и те же переменные в одних уравнениях являются эндогенными, а в других — регрессорами (*COU, Simultaneous equations*).

*Формы спецификации COU.* Спецификация моделей COU может быть представлена в структурной и приведенной формах:

1) *структурная* форма (результат формализации экономических закономерностей):

$$A \cdot Y_t + B \cdot X_t = V_t, \quad (18.1)$$

где

$$A_{m,m} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix}, \quad B_{m,k} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{pmatrix}$$

— матрицы структурных параметров,  $Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{mt})^T$  — вектор столбец значений эндогенных переменных,  $X_t = (X_{1t}, X_{2t}, \dots, X_{kt})^T$  — вектор столбец значений predetermined переменных, которые могут включать как экзогенные переменные (внешние по отношению к системе), так и лаговые значения эндогенных переменных,  $V_t = (v_{1t}, v_{2t}, \dots, v_{mt})^T$  — вектор столбец случайных возмущений,  $t$  — номер наблюдения,

2) *приведенная форма* (результат представления вектора эндогенных переменных в явном виде через вектор predetermined переменных):

$$Y_t = -A^{-1}BX_t + A^{-1}V_t = MX_t + U_t, \quad (18.2)$$

где

$$M = -A^{-1}B - \quad (18.3)$$

матрица коэффициентов приведенной формы,

$$U_t = A^{-1}V_t - \quad (18.4)$$

вектор случайных возмущений приведенной формы.

Рассмотрим систему, включающую два уравнения:

$$\begin{cases} Y_{1t} = a_{12}Y_{2t} + b_{11}X_{1t} + v_{1t} \\ Y_{2t} = a_{21}Y_{1t} + b_{22}X_{2t} + v_{2t} \end{cases} \quad (18.5)$$

с вектором эндогенных переменных  $Y_t = (Y_{1t}, Y_{2t})^T$  и вектором predetermined переменных  $X_t = (X_{1t}, X_{2t})^T$ , где  $Y_1$  — годовое потребление продукта на душу населения (в кг.);  $Y_2$  — оптовая цена за кг. (в долл.);  $X_1$  — доход на душу населения (в долл.);  $X_2$  — расходы по обработке продукта (в % к цене).

Система (18.5) получена в результате математической формализации экономических закономерностей, характеризующих моделируемый объект, и представляет собой *структурную форму спецификации*. Ее можно записать в матричном виде (18.1), предварительно перенеся все члены, кроме случайных возмущений, в левую часть:

$$\begin{cases} Y_{1t} - a_{12}Y_{2t} - b_{11}X_{1t} = v_{1t} \\ Y_{2t} - a_{21}Y_{1t} - b_{22}X_{2t} = v_{2t} \end{cases}$$

Сформировав вектор эндогенных переменных модели, включающий два элемента ( $m = 2$ ),  $Y_t = (Y_{1t}, Y_{2t})^T$ , и вектор<sup>3</sup> predetermined переменных ( $k = 2$ ),  $X_t = (X_{1t}, X_{2t})^T$  заполним матрицы структурных параметров ( $A$  — при векторе  $Y$ ,  $B$  — при векторе  $X$ )

$$\begin{pmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} + \begin{pmatrix} -b_{11} & 0 \\ 0 & -b_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}. \quad (18.6)$$

Матрица  $A(m \times m)$  — квадратная (число строк равно числу уравнений системы, число столбцов равно числу эндогенных переменных<sup>4</sup>). Матрица  $B(m \times k)$  — прямоугольная (число строк равно числу уравнений, число столбцов равно размерности вектора predetermined переменных). Вектор эндогенных переменных системы (18.5) не выражен в явном виде через вектор predetermined переменных, поэтому для решения задачи оценки или прогнозирования необходимо составить *приведенную* форму спецификации (18.2):

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \quad (18.7)$$

Элементы матрицы  $M$  приведенных параметров связаны с элементами матриц  $A$  и  $B$  структурных параметров соотношением (18.3).

*Проблема идентифицируемости структурных параметров.* Структурная форма спецификации отражает все взаимосвязи переменных в системе, однако оценка ее параметров приводит к проблеме эндогенности: переменные, являющиеся в одних уравнениях эндогенными, а в других — регрессорами, коррелируют с возмущениями этих уравнений, что приводит к смещенности и несостоятельности оценок параметров. В приведенной форме эта проблема отсутствует, так как по своей структуре это система независимых уравнений, и поэтому для оценки приведенных параметров применим МНК. Возникает вопрос о возможности определения структурных параметров через МНК-оценки приведенных

<sup>3</sup> Вектор predetermined переменных включает дополнительный элемент — единицу, если хотя бы в одном из уравнений присутствует свободный член.

<sup>4</sup> В правильно составленной спецификации число уравнений равно числу эндогенных переменных.

параметров. Если такая возможность есть, уравнение называется *идентифицируемым*, если нет — *неидентифицируемым*.

Проверка идентифицируемости отдельных уравнений системы выполняется при помощи *порядкового* и *рангового* условий. Порядковое условие идентифицируемости (необходимое условие):

$$k - p \geq q - 1, \quad (18.8)$$

где  $k$  — число predetermined переменных в системе,  $p$  — число predetermined переменных в уравнении,  $m$  — число эндогенных переменных в системе,  $q$  — число эндогенных переменных в уравнении. Если условие (18.8) выполняется со знаком строгого неравенства, уравнение называется *сверхидентифицируемым*, если со знаком равенства — *точно идентифицируемым*. Если неравенство не выполняется, то уравнение *неидентифицируемо*, и нет возможности определения структурных параметров по приведенным.

*Методы оценки параметров СОУ.* Для оценки параметров СОУ разработаны специальные эконометрические методы: косвенный метод наименьших квадратов (КМНК, *indirect least squares method, ILS*), двухшаговый метод наименьших квадратов (ДМНК, *two stage least squares method, 2SLS*), которые позволяют решить проблему эндогенности регрессоров, возникающую при оценке структурных параметров системы. Повысить точность ДМНК-оценок структурных параметров можно в рамках трёхшагового метода наименьших квадратов (ТМНК, *three stage least squares method, 3SLS*), учитывающего корреляцию возмущений отдельных уравнений системы.

*Косвенный метод наименьших квадратов (КМНК).* КМНК применяется для *точно идентифицируемых уравнений*, и состоит из следующих шагов:

- 1) по структурной форме модели строится приведенная форма;
- 2) определяются МНК-оценки параметров приведенной формы;



3) по МНК-оценкам параметров приведенной формы вычисляются оценки параметров структурной формы, с учетом их взаимосвязи (18.3).

Запишем равенство (18.3) через расширенную матрицу  $\bar{A} = (A|B)$  структурной формы

$$\bar{A} \cdot \begin{pmatrix} M \\ I \end{pmatrix} = 0, \quad (18.13)$$

где  $I$  — единичная матрица  $k \times k$ ,  $k$  — число столбцов матрицы  $M$ . Из решения системы (18.13) и находятся КМНК-оценки структурных параметров. Если значения элементов матрицы  $M$  приведенной формы неизвестны, то в системе (18.13) используются их МНК-оценки. МНК к оценке структурных параметров применяется косвенно, отсюда и название метода.

*Двухшаговый метод наименьших квадратов (ДМНК).* Двухшаговый МНК обладает большей универсальностью по сравнению с КМНК и применяется как для точно идентифицируемых уравнений (в этом случае оценки ДМНК и КМНК совпадают), так и для сверхидентифицируемых уравнений. По существу, ДМНК является методом инструментальных переменных (МИП), предназначенным для решения проблемы эндогенности регрессоров: регрессоры, коррелирующие с возмущениями в регрессионном уравнении, заменяются *инструментами* — переменными, которые сильно коррелируют с замещаемыми регрессорами и не коррелируют с возмущениями данного уравнения. В ДМНК, в качестве инструментов, используют оценки эндогенных регрессоров, полученные в рамках оценивания модели *приведенной формы*. Поэтому процедура оценки параметров СОУ, в ДМНК, разбивается на два шага: *на первом шаге* определяются *эндогенные регрессоры* оцениваемого уравнения СОУ (регрессоры данного уравнения, которые являются элементами вектора эндогенных переменных системы). Выполняется оценка этих переменных по всем предопределенным переменным системы; *на втором шаге*

оценивается структурное уравнение СОУ, в котором «проблемные» регрессоры заменяются их оценками, полученными на первом шаге.

## 19. Антагонистические игры, их применение .

Теория игр является одной из наук, предоставляющей возможность математического описания постановок различных задач по принятию решений и математическое обоснование подходов к их анализу. Теория антагонистических игр представляет собой теоретические основы математических моделей принятия оптимальных решений в конфликтных рыночных отношениях, носящих характер конкурентной борьбы.

Использование теории игр помогает лицу, принимающему решение, произвести критический анализ ситуации и в результате более обосновано и последовательно проводить определенную политику или стратегию поведения при решении сложных, комплексных проблем.

Пусть  $S_A^C = \{A_1, A_2, \dots, A_m\}$  и  $S_B^C = \{B_1, B_2, \dots, B_n\}$  – множества чистых стратегий соответственно игроков  $A$  и  $B$  в парной антагонистической игре (с нулевой суммой выигрышей). Эти стратегии называются чистыми, в отличие от смешанных стратегий, поскольку каждый из игроков выбирает одну из них определенным, а не случайным образом. При  $m = 1$  проблема выбора стратегии игроком  $A$  отсутствует. При  $n = 1$  проблема выбора стратегии игроком  $A$  тривиальна. Поэтому предполагают  $m \geq 2$  и  $n \geq 2$ .

Из выигрышей  $a_{ij} = F_A(i, j) = F_A(A_i, B_j)$  игрока  $A$  в игровых ситуациях  $(A_i, B_j)$ ,  $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ , представляющих собой числовые значения его выигрыш-функции  $F_A$ , определенной на декартовом произведении  $S_A^C \times S_B^C$ , формируют матрицу выигрышей игрока  $A$  (платежную матрицу или матрицу игры)

$$A = \begin{array}{c|ccccc} & B_j & B_1 & B_2 & \dots & B_n \\ \hline A_i & & & & & \\ \hline A_1 & a_{11} & a_{12} & \dots & a_{1n} \\ A_2 & a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ A_m & a_{m1} & a_{m2} & \dots & a_{mn} \end{array} \quad (19.1)$$

Если игрок  $B$  – противник игрока  $A$  и  $b_{ji} = F_B(j, i) = F_B(B_j, A_i)$ ,  $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ , – выигрыши игрока  $B$ , то матрица выигрышей игрока  $B$  имеет вид

$$B = \begin{array}{c|ccccc} & A_i & A_1 & A_2 & \dots & A_m \\ \hline B_j & & & & & \\ \hline B_1 & b_{11} & b_{12} & \dots & b_{1m} \\ B_2 & b_{21} & b_{22} & \dots & b_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ B_n & b_{n1} & b_{n2} & \dots & b_{nm} \end{array} \quad (19.2)$$

Так как игра – антагонистическая, то выигрыш-функции игроков  $A$  и  $B$  и их матрицы выигрышей связаны между собой следующим образом:

$$b_{ji} = F_B(B_j, A_i) = -F_A(A_i, B_j) = -a_{ij}, \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n,$$

т.е.  $B = A^T$ , где  $T$  – значок транспонирования матрицы.

*Показателем эффективности стратегии  $A_i$*  назовем минимальный выигрыш при этой стратегии (т.е. минимальный элемент  $i$ -й строки матрицы  $A$ ):

$$\alpha_i = \min_{1 \leq j \leq n} a_{ij}, \quad i = 1, 2, \dots, m.$$

*Максимумом или нижней ценой игры в чистых стратегиях* называется наибольший из показателей эффективности стратегий  $A_i$ ,  $i = 1, 2, \dots, m$ ,

$$\alpha = \max_{1 \leq i \leq m} \alpha_i = \max_{1 \leq i \leq m} \min_{1 \leq j \leq n} a_{ij}. \quad (19.3)$$

Стратегия  $A_k$ , показатель эффективности которой совпадает с максимумом:  $\alpha_k = \alpha$ , называется *максиминной стратегией* игрока  $A$ . Множество всех (чистых) максиминных стратегий игрока  $A$  обозначим через  $(S_A^C)^{\max\min}$ . Принцип выбора игроком  $A$  максиминной стратегии в качестве эффективной называется *максиминным принципом*. Если игрок  $A$  придерживается максиминного принципа выбора стратегий, то ему при любой игре противника  $B$  гарантирован выигрыш в чистых стратегиях, не меньший максимина  $\alpha$ .

*Показателем неэффективности стратегии  $B_j$*  назовем максимальный проигрыш игрока  $B$  при этой стратегии (т.е. максимальный элемент  $j$ -го столбца матрицы  $A$ ):

$$\beta_j = \max_{1 \leq i \leq m} a_{ij}, \quad j = 1, 2, \dots, n.$$

*Минимаксом* или *верхней ценой игры в чистых стратегиях* называется наименьший из показателей неэффективности стратегий  $B_j$ ,  $j = 1, 2, \dots, n$ :

$$\beta = \min_{1 \leq j \leq n} \beta_j = \min_{1 \leq j \leq n} \max_{1 \leq i \leq m} a_{ij}. \quad (19.4)$$

Стратегия  $B_l$ , показатель неэффективности которой совпадает с минимаксом:  $\beta_l = \beta$ , называется *минимаксной стратегией* игрока  $B$ . Множество всех (чистых) минимаксных стратегий игрока  $B$  обозначим через  $(S_B^C)^{\min\max}$ . Принцип выбора игроком  $B$  минимаксной стратегии в качестве эффективной называется *минимаксным принципом*. Если игрок  $B$  придерживается минимаксного принципа выбора стратегий, то он при любой игре противника  $A$  не может проиграть больше минимакса  $\beta$ .

Для нахождения нижней и верхней цен игры в чистых стратегиях удобно дополнить матрицу игры  $A$  столбцом показателей эффективности  $\alpha_i$  стратегий  $A_i$ ,  $i = 1, 2, \dots, m$ , и строкой показателей неэффективности  $\beta_j$  стратегий  $B_j$ ,  $j = 1, 2, \dots, n$ . В результате получим матрицу

$B_j$	$B_1$	$B_2$	...	$B_n$	$\alpha_i$
$A_i$					
$A_1$	$a_{11}$	$a_{12}$	...	$a_{1n}$	$\alpha_1$
$A_2$	$a_{21}$	$a_{22}$	...	$a_{2n}$	$\alpha_2$
...	...	...	...	...	...
$A_m$	$a_{m1}$	$a_{m2}$	...	$a_{mn}$	$\alpha_m$
$\beta_j$	$\beta_1$	$\beta_2$	...	$\beta_n$	$\beta$

Для элементов матрицы  $A$  имеют место неравенства  $\alpha_i \leq a_{ij} \leq \beta_j$ ,  $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ , и, следовательно, нижняя цена игры в чистых стратегиях не больше ее верхней цены:  $\alpha \leq \beta$ .

Ситуация  $(A_k, B_l)$ , сложившаяся в результате выбора игроками  $A$  и  $B$  соответственно стратегий  $A_k$  и  $B_l$ ,  $k \in \{1, 2, \dots, m\}$ ,  $l \in \{1, 2, \dots, n\}$  называется *удовлетворительной (приемлемой, допустимой) для игрока  $A$* , если  $a_{il} \leq a_{kl}$ ,  $i = 1, 2, \dots, m$ , и *удовлетворительной для игрока  $B$* , если  $a_{kl} \leq a_{kj}$ ,  $j = 1, 2, \dots, n$ .

Ситуация  $(A_k, B_l)$  называется *равновесной (ситуацией равновесия, устойчивой)*, или *седловой точкой выигрыш-функции игрока  $A$* , если она удовлетворительна для каждого из игроков  $A$  и  $B$ , т.е.  $a_{il} \leq a_{kl} \leq a_{kj}$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, n$ , или эквивалентным образом  $\alpha_k = a_{kl} = \beta_l$ .

Выигрыш  $a_{kl}$ , соответствующий ситуации равновесия  $(A_k, B_l)$ , т.е. значение выигрыш-функции игрока  $A$  на аргументе  $(A_k, B_l)$ , называется *седловой точкой матрицы игры*. Таким образом, элемент  $a_{kl}$ , являющийся седловой точкой матрицы игры, является минимальным в  $i$ -й строке и максимальным в  $l$ -м столбце. Игра, матрица которой содержит хотя-бы один такой элемент, называется *игрой с седловой точкой*.

Стратегии  $A_k$  и  $B_l$  соответственно игроков  $A$  и  $B$ , создающие равновесную ситуацию  $(A_k, B_l)$ , или, другими словами, соответствующие

седловой точке  $a_{kl}$ , называются *оптимальными*. Обозначим через  $(S_A^c)^o$  и  $(S_B^c)^o$  – множества чистых оптимальных стратегий соответственно игроков  $A$  и  $B$ .

Если нижняя цена игры  $\alpha$  равна верхней цене игры  $\beta$ , то их общее значение  $\gamma = \alpha = \beta$  называется *ценой игры в чистых стратегиях*. Совокупность  $\{(S_A^c)^o, (S_B^c)^o, \gamma\}$  множеств  $(S_A^c)^o$  и  $(S_B^c)^o$  чистых оптимальных стратегий соответственно игроков  $A$  и  $B$  и цены игры в чистых стратегиях  $\gamma$  называется *полным (или общим) решением игры в чистых стратегиях*, а совокупность  $\{A_k, B_l, \gamma\}$  какой-либо пары чистых оптимальных стратегий  $A_k$  и  $B_l$  и цены игры в чистых стратегиях  $\gamma$  называется *частным решением игры в чистых стратегиях*.

## **20. Игры с природой. Матрица рисков. Принятие решений в условиях риска и неопределенности.**

В экономической практике во многих задачах принятия решений существенным элементом является неопределённость, заключающаяся в недостаточной информированности лица, принимающего решение, об объективных условиях, в которых будет приниматься решение. Неопределённость такого рода может порождаться различными причинами: меняющимся объёмом перевозок, рыночной конъюнктурой, политикой правительства, надёжностью партнёра, курсом валюты, уровнем инфляции, налоговой политикой и др.

Многие задачи указанного типа обладают тем свойством, что в них проявляется взаимодействие стороны, принимающей решение, с окружающей средой, являющейся объективной действительностью и называемой *природой*. Математическая модель подобных ситуаций называется *игрой с природой*. Заинтересованная сторона и природа в такой игре называются *игроками*. Сторона, принимающая решение, для достижения своей цели имеет возможность действовать различными

допустимыми способами, результативность которых зависит от *состояний природы*.

Таким образом, в игре с природой осознанно действует только один из двух игроков - лицо, принимающее решение; обозначим его через  $A$ . Природа, обозначим ее через  $\Pi$ , являясь вторым участником игры, не является ни противником, ни союзником игрока  $A$ , ибо она не действует осознанно злонамеренно против игрока  $A$  или за одно с ним, а принимает неопределенным образом то или иное свое состояние, не преследуя конкретной цели и абсолютно безразлично к результату игры. Игрок  $A$  на состояния природы не может оказывать никакого влияния

Одним из важных предположений в теории игр с природой является предположение о том, что в любой момент времени природа  $\Pi$  находится только в одном (но не известно в каком) из  $n$  своих состояний,  $\Pi_1, \Pi_2, \dots, \Pi_n$  т.е. состояния природы альтернативны, и каждое из них должно быть описано качественно (содержательно). Совокупность  $S_\Pi = \{\Pi_1, \Pi_2, \dots, \Pi_n\}$  состояний природы  $\Pi$  формируется либо на основе имеющегося опыта анализа состояний природы, либо в результате предположений и интуиции экспертов.

Если известны вероятности  $q_1, q_2, \dots, q_n$  соответственно состояний  $\Pi_1, \Pi_2, \dots, \Pi_n$  природы  $\Pi$  или принята какая-либо гипотеза о распределении этих вероятностей (например, гипотеза об их относительных величинах), то в этих случаях говорят о *принятии решения в условиях риска*.

Так как в каждый момент времени природа  $\Pi$  может случайным образом находиться только в одном из своих состояний, то события, состоящие в том, что природа  $\Pi$  находится в состоянии  $\Pi_j, j = 1, 2, \dots, n$ , случайны, несовместны и образуют полную группу. Поэтому, как известно из теории вероятностей, сумма вероятностей этих событий равна единице.

Состояние природы, вероятность которого равна нулю, не играет существенной роли в анализе ситуации и потому его можно исключить из рассмотрения. Поэтому вероятности  $q_1, q_2, \dots, q_n$  соответственно состояний  $\Pi_1, \Pi_2, \dots, \Pi_n$  удовлетворяют условиям

$$q_j > 0, \quad j = 1, 2, \dots, n, \quad \sum_{j=1}^n q_j = 1. \quad (20.1)$$

В случае же, когда вероятности состояний природы неизвестны и нет никакой возможности получить о них какую-либо статистическую информацию, то говорят о «*принятии решения в условиях (полной) неопределенности*».

Количественная характеристика результатов выбора игроком  $A$  своих чистых стратегий при различных состояниях природы задается его выигрыш-функцией в чистых стратегиях  $F : (S^C \times S_\Pi) \rightarrow R$ , определённой на декартовом произведении  $S^C \times S_\Pi$  и ставящей в соответствие каждой ситуации  $(A_i, \Pi_j) \in (S^C \times S_\Pi)$  некоторое действительное число  $a_{ij} = F(i, j) == F(A_i, \Pi_j)$ , называемое выигрышем игрока  $A$  в ситуации  $A_i, \Pi_j$ . Поскольку выигрыши  $a_{ij}$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, n$  снабжены двойной индексацией, то их массив удобно представлять в виде матрицы:

$$A = \begin{array}{c|ccccc} & \Pi_j & \Pi_1 & \Pi_2 & \dots & \Pi_n \\ \hline A_i & & & & & \\ \hline A_1 & a_{11} & a_{12} & \dots & a_{1n} \\ \hline A_2 & a_{21} & a_{22} & \dots & a_{2n} \\ \hline \dots & \dots & \dots & \dots & \dots \\ \hline A_m & a_{m1} & a_{m2} & \dots & a_{mn} \end{array} \quad (2.2)$$

Матрица  $A$  называется матрицей выигрышей игрока  $A$ , или *платежной матрицей*, или *матрицей игры*.

При выборе возможной стратегии в игре с природой игрок  $A$  должен исходить из матрицы выигрышей  $A$  (20.2). Но она не всегда адекватно отражает имеющуюся ситуацию. На выбор стратегии влияют не только



выигрыши, составляющие матрицу  $A$ , но и показатели благоприятности состояний природы для увеличения выигрыша и индикаторы «удачности» или «неудачности» выбора данной стратегии при данном состоянии природы.

Показателем благоприятности состояния  $\Pi_j$  для увеличения выигрыша в чистых стратегиях называется наибольший выигрыш при этом состоянии, т. е. наибольший элемент в  $j$ -м столбце платёжной матрицы:

$$\beta_j = \beta_j(S^C) = \max_{1 \leq i \leq m} a_{ij}, \quad j = 1, 2, \dots, n. \quad (20.3)$$

Для описания степени удачности выбора игроком  $A$  стратегии  $A_i$  при состоянии природы  $\Pi_j$  вводят понятие «риска» (игрового риска или сожаления).

Риском  $r_{ij}$  игрока  $A$  при выборе им стратегии  $A_i$  и при условии, что природа находится в состоянии  $\Pi_j$ , называется разность между показателем благоприятности  $\beta_j$  и выигрышем  $a_{ij}$ , т. е. разность между выигрышем, который игрок  $A$  получил бы, если бы знал заранее, что природа примет состояние  $\Pi_j$ , и выигрышем, который он получит при этом же состоянии  $\Pi_j$ , выбрав стратегию  $A_i$ :

$$r_{ij} = \beta_j - a_{ij}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n. \quad (20.4)$$

Таким образом, риск  $r_{ij}$  игрока  $A$  — это упущенная им возможность максимального выигрыша  $\beta_j$  при этом состоянии природы. Величину риска можно ещё интерпретировать как своеобразную плату за отсутствие информации о состоянии природы.

Риск измеряется в тех же единицах, что и выигрыши.

Из рисков можно составить матрицу рисков (сожалений)  $R = R(A)$ , которая имеет следующий вид:

$\Pi_j \backslash A_i$	$\Pi_1$	$\Pi_2$	$\dots$	$\Pi_n$
$A_1$	$r_{11}$	$r_{12}$	$\dots$	$r_{1n}$

$$R(A)= \begin{array}{|c|c|c|c|c|} \hline A_2 & r_{21} & r_{22} & \dots & r_{2n} \\ \hline \dots & \dots & \dots & \dots & \dots \\ \hline A_m & r_{m1} & r_{m2} & \dots & r_{mn} \\ \hline \end{array} \quad (20.5)$$

Из определения рисков (20.4) следует, что каждый столбец матрицы рисков  $R$  содержит, по меньшей мере, один нуль. Следовательно, существуют строки матрицы  $R$ , содержащие нули.

Матрица выигрышей  $A$  однозначно порождает матрицу рисков (20.5), поскольку каждый риск  $r_{ij}$  однозначно определяется по формуле (20.4). Обратное неверно: одна и та же матрица рисков может порождаться различными матрицами выигрышей.

Равные выигрыши при одной и той же стратегии и при различных состояниях природы могут быть неравноценными в смысле рисков. Одинаковые же выигрыши при разных стратегиях, но при одном и том же состоянии природы всегда равноценны.

## **21. Бескоалиционные неантагонистические игры, их применение.**

Более сложными теоретико-игровыми моделями, выходящими за некоторые ограничения антагонистических игр, являются игры с переменной суммой. Такие игры описывают ситуации, в которых интересы игроков хотя и не совпадают, но не обязательно являются противоположными. Класс игр, в которых каждый игрок принимает решение изолированно, т.е. без договоренностей, переговоров, соглашений или коалиций с другими игроками, называется *бескоалиционные игры*. Поскольку такие игры представляются как математическая модель взаимодействия нескольких сторон, в процессе которого игроки не могут кооперироваться и координировать свои действия, то их ещё называют *некооперативными играми*.

Бескоалиционные игры называются *статическими*, если в них количество ходов каждого игрока равно единице. Такие игры обычно задаются в нормальной форме. *Динамические* игры предполагают серии ходов игроков, для описания которых зачастую используется дерево решений (развернутая форма игры). Бескоалиционная игра *конечная*, если конечны все множества стратегий конечного числа игроков (в противном случае игра *бесконечная*).

В нормальной форме некооперативная игра  $n$  участников – это система, состоящая из множества игроков, множества их стратегий и наборов их функций выигрыша:

$$\Gamma = \langle I, S = \prod_{i \in I} \{S_i\}, \{H_i\}_{i \in I} \rangle,$$

где  $I = \{1, \dots, n\}$  – множество игроков,  $S_i$  – множество стратегий,  $H_i: S = \prod_{i \in I} S_i \rightarrow R^1$  – функция выигрышей игрока  $i$ , ставящая в соответствие каждому набору стратегий выигрыш данного игрока.

Некооперативная игра в развёрнутой форме с множеством игроков  $I$  представляется с использованием ориентированного дерева (дерева игры) следующим образом.

Вершины дерева представляют собой *состояния (позиции)*, в которых может оказываться игра, рёбра – *ходы*, которые могут использовать игроки. Предполагается, что в каждой позиции может совершать ход не более одного игрока. Выделяется три вида позиций в игре:

- начальная, представляемая корнем дерева (вершиной, не имеющей входящих рёбер);
- промежуточные, имеющие входящие и выходящие рёбра;
- терминальные, имеющие только входящие ребра.

Начальная и промежуточные позиции образуют множество *нетерминальных* позиций.

Для каждой вершины дерева  $v$ , соответствующей нетерминальной позиции, определен игрок  $i$ , совершающий в ней ход и множество ходов

этого игрока  $X_v$ . Каждому ходу  $x \in X_v$ , соответствует ребро, выходящее из вершины  $v$ .

Для каждой вершины  $v$ , соответствующей терминальной позиции, определены функции выигрыша всех игроков  $H_i(v)$ .

Игра предлагает следующий порядок разыгрывания:

1. Игра начинается из начальной позиции.
2. В любой нетерминальной позиции  $v$  игрок, имеющий в ней право хода, выбирает ход  $x \in X_v$ , в результате чего игра попадает в следующую позицию, в которую входит ребро, соответствующее ходу  $x$ .

Если эта позиция является нетерминальной, то повторяется п. 2.

Если игра попадает в терминальную позицию  $v$ , то все игроки получают выигрыши  $H_i(v)$ , и игра завершается.

Наиболее простым случаем бескоалиционных игр с переменной суммой является неантагонистическая игра двух лиц с конечным числом стратегий (например, прототипные игры «Борьба за рынок», «Семейный спор», «Дилемма узников» и др.). В таких играх каждый из игроков делает один ход – выбирает одну стратегию из имеющегося у него конечного числа стратегий, и после этого он получает свой выигрыш согласно определённым для каждого из них матрицам выигрышей. То есть подобная игра полностью определяется двумя матрицами выигрышей для двух игроков. Поэтому такие игры называются *биматричными*.

Пусть у игрока 1 имеется  $m$  стратегий,  $i = \overline{1, m}$ , у игрока 2 имеется  $n$  стратегий,  $j = \overline{1, n}$ . Выигрыши игроков 1 и 2 соответственно задаются матрицами:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}; \quad B = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \dots & \dots & \dots \\ b_{m1} & \dots & b_{mn} \end{pmatrix}. \quad (21.1)$$

Если считать смешанной стратегией игрока 1 полный набор вероятностей применения 1 игроком своих чистых стратегий  $x = (x_1, \dots,$

$x_m$ ), а  $y = (y_1, \dots, y_n)$  – смешанной стратегией игрока 2, то их средние выигрыши будут соответственно равны

$$\begin{cases} H_1(A, x, y) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j = xAy^T \\ H_2(B, x, y) = \sum_{i=1}^m \sum_{j=1}^n b_{ij} x_i y_j = xBy^T \end{cases} \quad (21.2)$$

Ситуация  $s^*$  является *равновесием по Нэшу*, если изменение своей стратегии не выгодно ни одному игроку, то есть для любого игрока  $i$  и любой его стратегии  $s \in S$  выполняется неравенство  $H_i(s^*) \geq H_i(s_i, s_{-i}^*)$ . При такой ситуации любому игроку не выгодно отклоняться от неё, если все остальные игроки её придерживаются.

Нэш доказал, что если разрешить смешанные стратегии, тогда в каждой игре для  $n$  игроков будет хотя бы одно равновесие Нэша. Разумеется, это в полной мере распространяется и на биматричные игры. Если равновесие по Нэшу в биматричной игре существует в чистых стратегиях, то решать систем (21.2) не имеет смысла – оптимальные стратегии и выигрыши игроков могут быть получены по принципу доминирования. Чтобы определить равновесие по Нэшу для игры с матрицами (21.1) (их для удобства записывают в виде одной матрицы с элементами  $(a_{ij}, b_{ij})_{n \times m}$ ) нужно для каждого столбца отметить максимальный элемент по первой координате, для каждой строки – максимальный элемент по второй координате. Ситуация равновесия  $(i, j)$  соответствует элементам матрицы, у которых выделены обе координаты.

Если в равновесии по Нэшу стратегия каждого из игроков является лучшим ответом на действия другого игрока, то **равновесием по Парето** называется такая ситуация, при которой невозможно увеличить выигрыш одного игрока без нанесения ущерба другому. *Парето-оптимальное состояние рынка* – ситуация, когда нельзя улучшить положение любого участника экономического процесса, одновременно не снижая благосостояния как минимум одного из остальных.

## 22. Производственная функция и ее свойства. Производственная функция с постоянной эластичностью замещения факторов производства и ее частные случаи.

Производственная функция  $y = f(x_1, x_2, \dots, x_n)$  – это функция, независимые переменные которой принимают значения объемов затрачиваемых ресурсов, а значение функции равно объёму выпускаемой продукции. Вектор  $\vec{x} = (x_1, x_2, \dots, x_n)$  называется конфигурацией ресурсов.

### Примеры производственных функций ( $n = 2$ )

Функция Кобба – Дугласа:

$$y = \alpha_0 \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2}, \quad (22.1)$$

$$\alpha_0 > 0, \alpha_1 > 0, \alpha_2 > 0, \alpha_1 + \alpha_2 = 1. \quad (22.1. a)$$

Линейная функция:

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2, \quad (22.2)$$

$$b_0 \geq 0, b_1 > 0, b_2 > 0. \quad (22.2. a)$$

Функция Леонтьева:

$$y = \min(c_1 \cdot x_1; c_2 \cdot x_2), \quad (22.3)$$

$$c_1 > 0, c_2 > 0. \quad (22.3. a)$$

Основные свойства производственной функции вводятся при помощи двух аксиом:

*Аксиома 1.* В пространстве  $R_+^n$  факторов производства существует экономическая область  $S$ , в которой производственная функция не убывает по каждому аргументу, т.е.:

$$\Delta y = M_y(x_i) = f(x_1, x_2, \dots, x_i + 1, \dots, x_n) - f(x_1, x_2, \dots, x_i, \dots, x_n) \geq 0. \quad (22.4)$$

Дополнительная единица каждого фактора производства приводит к дополнительному уровню  $\Delta y$  выпуска продукции. Величина  $\Delta y = M_y(x_i)$  называется предельным продуктом  $i$ -го фактора производства:

$$M_y(x_i) \approx \frac{\partial f(x_1, x_2, \dots, x_i, \dots, x_n)}{\partial x_i}. \quad (22.5)$$

*Аксиома 2.* Производственная функция выпукла вверх, т.е. каждая последующая дополнительная единица фактора гарантирует меньше дополнительной продукции, чем предыдущая дополнительная единица:

$$M_y(x_i) \rightarrow 0 \text{ при } x_i \rightarrow +\infty. \quad (22.6)$$

$$M_y(x_i^0) > M_y(x_i^0 + 1) \quad (22.7)$$

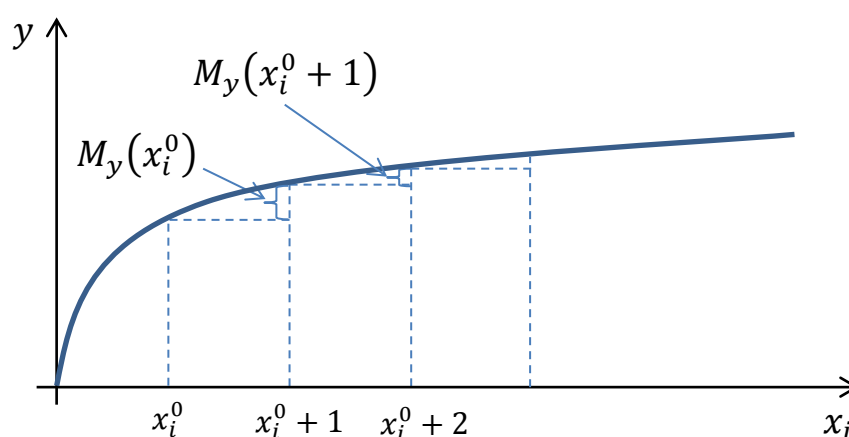


Рис. 4. Закон убывающей полезности фактора производства.

### **Свойства производственной функции.**

*Свойство 1.* Необходимость каждого фактора производства:

$$f(x_1, x_2, \dots, x_i = 0, \dots, x_n) = 0 \quad \forall i = 1, 2, \dots, n. \quad (22.8)$$

*Свойство 2.* Неотрицательность предельного продукта  $i$ -го фактора:

$$\forall x_i > 0 \quad \frac{\partial f(x_1, x_2, \dots, x_i, \dots, x_n)}{\partial x_i} \geq 0, \quad (22.9)$$

т.е., при увеличении затрат  $x_i$   $i$ -го ресурса при условии постоянства затрат прочих факторов объем выпуска продукции возрастает (не убывает).

*Свойство 3.* Невозрастающая полезность фактора производства:

$$\forall x_i > 0 \quad \frac{\partial^2 f(x_1, x_2, \dots, x_i, \dots, x_n)}{\partial x_i^2} \leq 0, \quad (22.10)$$

т.е., при увеличении затрат  $x_i$   $i$ -го ресурса при условии неизменных затрат прочих ресурсов предельный продукт  $i$ -го фактора производства снижается (не возрастает).

*Свойство 4. Однородность производственной функции.*

$\exists \rho > 0$ , что для  $\forall \vec{x} = (x_1, x_2, \dots, x_n)$  справедливо равенство:

$$f(\mu x_1, \mu x_2, \dots, \mu x_n) = \mu^\rho f(x_1, x_2, \dots, x_n). \quad (22.11)$$

Перечисленные свойства справедливы в экономической области производственной функции. За её пределами поведение производственной функции может отличаться от перечисленных свойств.

### **Эластичность замещения факторов производства ( $n = 2$ )**

Определим понятие эластичности замещения факторов производства на примере производственной функции с двумя факторами производства  $y = f(x_1, x_2)$ , т.е. для случая  $n = 2$ .

Эластичность  $E_x^y$  замещения факторов производства показывает, на сколько процентов необходимо заменить соотношение объемов  $x_1$  и  $x_2$  факторов производства, чтобы предельная норма замещения ( $MRS^5$ ) изменилась на 1%.

Эластичность переменной  $y$  по переменной  $x$  описывается формулой относительной производной:

$$E_x^y = \frac{dy/y}{dx/x} = \frac{dy}{dx} \cdot \frac{x}{y}. \quad (22.12)$$

Используя (22.12), определим эластичность замещения факторов производства:

$$E_2^1 = - \frac{d\left(\frac{x_1}{x_2}\right)}{dMRS_{12}} \cdot \frac{MRS_{12}}{\frac{x_1}{x_2}}. \quad (22.13)$$

$$MRS_{12} = \frac{\partial f(x_1, x_2)}{\partial x_1} / \frac{\partial f(x_1, x_2)}{\partial x_2} = \frac{M_y(x_1)}{M_y(x_2)}, \quad (22.13.1)$$

$$y = f(x_1, x_2). \quad (22.13.2)$$

---

<sup>5</sup> Marginal Rate of Substitution.



Рассмотрим (22.13) более подробно:

$$\begin{aligned}
 E_2^1 &= - \frac{d\left(\frac{x_1}{x_2}\right)}{d\left[\frac{\partial f(x_1, x_2)}{\partial x_1} / \frac{\partial f(x_1, x_2)}{\partial x_2}\right]} \cdot \frac{\frac{\partial f(x_1, x_2)}{\partial x_1} / \frac{\partial f(x_1, x_2)}{\partial x_2}}{\frac{x_1}{x_2}} = \\
 &= - \frac{d\left(\frac{x_1}{x_2}\right) / \frac{x_1}{x_2}}{d\left[\frac{\partial f(x_1, x_2)}{\partial x_1} / \frac{\partial f(x_1, x_2)}{\partial x_2}\right] / \left[\frac{\partial f(x_1, x_2)}{\partial x_1} / \frac{\partial f(x_1, x_2)}{\partial x_2}\right]} = \\
 &= - \frac{d \ln \frac{x_1}{x_2}}{d \ln \frac{\partial f(x_1, x_2)}{\partial x_1} / \frac{\partial f(x_1, x_2)}{\partial x_2}}.
 \end{aligned} \tag{22.14}$$

Можно показать, что

$$E_1^2 = - \frac{d \ln \frac{x_2}{x_1}}{d \ln \left[\frac{\partial f(x_1, x_2)}{\partial x_2} / \frac{\partial f(x_1, x_2)}{\partial x_1}\right]}. \tag{22.15}$$

Очевидно, что  $E_2^1 = E_1^2$ .

### Производственная функция с постоянной эластичностью замещения (CES<sup>6</sup>-функция)

Рассмотрим производственную функцию вида:

$$y = \alpha_0 \cdot (a_1 x_1^{-\alpha} + a_2 x_2^{-\alpha})^{-h/\alpha}, \tag{22.16}$$

$$\alpha_0 > 0, a_1 > 0, \alpha > 0, h > 0. \tag{22.16. a}$$

Найдём для (22.16) эластичность замещения первого фактора вторым:

$$E_2^1 = - \frac{d \ln \frac{x_1}{x_2}}{d \ln \left[\frac{a_1}{a_2} \left(\frac{x_1}{x_2}\right)^{-1-\alpha}\right]} = \frac{1}{1 + \alpha}. \tag{22.17}$$

Функция (22.16) называется *CES-функцией*, или функцией с постоянной эластичностью замещения. Функции (22.1), (22.2) и (22.3) являются частными случаями функции (22.16).

---

<sup>6</sup> Constant Elasticity of Substitution.

Покажем, что производственная функция (22.1) Кобба-Дугласа является частным случаем функции (22.16) при  $\alpha \rightarrow 0$ :

$$y = \alpha_0 \cdot (a_1 x_1^{-\alpha} + a_2 x_2^{-\alpha})^{-h/\alpha} = \frac{\alpha_0 x_2^h}{\left[ \left( \frac{x_2}{x_1} \right)^\alpha a_1 + a_2 \right]^{\frac{h}{\alpha}}}. \quad (22.18)$$

Если  $a_1 + a_2 > 1$ ,  $h > 0$ , то при  $\alpha \rightarrow 0$

$$\left[ \left( \frac{x_2}{x_1} \right)^\alpha a_1 + a_2 \right]^{\frac{h}{\alpha}} \rightarrow (a_1 + a_2)^\infty = \infty \Rightarrow y = 0. \quad (22.19)$$

Если  $a_1 + a_2 < 1$ ,  $h > 0$ , то при  $\alpha \rightarrow 0$

$$\left[ \left( \frac{x_2}{x_1} \right)^\alpha a_1 + a_2 \right]^{\frac{h}{\alpha}} \rightarrow (a_1 + a_2)^\infty = 0 \Rightarrow y = \infty. \quad (22.20)$$

Если  $a_1 + a_2 = 1$ , то

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \left[ \left( \frac{x_2}{x_1} \right)^\alpha a_1 + a_2 \right]^{\frac{h}{\alpha}} &= (1^\infty) = \lim_{\alpha \rightarrow 0} e^{\ln \left[ \left( \frac{x_2}{x_1} \right)^\alpha a_1 + a_2 \right]^{\frac{h}{\alpha}}} = \\ &= e^{\lim_{\alpha \rightarrow 0} \frac{h \cdot \ln \left[ \left( \frac{x_2}{x_1} \right)^\alpha a_1 + a_2 \right]}{\alpha}} = e^{\frac{0}{0}} = e^{h \cdot \lim_{\alpha \rightarrow 0} \frac{\ln \left[ \left( \frac{x_2}{x_1} \right)^\alpha a_1 + a_2 \right]}{\alpha}} = \\ &= e^{h \cdot \lim_{\alpha \rightarrow 0} \frac{a_1 \left( \frac{x_2}{x_1} \right)^\alpha \ln \frac{x_2}{x_1}}{\left( \frac{x_2}{x_1} \right)^\alpha a_1 + a_2}} = e^{\ln \left( \frac{x_2}{x_1} \right)^{ha_1}} = \left( \frac{x_2}{x_1} \right)^{ha_1}, \end{aligned} \quad (22.21)$$

ПОЭТОМУ

$$\lim_{\alpha \rightarrow 0} \alpha_0 \cdot (a_1 x_1^{-\alpha} + a_2 x_2^{-\alpha})^{-h/\alpha} = \frac{\alpha_0 x_2^h}{\left( \frac{x_2}{x_1} \right)^{ha_1}} = \alpha_0 x_1^{ha_1} x_2^{h-ha_1}. \quad (22.22)$$

При  $\alpha \rightarrow \infty$  и  $h = 1$  частным случаем CES-функции (22.16) является функция Леонтьева (22.3).

При  $\alpha = -1$  и  $h = 1$  частным случаем CES-функции (22.16) является линейная функция (22.2).

### 23. Задача управления производственным сектором экономики и структурная модель Леонтьева «затраты – выпуск».

Модель Леонтьева предназначена для управления производственным сектором экономики, который состоит из  $n$  отраслей. Исходными данными (экзогенными переменными) задачи считаются уровни  $Y$  конечной продукции по отраслям за планируемый отрезок времени

$$\vec{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}. \quad (23.1)$$

Искомыми величинами (эндогенными переменными) являются:

1. валовые выпуски  $\vec{X}$  отраслей на том же отрезке времени, что и для  $\vec{Y}$ :

$$\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}; \quad (23.2)$$

2. межотраслевые поставки  $X = \|x_{ij}\|, i, j = 1, 2, \dots, n$ .

Межотраслевые поставки – части валовых выпусков отраслей, которые предназначены производственному сектору для обеспечения технологических процессов. Например, символом  $x_{61}$  обозначено количество продукции шестой отрасли, необходимое первой отрасли для обеспечения искомого валового выпуска  $x_1$ .

Структурная форма модели Леонтьева имеет вид:

$$\vec{X} = A\vec{X} + \vec{Y}. \quad (23.3)$$

В (23.3)  $A = \|a_{ij}\|$  – матрица технологических коэффициентов – параметров модели. Величина  $a_{ij}$  показывает среднюю величину непосредственных затрат продукции  $i$ -й отрасли на выпуск единицы продукции  $j$ -й отрасли. Например, если  $a_{ij} = 0.31$ , то из этого следует, что  $j$ -й отрасли в процессе валового выпуска её продукции на 1 денежную единицу требовалось 0.31 денежной единицы продукции  $i$ -й отрасли.

Приведённую форму модели Леонтьева получим, выразив вектор  $\vec{X}$  через вектор  $\vec{Y}$ :

$$\vec{X} = (E - A)^{-1}\vec{Y}, \quad (23.4)$$

$E$  – единичная матрица.

После того, как получены конкретные значения  $x_i$ , можно рассчитать элементы матрицы межотраслевых поставок по формуле:

$$x_{ij} = a_{ij}x_j. \quad (23.5)$$

Матрицу  $B = (E - A)^{-1} = \|b_{ij}\|$  называют мультипликатором Леонтьева, или матрицей коэффициентов полных материальных затрат.

Коэффициент  $b_{ij}$  отражает количество валовой продукции  $i$ -й отрасли, необходимое для выпуска единицы конечной продукции  $j$ -й отрасли.

Запишем уравнение (23.4) для  $i$ -й отрасли:

$$x_i = \sum_{j=1}^n b_{ij}y_j, \quad j = 1, 2, \dots, n. \quad (23.6)$$

Зафиксируем номер  $j$  и примем следующие значения величин конечного выпуска по отраслям:

$$y_1 = 0, \dots, y_{j-1} = 0, y_j = 1, y_{j+1} = 0, \dots, y_n = 0. \quad (23.7)$$

Подставляя (23.7) в (23.6), получим,

$$x_i = b_{ij}. \quad (23.8)$$

Поскольку конечный спрос  $y_j = 1$  продукции  $j$ -й отрасли (при нулевых конечных спросах продукции других отраслей) обеспечивается валовыми выпусками

$$x_1 = b_{1j}, \dots, x_i = b_{ij}, \dots, x_n = b_{nj}. \quad (23.9)$$

всех отраслей производственного сектора, то коэффициенты  $b_{ij}$  именуются коэффициентами полных материальных затрат.

В своей работе «Экономические эссе»<sup>7</sup> В. Леонтьев, автор модели «Затраты – Выпуск», в качестве примера рассматривает 8-отраслевую экономику. В этом примере рассматриваются направления экономической деятельности, условно названные следующим образом:

1. продукты питания и лекарства;
2. ткани, одежда, мебель;
3. оборудование и машины;
4. транспортные средства и бытовая техника;
5. строительство;
6. металлы;
7. энергия;
8. химические продукты.

В рамках модели Леонтьева, например, для шестой и восьмой отраслей можно записать следующие соотношения:

$$\begin{cases} x_6 = a_{66}x_6 + a_{68}x_8 + y_6, \\ x_8 = a_{86}x_6 + a_{88}x_8 + y_8. \end{cases} \quad (23.10)$$

Зная величины конечного продукта отраслей  $y_6$ ,  $y_8$  и технологические коэффициенты  $a_{66}$ ,  $a_{68}$ ,  $a_{86}$ ,  $a_{88}$  можно из (23.10) определить значения валовых продуктов  $x_6$  и  $x_8$ .

Величина  $c_j$  материальных затрат  $j$ -й отрасли равна сумме межотраслевых потоков в эту отрасль, т.е.

$$c_j = \sum_{i=1}^n x_{ij}. \quad (23.11)$$

Добавленная стоимость  $v_j$   $j$ -й отрасли рассчитывается по формуле:

$$v_j = x_j - c_j. \quad (23.12)$$

Модель Леонтьева включает тождество межотраслевого баланса:

$$\sum_{i=1}^n y_i = \sum_{j=1}^n v_j. \quad (23.13)$$

---

<sup>7</sup> Леонтьев В. Экономические эссе. Теории, исследования, факты и политика. – М.: Политиздат, 1990, С. 190 – 191.

Рассмотрим понятие продуктивности матрицы  $A$  технологических коэффициентов. Матрица  $A$  называется продуктивной, если существует матрица  $B$  с неотрицательными коэффициентами. Достаточным условием продуктивности является:

$$\max |\lambda_i(A)| < 1, \quad (23.14)$$

где  $\lambda_i(A)$  – собственное число матрицы  $A$ .

Для проверки условия (23.14) необходимо построить характеристический полином матрицы и решить соответствующее уравнение относительно  $\lambda$ :

$$|A - \lambda E| = 0, \quad (23.15)$$

$\lambda$  – вектор собственных чисел матрицы  $A$ .

Модель «Затраты – Выпуск» может быть использована также для расчета уровней конечного спроса  $\vec{Y}$  по известному совокупному спросу  $\vec{X}$ :

$$\vec{Y} = (E - A)\vec{X}. \quad (23.16)$$

Рассмотрим условную экономику, состоящую из трёх отраслей. Известны коэффициенты прямых материальных затрат (матрица  $A$ ) и совокупный валовой продукт  $\vec{X}$ . Какой уровень конечного потребления продукции обеспечивает экономика при заданных технологических коэффициентах и заданных уровнях совокупного спроса?

Отрасли народного хозяйства	Коэффициенты прямых материальных затрат			Совокупный спрос, млрд. руб.
	1	2	3	
1	$a_{11}$	$a_{12}$	$a_{13}$	$x_1$
2	$a_{21}$	$a_{22}$	$a_{23}$	$x_2$
3	$a_{31}$	$a_{32}$	$a_{33}$	$x_3$

Уровни конечной продукции (конечного потребления) в экономике по отраслям вычисляются по формулам:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 - a_{11} & -a_{12} & -a_{13} \\ -a_{21} & 1 - a_{22} & -a_{23} \\ -a_{31} & -a_{32} & 1 - a_{33} \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \quad (23.17)$$

$$y_1 = (1 - a_{11})x_1 - a_{12}x_2 - a_{13}x_3, \quad (23.18a)$$

$$y_2 = -a_{21}x_1 + (1 - a_{22})x_2 - a_{23}x_3, \quad (23.18b)$$

$$y_3 = -a_{31}x_1 - a_{32}x_2 + (1 - a_{33})x_3. \quad (23.18c)$$

Сопутствующие материальные затраты по отраслям:

$$X = \begin{pmatrix} a_{11}x_1 & a_{12}x_2 & a_{13}x_3 \\ a_{21}x_1 & a_{22}x_2 & a_{23}x_3 \\ a_{31}x_1 & a_{32}x_2 & a_{33}x_3 \end{pmatrix}, \quad (23.19)$$

где  $X$  – матрица межотраслевых поставок.

$$\vec{C} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{21}x_1 + a_{31}x_1 \\ a_{12}x_2 + a_{22}x_2 + a_{32}x_2 \\ a_{13}x_3 + a_{23}x_3 + a_{33}x_3 \end{pmatrix}, \quad (23.20)$$

$$c_1 = (a_{11} + a_{21} + a_{31})x_1, \quad (23.21a)$$

$$c_2 = (a_{12} + a_{22} + a_{32})x_2, \quad (23.21b)$$

$$c_3 = (a_{13} + a_{23} + a_{33})x_3, \quad (23.21c)$$

где  $\vec{C}$  – вектор суммарных материальных затрат по отраслям.

**24. Реляционная модель данных. Нормализация отношений. Функциональные зависимости. Нормальные формы. Алгоритмы нормализации. Влияние нормализации на эффективность работы с базой данных.**

*Процесс разработки корректной схемы реляционной БД называется логическим проектированием БД. В реляционных БД логическое проектирование приводит к разработке схемы БД, то есть совокупности схем отношений, которые адекватно моделируют абстрактные объекты предметной области и семантические связи между этими объектами.*

При представлении концептуальной схемы в виде реляционной модели возможны различные варианты выбора схем отношений, то есть одну и ту же информацию можно задать разным набором таблиц.

От правильного выбора схем отношений, представляющих концептуальную схему, в значительной степени будет зависеть эффективность функционирования базы данных.

Рассмотрим *пример* конкретной схемы отношений и проанализируем её недостатки.

Предположим, что данные об абитуриентах, специальностях, формах обучения включены в одну таблицу, содержащую данные об атрибутах. Схема отношения имеет следующий вид:

АБИТУРИЕНТ (код абитуриента, Ф.И.О., факультет, специальность, форма обучения).

Эта схема отношений обуславливает следующие недостатки соответствующей базы данных:

1. Дублирование информации (избыточность). У абитуриентов, поступающих на один факультет, будет повторяться название факультета. Для разных факультетов будут повторяться одинаковые формы обучения, специальности.
2. Потенциальная противоречивость (аномалии обновления). Если, например, изменится название специальности, то, изменяя её в одном кортеже (у одного абитуриента), необходимо изменять и во всех других кортежах, где она присутствует.
3. Потенциальная возможность потери сведений (аномалии удаления). При удалении всех абитуриентов, поступающих на определенную специальность, мы теряем все сведения об этой специальности.
4. Потенциальная возможность не включения информации в базу данных (аномалии включения). В базе данных будут отсутствовать сведения о специальности, если на нее абитуриенты не подали заявления.

В теории реляционных баз данных существуют формальные методы построения реляционной модели базы данных, в которой отсутствует избыточность и аномалии обновления, удаления и включения.

Э.Ф. Кодд доказал, что, следуя при создании таблиц и связей между ними только немногим формализованным правилам, можно обеспечить корректность манипулирования данными.



Исходной точкой является представление предметной области в виде одного отношения, и на каждом шаге проектирования производится некоторый набор схем отношений, обладающих «улучшенными» свойствами.

Построение рационального варианта схем отношений (обладающих лучшими свойствами при операциях включения, модификации и удаления данных, чем все остальные наборы схем, а также отсутствует избыточность) осуществляется путем нормализации схем отношений.

В теории реляционных баз данных обычно выделяется следующая **последовательность нормальных форм:**

- первая нормальная форма (1NF);
- вторая нормальная форма (2NF);
- третья нормальная форма (3NF);
- нормальная форма Бойса-Кодда (BCNF);
- четвертая нормальная форма (4NF);
- пятая нормальная форма, или нормальная форма проекции-соединения (5NF или PJ/NF).

*Каждая следующая нормальная форма обладает свойствами, в некотором смысле, лучшими, чем предыдущая.*

*Каждой нормальной форме соответствует определенный набор ограничений, и отношение находится в некоторой нормальной форме, если удовлетворяет свойственному ей набору ограничений.*

*При переходе к следующей нормальной форме свойства предыдущих нормальных форм сохраняются.*

Сначала были предложены только первая (1НФ), вторая (2НФ) и третья (3НФ) нормальные формы. Затем Бойсом и Коддом было сформулировано более строгое определение третьей нормальной формы, которое получило название нормальной формы Бойса-Кодда (НФБК). Все эти нормальные формы предназначены для устранения различных видов

нежелательных функциональных зависимостей между атрибутами отношения.

На начальном этапе схема отношений должна находиться в первой нормальной форме (1НФ). Отношение находится в первой нормальной форме, если все атрибуты отношения принимают простые значения (атомарные или неделимые), не являющимися множеством или кортежем из более элементарных составляющих. Далее отношение, представленное в первой нормальной форме, последовательно преобразуется во вторую и третью нормальные формы.

В большинстве случаев достижение третьей нормальной формы или формы Бойса—Кодда считается достаточным для реальных проектов баз данных (при некоторых предположениях о данных третья нормальная форма является искомым наилучшим вариантом).

Если эти предположения не выполняются, то процесс нормализации продолжается и отношение преобразуется в четвертую нормальную форму.

Пятая нормальная форма имеет больше теоретическое значение.

*Декомпозиция схемы отношения.* Последовательный переход от одной нормальной формы к другой при нормализации схем отношений осуществляется с помощью декомпозиции. Основной операцией, с помощью которой осуществляется декомпозиция, является *проекция*.

Декомпозицией схемы отношения  $R = \{A_1, A_2, \dots, A_n\}$  называется замена ее совокупностью подмножеств  $R_i$ , таких, что их объединение дает  $R$ . При этом допускается, чтобы подмножества были пересекающимися.

Декомпозиция должна обеспечить то, что запросы (выборка данных по условию) к исходному отношению и отношениям, получаемым в результате декомпозиции, дадут одинаковый результат. Соответствующее условие будет выполняться, если каждый кортеж отношения  $R$  может быть представлен как естественное соединение его проекций на каждое из подмножеств. В этом случае говорят, что декомпозиция обладает *свойством соединения без потерь*.

Вторым важнейшим желательным свойством декомпозиции является *свойство сохранения функциональных зависимостей*. Одной из основных концепций нормализации является функциональная зависимость, которая описывает связь между атрибутами отношения.

*Определение.* Пусть  $A$  и  $B$  это атрибуты некоторого отношения  $R$ . Атрибут  $B$  функционально зависит от атрибута  $A$  (что обозначается как  $A \rightarrow B$ ) если каждое значение  $A$  связано с одним значением  $B$ . (Причем каждый из атрибутов  $A$  и  $B$  может состоять из одного или нескольких атрибутов.)

*Определение.* Пусть  $R(A_1, A_2, \dots, A_n)$  – схема отношения, а  $X$  и  $Y$  – подмножества  $\{A_1, A_2, \dots, A_n\}$ . Тогда  $X$  функционально определяет  $Y$  или  $Y$  функционально зависит от  $X$  (обозначается  $X \rightarrow Y$ ) тогда и только тогда, когда каждое значение множества  $X$  отношения  $R$  связано с одним значением множества  $Y$  отношения  $R$ . Иначе говоря, если два кортежа  $R$  совпадают по значению  $X$ , они совпадают и по значению  $Y$ . Такая функциональная зависимость (ФЗ) обозначается как  $F: X \rightarrow Y$ .

*Определение.* Функциональной зависимостью набора атрибутов  $Y$  отношения  $R$  от набора атрибутов  $X$  того же отношения, обозначаемой как  $R: X \rightarrow Y$  или  $X \rightarrow Y$  называется такое соотношение проекций  $R[X]$  и  $R[Y]$ , при котором в каждый момент времени любому элементу проекции  $R[X]$  соответствует только один элемент проекции  $R[Y]$ , входящий вместе с ним в какой-либо кортеж отношения  $R$ .

**Детерминантом** называется любой атрибут, от которого полностью функционально зависит какой-то другой атрибут.

В определении функциональной зависимости термин "детерминант" характеризует один или несколько атрибутов, расположенных с левой стороны от стрелки.

*Пример* функциональной зависимости:




ГРАФИК\_ПОЛЕТОВ (Пилот, Рейс, Дата\_вылета, Время\_вылета).

Пилот	Рейс	Дата вылета	Время вылета
Иванов	100	8.07	10:20
Иванов	102	9.07	13:30
Исаев	90	7.07	6:00
Исаев	100	11.07	10:20
Исаев	103	10.07	19:30
Петров	100	12.07	10:20
Петров	102	11.07	13:30
Фролов	90	8.07	6:00
Фролов	90	12.07	6:00
Фролов	104	14.07	13:30

Известно, что:

- каждому рейсу соответствует определенное время вылета;
- для каждого пилота, даты и времени вылета возможен только один рейс;
- на определенный день и рейс назначается определенный пилот.

Следовательно:

- "Время\_вылета" функционально зависим от "Рейс": "Рейс"  "Время\_вылета";
- "Рейс" функционально зависим от {"Пилот", "Дата вылета", "Время вылета"}: {"Пилот", "Дата вылета", "Время вылета"}  "Рейс";
- "Пилот" функционально зависим от {"Рейс", "Дата вылета"}: {"Рейс", "Дата вылета"}  "Пилот".

Эти функциональные зависимости отражают взаимосвязи, обнаруженные между объектами предметной области. Они являются дополнительными ограничениями, определяемыми предметной областью.

Функциональная зависимость – это семантическое понятие. Она возникает, когда по значениям одних данных в предметной области можно определить значения других данных.

## Транзитивная функциональная зависимость

*Определение.* Функциональная зависимость  $R.A \twoheadrightarrow R.B$  называется **транзитивной**, если существует набор атрибутов  $C$  такой, что:

- $C$  не является подмножеством  $A$ .
- $C$  не включает в себя  $B$ .
- Существует функциональная зависимость  $R.A \twoheadrightarrow R.C$ .
- Не существует функциональной зависимости  $R.C \twoheadrightarrow R.A$ .
- Существует функциональная зависимость  $R.C \twoheadrightarrow R.B$ .

*Пример отношения СотрудникиОтделыПроекты*

(первичный ключ Номер Сотрудника, Номер Проекта).

Номер Сотрудника	Зарплата Сотрудника	Номер Отдела	Номер Проекта	Задание Сотрудника
1	11000	12	1	1
1	9000	12	2	1
2	15000	4	1	2
3	12000	12	1	3
3	16000	12	2	2

Функциональные зависимости:

- номерСотрудника  $\rightarrow$  зарплатаСотрудника
- номерСотрудника  $\rightarrow$  номерОтдела
- номерОтдела  $\rightarrow$  зарплатаСотрудника
- номерСотрудника, номерПроекта  $\rightarrow$  заданиеСотрудника

Для рассмотренного отношения СотрудникиОтделыПроекты функциональная зависимость «Номер Сотрудника  $\rightarrow$  Зарплата Сотрудника»

является **транзитивной**: она является следствием функциональных зависимостей

Номер Сотрудника  $\rightarrow$  Номер Отдела

Номер Отдела → Зарплата Сотрудника

и при этом Номер Сотрудника не определяется атрибутом Номер Отдела.

Это означает, что заработная плата сотрудника на самом деле является характеристикой не сотрудника, а отдела, в котором он работает.

Что плохого в наличии транзитивной функциональной зависимости?

1. Не удастся занести в БД информацию, характеризующую заработную плату отдела до тех пор, пока в этом отделе не появится хотя бы один сотрудник (первичный ключ не может содержать неопределённое значение).

2. При удалении кортежа, описывающего последнего сотрудника данного отдела, мы лишимся информации о заработной плате в отделе.

3. Чтобы согласованным образом изменить заработную плату отдела, мы будем вынуждены предварительно найти все кортежи, описывающие сотрудников этого отдела.

Таким образом, транзитивные функциональные зависимости в отношении порождают аномалии, которые нужно устранить.

*Замечание.* Функциональные зависимости характеризуют все отношения, которые могут быть значениями схемы отношения R в принципе. Поэтому единственный способ определить функциональные зависимости — внимательно проанализировать семантику (смысл) атрибутов.

Функциональные зависимости являются, в частности, ограничениями целостности, поэтому целесообразно проверять их при каждом обновлении базы данных.

Первая нормальная форма

*Определение.* Таблица находится в **первой нормальной форме**, тогда и только тогда, когда если значения всех ее полей атомарные (на пересечении каждого столбца и каждой строки находятся только

элементарные значения атрибутов), и в ней отсутствуют повторяющиеся группы полей.

Отношения, находящиеся в первой нормальной форме, часто называют просто **нормализованными отношениями**.

Соответственно, ненормализованные отношения с не атомарными полями могут интерпретироваться как таблицы с неравномерным заполнением. Например, таблица "Расписание", которая имеет вид:

Преподаватель	День недели	Номер пары	Название дисциплины	Тип занятий	Группа
Петров В. И.	Понедельник	1	Языки программирования	Лекция	4906
	Вторник	1	Компьютерная графика	Лабор. работа	4907
	Вторник	2	Компьютерная графика	Лабор. работа	4906
Киров В. А.	Понед.	2	Алгебра	Лекция	4906
	Вторник	3	Программирование на C++	Лабор. работа	4907
	Вторник	4	Программирование на C++	Лабор. работа	4906
Серов А. А.	Понед.	3	Защита информации	Лекция.	4944
	Среда	3	Программирование на VB	Лабор. работа	4942
	Четверг	4	Программирование на VB	Лабор. работа	4922

Здесь на пересечении одной строки и одного столбца находится целый набор элементарных значений, соответствующих набору дней, перечню пар, набору дисциплин, по которым проводит занятия один преподаватель.

Для приведения отношения "Расписание" к требуемому виду (атомарность) необходимо дополнить каждую строку фамилией преподавателя.

Преподаватель	День недели	Номер пары	Название дисциплины	Тип занятий	Группа
Петров В. И.	Понед.	1	Языки программирования	Лекция	4906

Петров В. И.	Вторник	1	Компьютерная графика	Лабор. работа.	4907
Петров В. И.	Вторник	2	Компьютерная графика	Лабор. работа	4906
Киров В. А.	Понед.	2	Алгебра	Лекция	4906
Киров В. А.	Вторник	3	Программирование на C++	Лабор. работа	4907
Киров В. А.	Вторник	4	Программирование на C++	Лабор. работа	4906
Серов А. А.	Понед.	3	Защита информации	Лекция.	4944
Серов А. А.	Среда	3	Программирование на VB	Лабор. работа	4942
Серов А. А.	Четверг	4	Программирование на VB	Лабор. работа	4922

Эта исходная таблица с атомарными значениями характеризуется избыточностью (повторяются фамилии, дни недели, номера пар, дисциплины, тип занятий, группа).

Для приведения таблицы с атомарными значениями атрибутов к 1НФ можно использовать следующий *алгоритм*:

1. Определить группы повторяющихся полей.
2. Вынести группы повторяющихся полей в отдельные таблицы (в данном примере Преподаватель, Название дисциплины ....).
3. Назначить первичные ключи в новых таблицах.

#### Преподаватели

КодПреподавателя	Преподаватель
1	Петров В. И.
2	Киров В. А.
3	Серов А. А.

#### Дисциплины

Коддисциплины	Название дисциплины
---------------	---------------------



1	Языки программирования
2	Компьютерная графика
3	Алгебра
4	Программирование на C++
5	Защита информации
6	Программирование на VB

#### Виды занятий

КодВидаЗанятия	Тип занятий
1	Лекция
2	Лабораторная работа

#### Группы

Кодгруппы	Группа
1	4906
2	4907
3	4944
4	4942
5	4922

*Примечание:* С одной стороны, номер группы уникален и может использоваться в качестве значения первичного ключа. С другой стороны, обычно в номер группы закладывается признак курса (1БИ1). В этом случае, при использовании номера группы в качестве первичного ключа, будет потеряна информация о группах прошлых лет.

Дни недели	
КодДняНедели	Деньнедели

1	Понед.
2	Вторник
3	Среда

4	Четверг
5	Пятница

1
2
3
4

Пары

Номер пары
------------

Занятие (Расписание)

КодЗанятия	КодПреподавателя	Коддисциплины	КодВидаЗанятия	Кодгруппы	Деньнедели	Номерпары
1	1	1	1	1	1	1
	.....	...	.....	...	...	...
	.	.....	.....	.....	..... ...	.....

Удовлетворение требованиям первой нормальной формы называется **структурной** или **синтаксической нормализацией**.

*Вторая нормальная форма (2НФ).* 2НФ применяется к таблицам, которые имеют составной ключ.

*Определение.* Отношение находится **во второй нормальной форме** тогда и только тогда, когда оно находится в первой нормальной форме и не содержит неполных функциональных зависимостей неключевых атрибутов от атрибутов первичного ключа (каждый неключевой атрибут зависит от первичного ключа (не зависит от части ключа)).

Или любое неключевое поле однозначно идентифицируется полным набором ключевых полей.

Поле, зависящее только от части ключа – **частично зависимое поле**.

Для перевода отношения в 2НФ необходимо, используя *операцию проекции*, разложить его на несколько отношений следующим образом:

1. Построить проекцию без атрибутов, находящихся в частичной функциональной зависимости от первичного ключа;
2. Построить проекции на части составного ключа и атрибуты, зависящие от этих частей.

2НФ применяется к отношениям (таблицам), которые имеют составной ключ. Любое неключевое поле однозначно идентифицируется полным набором ключевых полей.

#### Проекты

Код проекта
Название
Основная цель
Продолжительность
Код руководителя проекта
Фамилия руководителя
Телефон руководителя

Пусть таблица имеет составной ключ, который формируется по двум полям: «Код проекта», «Код руководителя проекта».

Поля "Название", "Основная цель", "Продолжительность", "Фамилия руководителя", "Телефон руководителя" являются **частично зависимыми**:

«Название», «Основная цель», «Продолжительность» зависят только от «Кода проекта», но не зависят от «Кода руководителя», т.е. однозначно идентифицируется частью ключа, а не полным набором ключевых полей.

Аналогично поля «Фамилия руководителя» и «Телефон руководителя» зависят только от «Кода руководителя проекта».

Для приведения к 2НФ необходимо:

1. Вынести все частично зависимые поля в отдельную таблицу.
2. Определить ключевые поля.
3. Установить отношения между таблицами.

#### Проекты

Код проекта
Название
Основная цель
Продолжительность
Код руководителя проекта

#### Руководители

Код руководителя проекта
Фамилия руководителя
Телефон руководителя

*Третья нормальная форма.*

*Определение.* **Отношение находится в 3НФ**, если оно находится во 2НФ и каждый неключевой атрибут нетранзитивно зависит от первичного ключа.

Отношение находится в 3НФ в том и только том случае, если все неключевые атрибуты отношения взаимно независимы и полностью зависят от первичного ключа.

Другими словами, в таблице нет полей, которые не зависят от ключа.

*Пример.*

Заказы

Код заказа
Фамилия покупателя
Дата продажи
Код менеджера
Фамилия менеджера

⇐ первичный ключ

Таблица не находится в 3НФ, так как неключевое поле «Фамилия менеджера» зависит от другого неключевого поля «Код менеджера». Транзитивная зависимость

Код заказа  $\rightarrow$  Код менеджера    Код менеджера  $\rightarrow$  Фамилия менеджера

Для приведения к 3НФ необходимо вынести поле "Фамилия менеджера" в отдельную таблицу.

Заказы		Менеджеры	
Код заказа	$\Leftarrow$ первичный ключ	Код менеджера	
Фамилия покупателя		Фамилия менеджера	
Дата продажи			
Код менеджера			

Любая схема отношений может быть приведена к 3НФ декомпозицией, обладающей свойствами соединения без потерь и сохраняющей зависимости.

Третья нормальная форма исключает избыточность и аномалии включения и удаления. Но 3НФ не предотвращает все возможные аномалии.

*Нормальная форма Бойса-Кодда (НФБК).* Отношения базы данных проектируются таким образом, чтобы исключить в них присутствие частичных или транзитивных зависимостей, поскольку эти зависимости приводят к появлению аномалий обновления, рассмотренных ранее.

Для получения второй и третьей нормальных форм требовалось найти и исключить частичные и транзитивные зависимости от первичного ключа. Однако не рассматривались такие же зависимости от потенциальных ключей отношения, если таковые имеются.

Нормальная форма Бойса-Кодда (НФБК) учитывает функциональные зависимости, в которых участвуют все потенциальные ключи отношения (уникальные комбинации), а не только его первичный ключ.

Для отношения с единственным потенциальным ключом его 3НФ и НФБК являются эквивалентными.

*Отношение находится в НФБК тогда и только тогда, когда каждый его детерминант является потенциальным ключом.*

Если в R для каждой функциональной зависимости  $X \rightarrow A$ , где A не принадлежит X, X включает в себя некоторый ключ, то говорят, что данное отношение находится в нормальной форме Бойса-Кодда.

Для проверки принадлежности отношения к НФБК необходимо найти все его детерминанты и убедиться в том, что они являются потенциальными ключами.

Детерминантом является один атрибут или группа атрибутов, от которой полностью функционально зависит другой атрибут.

НФБК является более строгой версией ЗНФ. Иными словами, *любое отношение, находящееся в НФБК, находится в ЗНФ*. Обратное неверно.

Нарушения требований НФБК происходят редко, поскольку это может случиться только тогда, когда:

- имеются два (или более) составных потенциальных ключа;
- эти потенциальные ключи перекрываются, то есть ими совместно используется, по крайней мере, один общий атрибут.

*Пример.* Описывается ситуация, когда отношение нарушает требования НФБК, и дается метод преобразования этого отношения в НФБК.

Рассмотрим отношение Client\_Interview, которое содержит сведения о собеседованиях клиентов с сотрудниками компании. Для проведения собеседования в тот день, на который назначена встреча с клиентом, в распоряжение сотрудника предоставляется особая комната. Однако в течение рабочего дня эта комната может использоваться несколькими разными сотрудниками. С клиентом проводится только одно собеседование в день, но он может участвовать в нескольких собеседованиях в разные дни.

Это отношение имеет следующий вид:

Client Interview (Client\_No, Interview\_Date, Interview\_Time, Staff\_No, Room\_No)

Client_No	Interview_Date	Interview_Time	Staff_No	Room_No
CR76	13-May-98	10.30	SG5	G101
CR56	13-May-98	12.00	SG5	G101
CR74	13-May-98	12.00	SG37	G102
CR56	1-Jul-98	10.30	SG5	G102

Обсуждаемое отношение имеет три потенциальных ключа:

(Client\_No, Interview\_Date) - клиентом проводится только одно собеседование в день.

(Staff\_No, Interview\_Date, Interview\_Time)

(Room\_No, Interview\_Date, Interview\_Time).

Следовательно, отношение Client Interview обладает тремя составными потенциальными ключами, которые перекрываются, то есть ими совместно используется один общий атрибут Interview\_Date.

В качестве первичного ключа данного отношения выбрана комбинация атрибутов (Client\_No, Interview\_Date).

Оно содержит следующие функциональные зависимости:

фз1 Client\_No, Interview\_Date  $\rightarrow$  Interview\_Time, Staff\_No, Room\_No

*Первичный ключ*

фз2 Staff\_No, Interview\_Date, Interview\_Time  $\rightarrow$  Client\_No

*Потенциальный ключ*

фз3 Room\_No, Interview\_Date, Interview\_Time  $\rightarrow$  Staff\_No, Client\_No

*Потенциальный ключ*

фз4 Staff\_No, Interview\_Date  $\rightarrow$  Room\_No

Рассмотрим эти функциональные зависимости для определения нормальной формы отношения Client Interview.

Поскольку функциональные зависимости фз1, фз2 и фз3 являются потенциальными ключами этого отношения, то они не вызовут никаких проблем.

Функциональная зависимость фз4 Staff\_No, Interview\_Date  $\rightarrow$  Room\_No.

Поскольку в этом отношении нет никаких частичных или транзитивных зависимостей от первичного ключа (Client\_No, Interview\_Date) и допускается наличие функциональной зависимости фз4, можно считать, что отношение Client Interview находится в ЗНФ.

Однако это отношение не находится в НФБК (более строгий вариант ЗНФ), так как в нем присутствует детерминант (Staff\_No, Interview\_Date), который не является потенциальным ключом этого отношения.

В НФБК требуется, чтобы все детерминанты отношения были его потенциальными ключами.

В противном случае отношение Client\_Interview может страдать от аномалий обновления. Например, 13 мая 1998 года (значение «13-May-98») при изменении номера комнаты, выделенной сотруднику 'SG5', потребуется обновить значения в двух строках. Если при этом будет обновлена только одна строка, то база данных будет приведена в противоречивое состояние.

Для преобразования отношения Client\_Interview в форму НФБК необходимо устранить нарушающую ее ограничения функциональную зависимость посредством создания двух новых отношений — Interview и Staff\_Room.

Отношения Interview и Staff\_Room имеют следующий вид (ключевые поля выделены:

Interview (Client\_No, Interview\_Date, Interview\_Time, Staff\_No)

Staff\_Room (Staff\_No, Interview\_Date, Room\_No)

Отношение Interview в НФБК

Client_No	<u>Interview_Date</u>	Interview_Time	Staff_No
CR76	13-May-98	10.30	SG5
CR56	13-May-98	12.00	SG5
CR74	13-May-98	12.00	SG37
CR56	1-Jul-98	10.30	SG5

Отношение Staff\_Room в НФБК

Staff_No	<u>Interview_Date</u>	Room_No
SG5	13-May-98	G101
SG37	13-May-98	G102
SG5	1-Jul-98	G102



Любое отношение, которое не находится в НФБК, можно декомпозировать с образованием НФБК-отношений, однако делать это не всегда желательно.

Например, декомпозиция будет нежелательна, если в результате ее выполнения утрачивается некоторая функциональная зависимость (то есть детерминант и определяемые им атрибуты помещаются в разные отношения). В этой ситуации будет трудно обеспечить исходную функциональную зависимость отношения и важное ограничение может быть утрачено. Если имеет место упомянутая ситуация, то лучше закончить процесс нормализации на этапе образования ЗНФ-отношений, в которых все требуемые зависимости всегда сохраняются.

Обратите внимание на то, что в рассмотренном примере при создании двух новых НФБК-отношений на основе исходного отношения Client\_Interview утрачивается следующая функциональная зависимость: Room\_No, Interview\_Date, Interview\_Time (зависимость фз3), поскольку детерминант этой зависимости больше не будет находиться в том же отношении, что и определяемые им атрибуты (потеряно ограничение, что для проведения собеседования в тот день, на который назначена встреча с клиентом, в распоряжение сотрудника предоставляется особая комната).

Однако, если не устранить функциональную зависимость Staff\_No, Interview\_Date  $\rightarrow$  Room\_No (зависимость фз4), то отношение Client\_Interview будет обладать избыточностью данных.

*Критерии для оценки влияние логического моделирования данных на качество физических моделей данных и производительность БД*

Сравнение нормализованных и ненормализованных моделей.

Критерий	Отношения слабо нормализованы (1НФ, 2НФ)	Отношения сильно нормализованы(3НФ)
Адекватность БД предметной области	ХУЖЕ (-)	ЛУЧШЕ (+)

Легкость разработки и сопровождения БД	СЛОЖНЕЕ (-)	ЛЕГЧЕ (+)
Скорость выполнения вставки, обновления, удаления	МЕДЛЕННЕЕ (-)	БЫСТРЕЕ (+)
Скорость выполнения выборки данных	БЫСТРЕЕ (+)	МЕДЛЕННЕЕ (-)

Если сравнивать нормализованные и ненормализованные модели отношений, то более сильно нормализованные оказываются лучше спроектированными. Они больше адекватны предметной области, для них быстрее выполняются операции модификации БД. У слабо нормализованных отношений преимущество заключается в том, что если к БД обращаться только с запросами на выборку данных, то для них такие запросы выполняются быстрее.

## **25. Оптимизация выполнения запросов на языке SQL. Непроцедурность запросов на языке SQL. Критерий и аргументы оптимизации запросов. Информация в базе данных для процесса оптимизации запросов. Средства отображения оптимальных планов в СУБД.**

Важным классификатором языков программирования является их процедурность или непроцедурность.

В процедурных языках программа явно описывает действия, которые необходимо выполнить, а результат задается только способом получения его при помощи некоторой процедуры (алгоритма), которая представляет собой определенную последовательность действий. Программа, написанная на процедурном языке, представляет собой последовательность команд, определяющих алгоритм решения задачи. Процедурное программирование - есть отражение фон Неймановской архитектуры компьютера. Программа

производит преобразование содержимого памяти, изменяя его от исходного состояния к результирующему.

В непроедурных языках (или декларативных языках) описывается не как получить, а что получить, а как - программа решает самостоятельно. Т.е. не существует однозначного алгоритма решения задачи (обработки данных).

Обработка данных в реляционных СУБД базируется на формальном аппарате реляционной алгебры и реляционного исчисления, предложенном Коддом (1971 г.) в качестве основы для создания реляционных языков.

Реляционная алгебра основывается на теории множеств, реляционное исчисление – на исчислении предикатов первого порядка. Выделяются два вида реляционного исчисления — исчисление кортежей и исчисление доменов.

Реляционная алгебра и реляционное исчисление эквивалентны друг другу, т.е. для каждого выражения алгебры существует эквивалентное выражение в реляционном исчислении (и наоборот).

Использование в реляционной модели данных реляционной алгебры и реляционного исчисления определяется тем, что они различаются уровнем процедурности. Выражения реляционной алгебры строятся на основе алгебраических операций (высокого уровня), и подобно тому, как интерпретируются арифметические и логические выражения, выражение реляционной алгебры также имеет процедурную интерпретацию. Другими словами, запрос, представленный на языке реляционной алгебры, может быть вычислен на основе выполнения элементарных алгебраических операций с учетом их приоритетов и возможного наличия скобок.

Для формулы реляционного исчисления однозначная вычислительная интерпретация, вообще говоря, отсутствует. Формула только ставит условия, которым должны удовлетворять кортежи результирующего отношения. Поэтому языки реляционного исчисления являются в большей степени непроедурными, или декларативными.

Язык называется **реляционно полным**, если он позволяет получить любое отношение, которое можно вывести с помощью реляционного исчисления. Большинство реляционных языков запросов является реляционно полными.

Реляционная алгебра и реляционное исчисление, как правило, не принимаются в качестве полной основы какого-либо языка БД. Обычно (например, в случае языка SQL) язык основывается на некоторой смеси алгебраических и логических конструкций. Кроме того, по сравнению с реляционной алгеброй и реляционным исчислением они обладают и другими, более широкими функциональными возможностями, поскольку в них предусмотрены дополнительные операции, способные выполнять вычислительные, обобщающие и упорядочивающие функции.

Реляционная алгебра и реляционное исчисление представляют собой формальные, а не дружественные пользователю языки. В реляционных базах данных они использовались в качестве основы для разработки других языков управления данными, более высокого уровня. Их можно разделить на две категории: языки на основе преобразований и графические языки.

Языки на основе преобразований являются непроцедурными языками, которые используют отношения для преобразования исходных данных к требуемому виду. Эти языки предоставляют простые в употреблении структуры для формулирования требуемого результата имеющимися средствами. Примером языка на основе преобразований является язык SQL. Запрос формулируется на естественном языке и не задает конкретный алгоритм обработки данных.

Рассмотрим пример запроса на языке SQL и возможные варианты алгоритма его выполнения.

*Запрос:* Выбрать всю информацию по менеджерам, которые работают в московских отделениях компании. В запросе используются таблицы Компания и Персонал.

*SELECT \* FROM Персонал As П, Компания As К*

*WHERE П.КодКомпании = К.КодКомпании AND -- К.КодКомпании -  
первичный ключ, П.КодКомпании –внешний ключ  
(П.Должность = 'Manager' AND К.Город = 'Москва');*

Исходя из правил эквивалентности выражений реляционной алгебры (существуют правила преобразований операций реляционной алгебры), приведенный запрос можно записать в следующих трех вариантах:

1.  $\sigma_{(П.Должность = 'Manager') \text{ AND } (К.Город = 'Москва') \text{ AND } (П.КодКомпании = К.КодКомпании)}$  (Персонал x Компания) - используется декартово произведение таблиц Персонал и Компания с последующей выборкой по заданным условиям.

2.  $\sigma_{(П.Должность = 'Manager') \text{ AND } (К.Город = 'Москва')}$  (Персонал ест. соедин П.КодКомпании = К.КодКомпании Компания) - используется естественное соединение таблиц Персонал и Компания по условию П.КодКомпании = К.КодКомпании с последующей выборкой по заданным условиям.

3.  $(П.Должность = 'Manager' \text{ (Персонал)}) \text{ ест. соедин } П.КодКомпании = К.КодКомпании \text{ (} \sigma_{К.Город = 'Москва'} \text{ (Компания))}$  - естественное соединение применяется к предварительной выборке по условию К.Город = 'Москва' (Компания) с последующей выборкой по заданному условию П.Должность = 'Manager'.

Чтобы произвести количественную оценку, предположим, что таблица Персонал содержит 1000 записей (кортежей), а таблица Компания — 50. Введем допущение, что пять из 50 менеджеров компании управляют отделениями, находящимися в Москве.

Сравнение трех вариантов выполнения одного и того же запроса можно провести с точки зрения количества операций доступа к диску, выполняемых в каждом случае. Количество операций доступа к внешней памяти определяют время выполнения запроса (чем больше операций, тем больше время выполнения запроса). Эффективность вариантов выполнения запроса будем оценивать стоимостью выполнения варианта, которая равна количеству обращения к внешней памяти.

Для упрощения будем считать, что ни одно из используемых отношений не имеет каких-либо индексов или ключей сортировки, а результаты любых промежуточных операций всегда записываются на диск.

Операции записи на диск окончательных результатов выполнения запроса игнорируется, поскольку их количество в каждом из трех рассматриваемых вариантов одинаково.

Дополнительно предположим, что по каждой дисковой операции осуществляется доступ только к одному кортежу (хотя на практике обмен с внешними устройствами осуществляется блоками, содержащими, как правило, несколько кортежей, см. предыдущий раздел), а оперативная память компьютера достаточно велика, чтобы разместить все отношения, используемые при выполнении любой из операций реляционной алгебры.

При обработке первого варианта запроса вычисляется декартово произведение таблиц Персонал и Компания, для чего потребуется  $(1000+50)$  операций доступа к диску при чтении исходных отношений и вывод на диск результирующей таблицы, содержащей  $(1000*50)$  строк. Затем потребуется вновь считать с диска все эти  $(1000*50)$  строки, чтобы проверить каждую из них на соответствие заданному предикату. В результате общая стоимость обработки данного варианта запроса составит:

$$(1000 + 50) + (1000 * 50) + (1000 * 50) = 101\ 050 \text{ дисковых операций}$$

Во втором варианте выполнения запроса сначала осуществляется соединение таблиц Персонал и Компания по полю КодКомпании (Персонал ест.  $\text{соед } \text{П.КодКомпании} = \text{К.КодКомпании}$  Компания), для чего потребуется выполнить  $1000+50$  обращений к диску при считывании кортежей исходных отношений.

Очевидно, что результирующая таблица операции соединения будет насчитывать 1000 кортежей - по одному, для каждого из сотрудников компании (поскольку каждый работник может быть приписан только к одному из отделений).

Последующая операция выборки потребует выполнения 1000 обращений к диску с целью считывания результатов операции соединения и последующей выборкой по заданным условиям

(П.должность = 'Manager') AND(К.Город = 'Москва') и общая стоимость выполнения запроса составит:

$$(1000 + 50) + 1000 + 1000 = 3\ 050 \text{ дисковых операций}$$

В третьем варианте на первом этапе обработки запроса каждый кортеж таблицы Персонал проверяется на соответствие указанному предикату (П.должность = 'Manager' (Персонал)), для чего потребуется выполнить 1000 операций чтения. Результирующий набор данных состоит всего 50 строк (50 компаний). На следующем этапе выполняется операция выборки из таблицы Компания сведений о тех отделениях, которые расположены в Москве. Считывается 50 исходных кортежей и записывается результирующая таблица из 5 строк. На последнем этапе выполняется соединение укороченных отношений Персонал и Компания, для чего потребуется 50+5 операций чтения. Следовательно, общая стоимость обработки запроса составит:

$$1000 + 50 + 50 + 5 + (50 + 5) = 1\ 160 \text{ дисковых операций}$$

Очевидно, что наиболее эффективным является третий вариант записи запроса, его выполнения в 87 раз меньше стоимости первого варианта. Интуитивно понятно, что выполнение операции декартового произведения или соединения таблиц связано с большими затратами, чем выполнение операций выборки. Поэтому основное преимущество третьего варианта обработки запроса состоит в том, что размеры обрабатываемых таблиц были существенно сокращены до выполнения операции соединения.

Из приведенного примера видно, что непроцедурность языка SQL (инструкция SELECT является непроцедурной и не определяет точные шаги, которые сервер базы данных должен предпринять, чтобы извлечь запрошенные данные) порождает различные алгоритмы выполнения запроса, которые могут иметь сильно отличающиеся стоимости выполнения

и, как следствие, будут определять общую эффективность работы с базой данных. Сервер базы данных должен проанализировать инструкцию SELECT для определения самого эффективного способа извлечения запрошенных данных. Выбор варианта выполнения запроса с наименьшей (или оптимальной) стоимостью возлагается на специальные средства СУБД, которые называются оптимизатором выполнения запросов.

Оптимизатор запросов является необходимой компонентой современных СУБД и от качества его работы существенно зависит эффективность работы приложений. На работу оптимизатора запросов по выбору вариантов выполнения с наименьшей стоимостью влияют много параметров базы данных, которыми можно управлять, обеспечивая повышение эффективности работы приложений.

*Процесс выполнения операторов SQL.* Процесс выполнения операторов SQL может быть условно разделен на 5 этапов, которые показаны на рисунке 5.

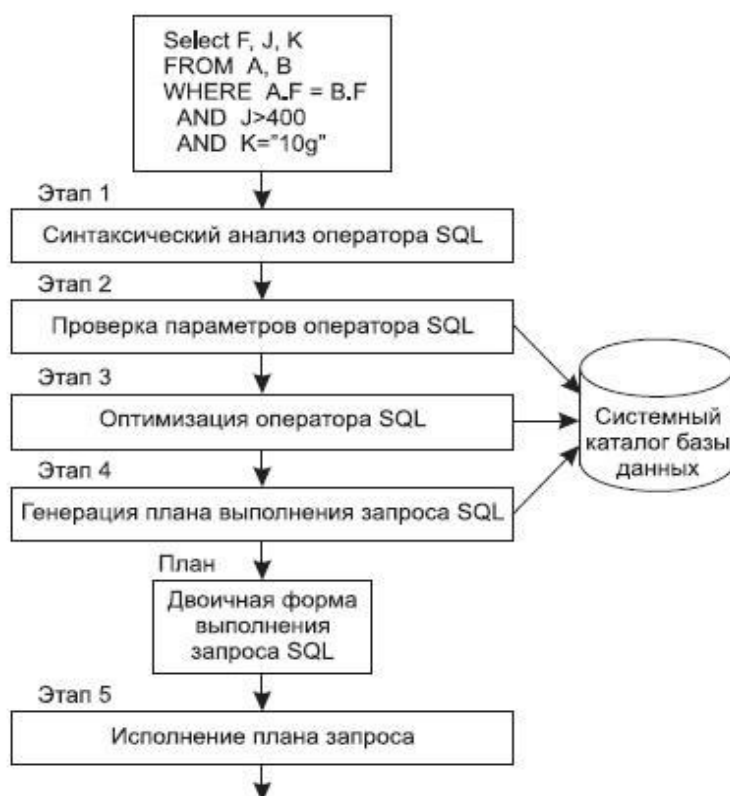


Рисунок 5. Процесс выполнения операторов SQL.

На первом этапе выполняется синтаксический анализ оператора SQL (проверяется корректность записи SQL-оператора в соответствии с



правилами синтаксиса) и производится декомпозиция запроса (запрос преобразуется в дерево).

Назначение декомпозиции состоит в преобразование запроса на языке высокого уровня (язык SQL) в выражение реляционной алгебры, с последующей проверкой его синтаксической и семантической корректности. Обычно декомпозиция включает стадии анализа, нормализации, семантического анализа, упрощения и реструктуризации запроса.

Запрос преобразуется к некоторому внутреннему представлению, более удобному для выполнения последующей обработки – выбирается та или иная форма дерева запроса. Дерево запроса строится следующим образом:

- Узлы листья создаются для каждого используемого в запросе базового отношения.
- Не листовые узлы создаются для каждого промежуточного отношения, создаваемого в результате выполнения некоторой операции реляционной алгебры.
- Корень дерева представляет результирующий набор данных запроса.
- Операции выполняются в направлении от листовых узлов к корню дерева.

Такие деревья называются деревьями реляционной алгебры.

На рисунке 6 показано дерево запроса, построенное для приведенного выше SQL-оператора:

```
SELECT * FROM Персонал As П, Компания As К  
WHERE П.КодКомпании = К.КодКомпании AND (П.Должность = 'Manager'  
AND К.Город = 'Москва').
```



Рисунок 6. Пример дерева реляционной алгебры.

На втором этапе проверяется корректность параметров оператора SQL: имен отношений, имен полей данных, привилегий пользователя по работе с указанными объектами. Здесь обнаруживаются семантические ошибки. Выполняется проверка всех объектов базы данных, на которые в запросе приводятся ссылки. Например, проверяется существование всех столбцов, на которые ссылается запрос, и определяются их идентификаторы. После процесса формируется окончательное дерево запроса.

На третьем этапе проводится оптимизация запроса. В качестве входных данных оптимизатор запросов получает скомпилированное дерево запроса, которое было сгенерировано на предыдущем шаге, и рассматривает стратегии доступа, прежде чем принять решение, как следует обрабатывать данный запрос. СУБД проводит разделение целостного запроса на ряд минимальных операций и оптимизирует последовательность их выполнения.

Для поиска наиболее эффективного плана выполнения запроса оптимизатор запросов вначале выполняет анализ запроса, в процессе которого он отыскивает аргументы поиска и операции соединения. Затем оптимизатор выбирает индексы, которые будут использоваться. Далее, если существуют операции соединения, оптимизатор выбирает порядок соединений и выбирает одну из техник обработки соединений.

На этом этапе строится несколько планов выполнения запроса и выбирается из них один — оптимальный для данного состояния БД.

Некоторые сложные инструкции SELECT имеют тысячи возможных планов выполнения.

Существуют эвристический подход к оптимизации запросов и оценка стоимости операций реляционной алгебры. Эвристические методы оптимизации запросов предусматривают использование правил трансформации для преобразования выражения реляционной алгебры в некоторую эквивалентную форму, обработка которой будет заведомо более эффективной.

Существуют определённые правила трансформации, позволяющие изменять порядок выполнения операций соединения и выборки таким образом, чтобы операции выборки выполнялись в первую очередь. Применение этих правил позволяют получить достаточно хорошие стратегии выполнения запросов, хотя и не всегда самые оптимальные.

Опираясь на правила трансформации (преобразования операций), оптимизатор получает возможность преобразовать некоторое выражение реляционной алгебры в эквивалентное ему выражение, обработка которого будет заведомо более эффективной.

Оценка стоимости операций реляционной алгебры основывается на том, что СУБД может поддерживать множество различных способов реализации операций реляционной алгебры. Назначение всего процесса оптимизации запросов состоит в выборе наиболее эффективного способа реализации операций, из которых состоит обработка запроса. Для этого система использует формулы, позволяющие оценить стоимость, нескольких различных вариантов обработки, после чего выбирает тот из них, стоимость которого минимальна. Как было показано раньше, стоимости реляционных операций определяется как стоимость необходимых операций доступа к внешней памяти.

Существуют различные варианты реализации основных операций реляционной алгебры и для каждого из них можно дать оценку их стоимости.

Существует два альтернативных варианта операций выборки - линейный поиск неотсортированного файла и двоичный поиск упорядоченного файла. Чтобы показать сложность рассматриваемой проблемы оптимизации, отметим только, что наряду с этими двумя способами рассматриваются варианты поиска по равенству первичного ключа, поиска по неравенству первичного ключа, поиска по равенству значения кластерного индекса, поиска по равенству значения некластерного индекса и др.

Стоимость операций зависит от кардинальности обрабатываемых отношений. Оценка кардинальности обрабатываемых отношений основывается на рассмотренной выше статистической информации о таблицах и индексах, хранимой в базе данных.

На четвертом этапе СУБД генерирует двоичную версию оптимального плана запроса, подготовленного на этапе 3. Двоичный план выполнения запроса в СУБД фактически является эквивалентом объектного кода программы.

На пятом этапе СУБД реализует (выполняет) разработанный план, выполняя оператор SQL.

Перечисленные этапы отличаются по числу обращений к БД и по процессорному времени, требуемому для их выполнения.

Синтаксический анализ проводится очень быстро, он не требует обращения к системным каталогам БД.

Семантический анализ уже требует работы с базой метаданных, то есть с системными каталогами БД, поэтому при выполнении этого этапа происходит обращение к системному каталогу и серьезная работа с ним.

Этап, связанный с оптимизацией плана запроса, требует работы не только с системным каталогом, но и со статистической информацией о БД, которая характеризует текущее состояние всех отношений, используемых в запросе, их физическое расположение на страницах и сегментах внешней памяти. Этап оптимизации наиболее трудоемкий и длительный в процессе выполнения запроса.

Однако если не проводить этап оптимизации, то стоимость (время) выполнения неоптимизированного запроса может в несколько раз превысить стоимость оптимизированного запроса. Время, потраченное на оптимизацию запроса, компенсирует затраты на выполнение неоптимизированного запроса.

Существует два варианта выполнения первых трех этапов процесса оптимизации запросов - динамическая и статическая оптимизация запросов.

В первом случае декомпозиция и оптимизация выполняется динамически, при каждом вызове запроса. Преимущество динамической оптимизации запросов состоит в том, что вся используемая в процессе выбора оптимальной стратегии информация является актуальной именно на текущий момент. Основной недостаток динамической оптимизации состоит, в относительном снижении эффективности обработки запросов, поскольку оптимизация выполняется при обработке каждого обращения к запросу. Более того, диапазон анализируемых стратегий может быть сокращен из-за необходимости обеспечения приемлемой производительности обработки, что может привести к выбору не совсем оптимальной стратегии.

Альтернативным вариантом является статическая оптимизация запросов, предусматривающая однократное выполнение оптимизации. Основное преимущество статической оптимизации состоит в устранении дополнительной нагрузки из-за выполнения процедур оптимизации в процессе выполнения запроса. В результате появляется возможность более тщательного подбора оптимальной стратегии за счет большего числа возможных вариантов. В случае многократно выполняемых запросов, дополнительные затраты времени на поиск самых оптимальных решений могут дать существенный выигрыш в достигаемой производительности. Основной недостаток этого метода состоит в том, что выбранная стратегия выполнения запроса, которая была оптимальной при компиляции запроса, в момент выполнения может оказаться уже не самой оптимальной. План может стать недействительным при изменении таблицы или представления,

на которые ссылаются запрос, изменения или удаления индексов, используемых планом выполнения, обновления статистик и др.

Для преодоления этого недостатка может использоваться комбинированный подход, предусматривающий повторную оптимизацию запроса в том случае, когда система обнаруживает, что статистические показатели базы данных существенно изменились с момента выполнения оптимизации данного запроса. Альтернативный вариант предусматривает автоматическую компиляцию запроса при первом его выполнении в каждом из сеансов, с последующим кэшированием оптимизированного плана запроса в течение всего времени существования данного сеанса. Это позволяет однократно выполнять оптимизации каждого из запросов, используемых в отдельном сеансе работы с базой данных. Рассмотрим, как этот механизм реализуется в MS SQL Server.

#### *Хранение и использование планов выполнения MS SQL Server.*

Скомпилированные планы запроса сохраняются в части памяти MS SQL Server, называемой кэшем плана. В MS SQL Server оптимизированные планы выполнения состоят из следующих основных компонентов - план запроса и контекст выполнения.

Тело плана выполнения является реентерабельной<sup>8</sup> структурой данных только для чтения, которая предназначена для использования любым числом пользователей. Она называется планом запроса. План запроса не содержит контекста пользователя. В памяти не может находиться более одной или двух копий плана запроса: одна — для всех последовательных

---

<sup>8</sup> Компьютерная программа в целом или её отдельная процедура называется реентерабельной, если она разработана таким образом, что одна и та же копия инструкций программы в памяти может быть совместно использована несколькими пользователями или процессами. При этом второй пользователь может вызвать реентерабельный код до того, как с ним завершит работу первый пользователь и это как минимум не должно привести к ошибке, а в лучшем случае не должно вызвать потери вычислений (то есть не должно появиться необходимости выполнять уже выполненные фрагменты кода).

выполнений, а другая — для всех параллельных выполнений. Одна параллельная копия обслуживает все параллельные выполнения независимо от степени параллелизма.

*Контекст выполнения.* Каждый пользователь, который в настоящий момент выполняет запрос, имеет структуру данных, которая содержит данные, относящиеся к данному выполнению, например, значения параметров. Эта структура данных называется контекстом выполнения (рисунок 7). Структуры данных контекста выполнения являются повторно используемыми. Если пользователь выполняет запрос и одна из структур не используется, она повторно инициализируется контекстом нового пользователя.



Рисунок 7. Пример контекста выполнения.

В MS SQL Server при выполнении инструкции SQL реляционный механизм сначала просматривает кэш процедур, проверяя, нет ли в нем плана выполнения для такой же инструкции. MS SQL Server повторно использует любой план выполнения, который найдет, снижая издержки на его повторную компиляцию. Если не найдено ни одного существующего плана, MS SQL Server формирует для этого запроса новый план.

MS SQL Server реализует эффективный алгоритм поиска существующих планов выполнения для любой инструкции SQL. В большинстве систем ресурсы, затрачиваемые на поиск готового плана, всегда меньше ресурсов, затрачиваемых на повторную компиляцию каждой инструкции SQL.

Алгоритмы поиска соответствия инструкции SQL существующему неиспользуемому плану выполнения в кэше требуют, чтобы все ссылки на объекты были полностью квалифицированы. Например, для инструкции `SELECT (SELECT * FROM Contact)` соответствие существующему плану не будет найдено, а для инструкции `SELECT (SELECT * FROM Person.Contact)`- будет.

После формирования плана выполнения он хранится в кэше процедур. MS SQL Server при необходимости освободить память удаляет из кэша старые, неиспользуемые процедуры. Каждый план запроса и контекст выполнения имеет соответствующий коэффициент стоимости, который указывает, насколько затратной является компиляция данной структуры. В ней присутствует также и поле возраста.

Каждый раз, когда соединение ссылается на объект, поле возраста увеличивается на коэффициент стоимости. Например, если план запроса имеет коэффициент стоимости 8 и на него ссылались дважды, его возраст становится равным 16. Процесс отложенной записи периодически просматривает список объектов, находящихся в кэше процедур. При каждом просмотре поле возраста уменьшается на единицу. В этом примере возраст плана запроса в кэше будет уменьшен до нуля за 16 проходов, если ни один из пользователей им не воспользуется.

Процесс отложенной записи освобождает объект в следующих условиях:

- Диспетчеру памяти требуется память, а вся доступная память уже занята.
- Поле возраста объекта равно 0.
- На объект в настоящий момент нет ссылок из соединений.

Поскольку поле возраста увеличивается каждый раз при ссылке на объект, для часто используемых объектов оно не успевает достигнуть значения 0, и поэтому они не удаляются из кэша. Редко используемые объекты вскоре будут доступны для освобождения, но фактически будут удалены только тогда, когда занимаемая ими память понадобится под другие объекты.



*Процесс оптимизации запроса.* Как было показано выше, оператор SELECT является непроцедурным, он не определяет точные шаги, которые сервер базы данных должен предпринять, чтобы извлечь запрошенные данные.

Инструкция SELECT определяет только следующее.

- Формат результирующего набора. Он указан, главным образом, в списке выбора. Однако другие предложения, например ORDER BY и GROUP BY, также затрагивают конечную форму результирующего набора.

- Таблицы, которые содержат исходные данные. Указываются в предложении FROM.

- Как таблицы логически связаны с целями инструкции SELECT. Это определяется в спецификациях соединения, которые могут появляться в предложении WHERE или в предложении ON, следующем за предложением FROM.

- Условия, которым строки в исходных таблицах должны соответствовать для выбора их инструкцией SELECT. Указываются в предложениях WHERE и HAVING.

План выполнения запроса представляет собой определение использования аргументов поиска, способов выбора информации из таблиц, порядка операций соединения, выбор техники (техник) для обработки операций соединения. Рассмотрим эти вопросы более подробно.

Приводимые рассуждения будут относиться к оптимизации запросов на основе оценка стоимости (выше были классифицированы два подхода к оптимизации: эвристический подход и оценка стоимости операций реляционной алгебры).

Оптимизатор запроса должен проанализировать возможные планы и выбрать один файл с самой низкой предполагаемой стоимостью. Некоторые сложные инструкции SELECT имеют тысячи возможных планов выполнения. В этих случаях оптимизатор запроса не анализирует все возможные комбинации. Вместо этого он использует сложные алгоритмы поиска плана выполнения, имеющего стоимость, близкую к минимальной

возможной стоимости. Оптимизатор запросов не выбирает только план выполнения с самой низкой стоимостью ресурсов; он выбирает такой план, который возвращает результаты пользователю при разумной стоимости ресурсов, но который возвращает результаты быстрее других. Например, параллельная обработка запроса обычно использует больше ресурсов, чем его последовательная обработка, но завершает выполнение запроса быстрее. Оптимизатор будет использовать параллельный план выполнения для возврата результатов, если это не окажет неблагоприятного влияния на загрузку сервера.

В начале процесса оптимизатор проверяет запрос на аргументы поиска, использование оператора OR, существование критериев соединения. Именно в этом порядке.

Аргумент поиска является частью запроса, которая ограничивает промежуточный результирующий набор запроса. Основным назначением аргументов поиска является то, что они позволяют использовать существующие индексы применительно к конкретному выражению.

Примеры аргументов поиска:

- emp\_fname = 'Иванов';
- salary >= 50000;
- emp\_fname = 'Иванов' AND salary >= 50000.

Существует несколько форм выражений, которые не могут быть использованы оптимизатором в качестве аргументов поиска.

К первой группе принадлежат выражения с оператором отрицания (NOT). Так же, если используется выражение в левой части оператора, существующее выражение не может быть использовано в качестве аргумента поиска.

Примеры выражений, которые не являются аргументами поиска:

- NOT IN ('d1', 'd2');
- emp\_no <> O. 9031;
- budget \* 0.59 > 55000.

Основным недостатком выражений, которые не могут быть использованы в качестве аргументов поиска, является то, что оптимизатор не может использовать существующие индексы в отношении выражения, чтобы повысить производительность соответствующего запроса. Иными словами, в этом случае единственным способом доступа к данным может быть только последовательное сканирование таблицы.

Использование оператора OR. Если один из термов в условии выборки содержит условие OR и для него не существует подходящего индекса или упорядочивания, то вся операция выборки выполняется без их использования.

Идентификация аргументов поиска позволяет оптимизатору принять решение о том, можно ли использовать один или более существующих индексов и определить способы выбора информации из таблиц данных.

На этой фазе оптимизатор проверяет каждый аргумент поиска на предмет, существуют ли подходящие индексы для соответствующего выражения. Если такой индекс существует, оптимизатор принимает решение, использовать его или нет.

Это решение зависит от селективности соответствующего выражения. Селективность выражения определяется как отношение количества строк, удовлетворяющих условию выбора, к общему количеству строк в таблице. Например, выбор по условию 10 строк из общего количества 1000 строк в таблице означает высокую селективность. Наоборот, выбор по условию из той же таблицы 900 строк означает низкую селективность.

Оптимизатор проверяет селективность выражения с индексированным столбцом, используя статистические данные, которые создаются для распределения значений в столбце (см. раздел Статистическая информация в MS SQL Server). Оптимизатор запросов использует эту информацию для определения оптимального плана запроса, оценивая стоимость использования индекса для выполнения запроса. В разделе Статистическая информация в MS SQL Server было сказано, что гистограмма отдельного

столбца — это набор значений данного поля, ограниченный до 200 значений.

Рассмотрим пример определения селективности для гистограммы на 10 интервалов для поля таблицы AGE (возраст сотрудника) таблицы Сотрудник (EMP). Пусть всего в организации работают сотрудники в возрасте от 10 до 60 лет. Тогда гистограмма, изображающая распределение значений поля AGE может иметь вид, показанный ниже на рисунке 8. Гистограмма построена исходя из разбиения диапазона значений поля AGE на 10 интервалов.

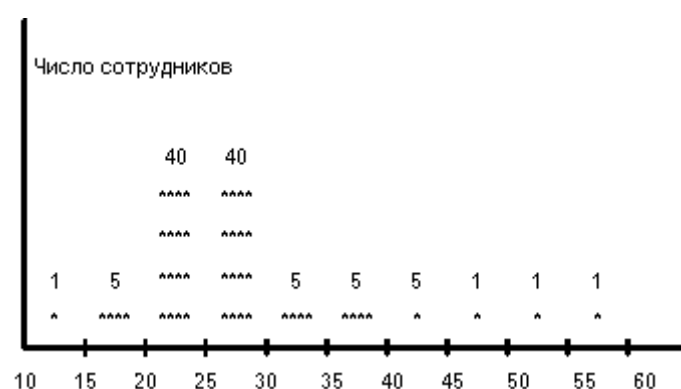


Рисунок 8. Пример гистограммы.

Селективность определяется следующим образом. Пусть в интервал значений AGE  $S_i$  попадает  $K_i$  значений. Тогда  $SEL (EMP.AGE = const)$ , если значение константы попадает в интервал значений  $S_i$ , можно оценить следующим образом:  $0 \leq SEL (EMP.AGE) \leq K_i/T$  ( $T$  - общее число кортежей в отношении EMP). Отсюда средняя оценка степени селективности предиката -  $K_i / (2 \cdot T)$ . Например,  $SEL (AGE = 29)$  оценивается в 0.2, а  $SEL (AGE = 16)$  оценивается в 0.025.

При обращении к таблице, у которой нет индексов, система выполняет последовательное сканирование таблицы. Оптимизатор использует индексы для повышения скорости выполнения запроса. При обращении к таблице, у которой есть индексы, оптимизатор может принять решение использовать или не использовать существующий индекс. Это зависит от селективности выражения с индексированным столбцом.

Если таблица без кластеризованного индекса, то система вначале исследует структуру некластеризованного индекса, а затем отыскивает строку, используя идентификатор этой строки. Т.е. обращение к страницам индекса, а затем обращение к нужной странице данных. Использование некластеризованного индекса требует больше операций ввода/вывода, чем при использовании соответствующего кластеризованного индекса. При использовании некластеризованного индекса нужно читать страницы данных после просмотра В-дерева некластеризованного индекса.

При большом количестве обращений к страницам данных для запроса при низкой селективности, обращение к страницам индекса является дополнительной затратой времени, т.к. нужно обратиться к большинству страниц данных. Если таблица имеет кластеризованный индекс, то проход по кластеризованной индексной структуре выполняется вслед за проходом по индексной структуре некластеризованного индекса таблицы.

Более эффективно последовательно сканировать страницы таблицы данных для поиска строк, которые принадлежат результирующему набору данных (оптимизатор принимает решение не использовать существующий индекс).

При использовании кластеризованного индекса для поиска данных система начинает поиск с корневого узла соответствующего В-дерева и после небольшого количества операций чтения (равно количеству уровней дерева) достигает узла листа, где хранятся данные. Когда система сканирует кластеризованный индекс, нет необходимости покидать структуру В-дерева для сканирования страниц данных, потому что такие страницы уже присутствуют на уровне листьев этого дерева.

По этой причине проход по индексной структуре кластеризованного индекса почти всегда выполняется значительно быстрее, чем проход по индексной структуре соответствующего некластеризованного индекса.

Можно сделать общий вывод, что ответ на вопрос, какой метод доступа (сканирование индекса или сканирование таблицы) является

наиболее быстрым, не является простым и зависит от селективности и типа индекса.

В общем случае, если необходимы только несколько строк с определенными ключевыми значениями, то сервер базы данных может использовать индекс. Если необходимы все строки в таблице, то сервер базы данных может проигнорировать индексы и выполнить просмотр таблицы. Если необходимы все строки в таблице, но есть индекс, ключевые столбцы которого находятся в ORDER BY запроса, то выполнение просмотра индекса вместо просмотра таблицы позволит избежать отдельной сортировки результирующего набора и может оказаться более эффективным. Если данных в таблице мало (таблица является очень маленькой), то просмотры таблицы могут быть самым эффективным методом для практически всех обращений к таблице.

Что касается выбора порядка соединения, то обычно порядок, в котором две или более соединяемые таблицы задаются в предложении FROM оператора SELECT, не оказывает влияния на решение, принимаемое оптимизатором относительно порядка их обработки. Множество различных факторов влияют на решение оптимизатора, касающееся того, к какой таблице в первую очередь будет выполняться обращение. Как правило, существует много последовательностей, в которых сервер базы данных может обращаться к базовым таблицам для построения результирующего набора.

Например, если инструкция SELECT ссылается на три таблицы, сервер базы данных сначала может обратиться к TableA, использовать данные из TableA для извлечения соответствующих строк из TableB и затем использовать данные из TableB для извлечения данных из TableC.

Другие последовательности, в которых сервер базы данных может обращаться к таблицам: TableC, TableB, TableA или TableB, TableA, TableC или TableB, TableC, TableA или TableC, TableA, TableB. Т.е. для трех таблиц, участвующих в соединении может быть четыре варианта

последовательности. Для более сложных запросов это количество возрастает.

Операция соединения является операцией, больше всего влияющей на время обработки запроса. Существуют разные способы реализации логической операции соединения или техники обработки соединения. Существуют три различные техники обработки соединения, которые могут быть выбраны оптимизатором в зависимости от статистических данных в каждой из таблиц, участвующих в соединении:

- вложенные циклы;
- слияние соединения;
- хеширование соединения.

Самым простым алгоритмом выполнения операций соединения является метод вложенных циклов, при котором соединение двух отношений выполняется по одному кортежу за цикл. При использовании вложенных циклов алгоритм соединения получает на вход  $n$  таблиц и условия соединения. Результатом его работы является набор строк с результатами соединения. Алгоритм состоит из произвольного числа вложенных итераций поиска данных в каждой из соединяемых таблиц.

Для двух таблиц алгоритм можно описать следующим образом: для каждой строки одной из таблиц (ведущей) выполняется поиск в другой таблице (ведомой) строк соответствующих условию соединения.

Во внешнем цикле выполняется последовательный перебор кортежей первого отношения  $R$  (ведущей таблицы), а во внутреннем цикле последовательно просматриваются все кортежи второго отношения  $S$  (ведомой таблицы). При считывании блока данных (см. раздел Базовые понятия хранения и обработки данных) используются два дополнительных цикла, предназначенных для обработки считанных с внешней памяти блоков (в MS SQL Server это страницы данных).

Алгоритм возвращается результирующую таблицу соединения отношений R и S и имеет следующий вид (блоки в обоих файлах последовательно пронумерованы, начиная с 1)<sup>9</sup>:

```
for iblock = 1 to nblocks(R) { // Внешний цикл
Rblock = readblock(R, iblock);
for jblock = T to nblocks(S) { // Внутренний цикл
Sblock = readblock(S, jblock)
for i = 1 to ntuples(Rblock) {
for j = 1 to ntuples (Sblock) {
if (Rblock.tuple[i]/Sblock.tuple[j]    <отвечает условию соединения>
then <помещение его в результирующую таблицу>;
}}}}
```

Поскольку каждый блок отношения R должен быть считан хотя бы один раз, а каждый блок отношения S должен быть считан для каждого блока отношения R, то оценка стоимости этого метода может быть выполнена по формуле:  $\text{nblocks}(R) + (\text{nblocks}(R) * \text{nblocks}(S))$ .

Из формулы видно, что значение второго терма этой формулы постоянно, тогда как значение первого терма зависит от того, какое из отношений выбрано для считывания во внешнем цикле. Очевидно, что во внешнем цикле должно считываться то из двух соединяемых отношений, которое занимает меньшее количество блоков во внешней памяти.

Внешним циклом выполняется поиск строк в ведущей таблице. Если часть или все ограничения для ведущей таблицы могут быть использованы для поиска по индексу, то на каждой итерации цикла в индексе ищутся расположения всех необходимых строк и выполняется прямой доступ к таблице. В противном случае таблица сканируется целиком. Оставшиеся ограничения используются для фильтрации выбранных строк. Для каждой оставшейся строки вызывается внутренний цикл.

---

<sup>9</sup> Tuples - кортеж



Внутренний цикл по условиям соединения и данным внешнего цикла ищет строки в ведомой таблице. Если часть или все ограничения для ведомой таблицы вместе с ограничениями, полученными от внешнего цикла, могут быть использованы для поиска по индексу, то на каждой итерации цикла в индексе ищутся расположения всех необходимых строк и выполняется прямой доступ к ведомой таблице. В противном случае таблица сканируется целиком. Оставшиеся ограничения используются для фильтрации выбранных строк.

На каждой итерации самого глубокого вложенного цикла выбранные из таблиц строки конкатенируются для получения одной строки итогового результата.

В общем случае циклы могут вкладываться произвольное число раз, в зависимости от числа таблиц, участвующих в соединении.

Преимущества этого алгоритма соединения является то, что это самый общий и поэтому незаменимый алгоритм соединения. При помощи соединения вложенными циклами можно реализовать любое условие соединения. Остальные алгоритмы имеют ограничения по реализуемым с их помощью условиям соединения, например, когда условие выражается неравенством.

Это также самый быстрый алгоритм, если необходимо получить только первую строку результата (например, в SQL выражениях типа EXISTS). Это следует из того, что если соединяемые атрибуты в соединении по эквивалентности образуют ключ внутреннего отношения, то внутренний цикл может завершаться, как только будет найдено первое соответствие.

Кроме того, при использовании поиска по индексу алгоритм лучше всех масштабируется. То есть при увеличении объёма данных в соединяемых таблицах время выполнения запроса увеличивается практически линейно при тех же аппаратных ресурсах.

Метод вложенных циклов является очень медленным, если не существует индекса для одного из соединяемых столбцов. По этой причине

оптимизатор запросов обычно выбирает этот метод в том случае, когда соединяемый столбец во внутренней таблице индексирован, и для внутренней таблицы не нужно выполнять сканирование для каждой строки внешней таблицы.

Алгоритм слияние соединения для соединения таблиц по эквивалентности является наиболее эффективным способ соединения. Это достигается в том случае, когда оба отношения отсортированы по соединяемым атрибутам. В этом случае можно найти подходящие кортежи отношений R и S посредством слияния двух отношений. Если исходные отношения не упорядочены, можно предварительно выполнить их сортировку. Поскольку кортежи отношений располагаются в порядке сортировки, те из них, которые содержат одинаковые значения соединяемого атрибута, гарантированно будут размещаться рядом друг с другом. Если предположить, что соединение выполняется для связи типа "многие ко многим", т.е. в обоих отношениях может существовать несколько кортежей с одним и тем же соединяемым значением, а также принять, что любой набор кортежей с одним и тем же соединяемым значением может быть полностью помещен в буфер базы данных, то каждый блок каждого из соединяемых отношений достаточно будет считать только один раз. Алгоритм получает на вход 2 таблицы и условие соединения. Результатом его работы является таблица с результатами соединения.

Алгоритм имеет следующий вид (процедуры чтения данных не приводятся):

// Соединяются отношения R и S по атрибуту A

// Отношения R и S предварительно сортируются (этого не требуется, если файлы уже отсортированы по атрибуту, используемому для соединения).

sort(R);

sort(S);

// Выполнение слияния

```

nextR = 1; nextS = 1;
while (nextR <= ntuples(R) and nextS <= ntuples(S) )
{
join_value = R.tuples[nextR].A; // Сканирование отношения S до нахождения
значения, меньшего чем текущее, используемое для соединения
while (S.tuples[nextS].A < join_value and nextS <= ntuples(S))
{ nextS = nextS + 1;
}
// Возможно, найдены соответствующие кортежи отношений R и S.
// Каждый кортеж в отношении S с некоторым значением join_value
    ставится в соответствие каждому кортежу в отношении R с этим же
значением join_value. (Предполагается соединение типа M:N).
while (S.tuples[nextS].A = join_value and nextS <= ntuples (S))
{ m = nextR;
while (R.tuples[m].A = join_value and m <= ntuples(R))
{ <вывод соответствующих кортежей S.tuples[nextS] и R.tuples[m] в таблицу
результата>
m = m + 1;
}
nextS = nextS + 1;
}
// Найдены все кортежи отношений R и S с одним и тем же значением
join_value.
// Выбирается из отношения R кортеж с другим значением join_value.
while {R.tuples[nextR].A = join_value and nextR <= ntuples(R))
{ nextR = nextR + 1; }
}

```

Соединение осуществляется за одно сканирование (проход по каждой из входных таблиц). То есть одна и та же строка считывается только один раз, что даёт преимущество перед соединением вложенными циклами.

Если объединяемые таблицы имеют индексы по сравниваемым полям, или одна или обе таблицы достаточно малы, чтобы быть отсортированными в памяти, то объединение с помощью слияния может быть выполнено достаточно эффективно. Оба отсортированных набора данных сканируются и в них ищутся одинаковые значения. За счёт сортировки слияние эффективнее вложенных циклов на больших объёмах данных, но план выполнения не может начинаться со слияния.

Соединение слиянием может использоваться для соединений с условиями отличными от равенства. Однако допустимы не любые условия соединения.

Главным недостатком алгоритма является необходимость в предварительной сортировке списков. Накладные расходы на сортировку могут быть неприемлемо высокими.

При реализации в СУБД, соединение слиянием требует больше памяти и менее гибко, чем алгоритм соединения вложенными циклами. Во многих СУБД соединение слиянием по умолчанию не используется оптимизатором запросов и должно быть включено явно.

Техника хеширования соединения<sup>10</sup> обычно используется, когда не существует никаких индексов для соединяемых столбцов. В случае техники хеширования соединения обе таблицы, которые должны быть соединены, рассматриваются как два потока ввода: компокуемый ввод и контрольный ввод (наименьшая таблица обычно представляет компокуемый ввод).

Этот процесс работает следующим образом:

1. Значение соединяемого столбца строки из компокуемого ввода сохраняется в хешированном сегменте памяти в зависимости от значения, полученного от алгоритма хеширования.

---

<sup>10</sup> Хэш-таблицей называется структура данных, позволяющая хранить пары вида (ключ, хэш-код) и поддерживающая операции поиска, вставки и удаления элемента. Задачей хэш-таблиц является ускорение поиска, например, при записи текстовых полей в базе данных может рассчитываться их хэш код и данные могут помещаться в раздел, соответствующий этому хэш-коду. Тогда при поиске данных надо будет сначала вычислить хэш-код текста и сразу станет известно, в каком разделе их надо искать, то есть, искать надо будет не по всей базе, а только по одному её разделу (это сильно ускоряет поиск).

2. Как только все строки из компонуемого ввода будут обработаны, начинается обработка строк из контрольного ввода.

3. Каждое значение соединяемого столбца строки из контрольного ввода обрабатывается с использованием того же алгоритма хеширования.

4. Отыскиваются соответствующие строки из хешированного сегмента памяти и используются для создания результирующего набора.

Техника хеширования соединения не требует никакого индекса. Поэтому данный метод хорошо применим в первую очередь для тех запросов, где не предполагается наличие индексов. При этом если оптимизатор использует данную технику, то это может служить подсказкой, что нужно создать дополнительные индексы для одного или более соединяемых столбцов.

Соединение хешированием не может применяться, если в ON есть условия, отличающиеся от равенства (например, больше, меньше и т.п.).

*Управление оптимизацией запросов.* В большинстве случаев оптимизатор запросов выбирает самый быстрый план выполнения. Однако существует несколько особых ситуаций, при которых оптимизатор, по ряду различных причин, не может найти оптимальное решение. В подобных случаях нужно применять “подсказки” (hint) оптимизатора, чтобы заставить его использовать конкретный план выполнения, который должен работать лучше.

Подсказки оптимизатора являются необязательной частью в операторе SELECT, которые указывают оптимизатору запросов, что нужно применить один из специфических вариантов поведения. При использовании подсказок оптимизатор не может использовать варианты, а следует указанным инструкциям.

Две причины, почему оптимизатор иногда выбирает не самый быстрый план выполнения:

- оптимизатор запросов не является безупречным;
- система не предоставляет оптимизатору подходящую информацию.

Подсказки оптимизатора могут помочь только в том случае, если выбранный оптимизатором план не является оптимальным. Если система не предоставляет оптимизатору подходящую для его работы информацию, нужно создавать и обновлять статистику базы данных (в MS SQL Server для этого используются опции базы данных `AUTO_CREATE_STATISTICS` и `AUTO_UPDATE_STATISTICS` для создания или изменения существующих статистических данных).

Использовать подсказки оптимизатора нужно только временно и только для тестирования. Их следует исключить из постоянной части любого запроса. Язык SQL и его реализации в MS SQL Server поддерживает различные “подсказки” для оптимизации, которые могут быть объединены в следующие группы:

- подсказки таблицы;
- подсказки соединения;
- подсказки запроса;
- структуры планов.

Подсказки таблицы можно применить к одной таблице. Поддерживаются следующие подсказки таблицы:

- `index`;
- `noexpand`;
- `forceseek` (в SQL начиная с версии Server 2008).

Подсказка `index` служит, для указания одного или более индексов, которые затем будут использованы в запросе. Эта подсказка задается в предложении `FROM` в запросе. Можно использовать эту подсказку, чтобы осуществить доступ к индексу, если оптимизатор по различным причинам выбрал последовательное сканирование таблицы для данного запроса. Подсказка `index` также может быть полезной для того, чтобы не позволить оптимизатору использовать другой конкретный индекс.

Например в запросе

```
SELECT * FROM new_addresses a WITH ( INDEX(i_stateprov))  
WHERE a.StateProvinceID = 9
```

явно указывается оптимизатору запросов использовать индекс i\_stateprov. Без этой подсказки оптимизатор мог, например, выбирать сканирование таблицы.

Другая форма подсказки для запроса INDEX, INDEX (0), требует, чтобы оптимизатор не использовал никаких существующих индексов. Например в запросе

```
SELECT * FROM Person.Address AS a  
WITH(INDEX(0))  
WHERE a.StateProvinceID = 32
```

указывается, что если кластеризованный индекс не существует, то INDEX (0) требует сканирования таблицы. При существовании кластеризованного индекса INDEX (0) определяет сканирование кластеризованного индекса.

Подсказка NOEXPAND указывает, что не будет раскрываться никакое индексированное представление для доступа к лежащим в его основе таблицам в то время, когда оптимизатор запросов обрабатывает запрос. Оптимизатор запросов трактует представление как таблицу без кластеризованного индекса.

Подсказка таблицы FORCESEEK заставляет оптимизатор запросов выполнять только операцию поиска по индексу в качестве пути доступа к данным в таблице (или представлении), на которую ссылается запрос.

Можно использовать эту подсказку таблицы для того, чтобы перекрыть план по умолчанию, выбранный оптимизатором запросов, для устранения проблем производительности, появившихся в результате неэффективного плана выполнения.

Например, если план содержит операторы сканирования таблицы, а соответствующая таблица требует большого количества чтений в процессе выполнения запроса, то задание операции выборки по индексу может дать лучшую производительность запроса.

Подсказки соединения дают указания оптимизатору запросов, как должны выполняться операции соединения в запросе. Они заставляют оптимизатор либо соединять таблицы в том порядке, в каком они указаны в предложении FROM оператора SELECT, либо использовать техники выполнения соединения, явно указанные в операторе.

Поддерживается несколько подсказок соединения:

- FORCE ORDER;
- LOOP;
- HASH;
- MERGE.

Подсказка FORCE ORDER заставляет оптимизатор выполнять соединение в том порядке, в котором они указаны в запросе.

При использовании этой подсказки соединения OPTION(FORCE ORDER), оптимизатор выполняет операцию соединения в том порядке, в котором таблицы появляются в запросе.

Например, в запросе

```
SELECT e.EmployeeID, e.LoginID, d.DepartmentID  
FROM HumanResources.Employee As e, HumanResources.Department As d,  
HumanResources.EmployeeDepartmentHistory As h  
WHERE d.DepartmentID = h.DepartmentID  
AND h.EmployeeID = e.EmployeeID  
AND h.EndDate IS NOT NULL  
OPTION(FORCE ORDER);
```

Соединение будут выполняться в порядке описания таблиц в предложении FROM - HumanResources.Employee, HumanResources.Department, HumanResources.EmployeeDepartmentHistory. Без этой подсказки оптимизатор может выбрать другую последовательность обработки.

Подсказки запроса LOOP, HASH и MERGE заставляют оптимизатор использовать метод соединения вложенный цикл, технику хеширования



соединений и технику слияния соединений соответственно. Эти три подсказки соединения могут применяться только в том случае, когда операция соединения соответствует стандарту SQL, т. е. когда соединение явно указано при помощи ключевого слова JOIN в предложении FROM оператора SELECT.

Например, в запросе используется подсказка OPTION (MERGE JOIN) использовать технику слияния соединений

```
SELECT * FROM Person.Address a JOIN Person.StateProvince s  
ON a.StateProvinceID = s.StateProvinceID  
OPTION (MERGE JOIN);
```

и оптимизатор выберет именно эту технику обработки соединения.

Выбранная подсказка соединения может быть записана в предложении FROM запроса или при использовании предложения option в конце этого запроса. Если нужно применять несколько различных подсказок вместе, то рекомендуется предложение OPTION.

Существует несколько подсказок запроса, которые используются для различных целей. Рассматриваются только две подсказки запроса:

- FAST  $n$  ;
- OPTIMIZE FOR.

Список других подсказок запроса можно найти в документации.

Подсказка FAST  $n$  указывает, что запрос оптимизируется для быстрого поиска первых  $n$  строк. После того, как первые  $n$  строк будут получены, запрос продолжит свое выполнение и создаст полный набор результата.

Эта подсказка может быть очень полезной, если имеется очень сложный запрос с множеством результирующих строк, требующих большого количества времени для их обработки. Обычно запрос выполняется полностью, а затем система отображает его результат. Эта подсказка запроса требует, чтобы система отображала первые  $n$  строк сразу после их обработки.

Подсказка `OPTIMIZE FOR` заставляет оптимизатор запросов использовать конкретное значение для локальной переменной, когда запрос компилируется и оптимизируется. Это значение используется только в процессе оптимизации запроса, но не в процессе его выполнения. Эта подсказка запроса может быть использована, когда задаются структуры планов (см. далее).

Например, в запросе показано использование подсказки запроса `OPTIMIZE FOR`.

```
DECLARE @city_name nvarchar(30)
SET @city_name = 'Newark'
SELECT * FROM Person.Address
WHERE City = @city_name
OPTION ( OPTIMIZE FOR (@city_name = 'Seattle') );
```

Хотя значение переменной `@city_name` установлено в `Newark`, подсказка `OPTIMIZE FOR` заставляет оптимизатор использовать значение `Seattle` для этой переменной при оптимизации запроса.

Подсказки явно задаются в операторе `SELECT` для оказания влияния на работу оптимизатора запросов. Однако иногда нельзя изменять текст в операторе `SELECT` напрямую.

В этом случае можно повлиять на выполнение запросов, используя структуры планов. Другими словами, структуры планов дают возможность использовать отдельную подсказку оптимизации каждый раз, когда выполняются ее условия, без изменения синтаксиса оператора.

В MS SQL Server структуры планов создаются при помощи системной процедуры `sp_create_plan_guide`. Эта процедура создает структуру плана для ассоциированных подсказок запроса или для фактических планов запросов для запросов к базе данных. Другая системная процедура, `sp_control_plan_guide`, включает, отключает или удаляет существующую структуру плана.

MS SQL Server поддерживает следующие типы структур планов:

- **SQL.** Соответствует запросам, которые выполняются в контексте одиночных операторов Transact-SQL или пакетам, которые не являются частью объекта базы данных;

- **OBJECT.** Соответствует запросам, которые выполняются в контексте подпрограмм и триггеров DML;

- **TEMPLATE.** Соответствует одиночным запросам, параметризованным к указанной форме.

Для редактирования информации, связанной со структурой планов, используется представление просмотра каталогов `sys.plan_guides`. Это представление содержит одну строку для каждой структуры плана в текущей базе данных. Наиболее важными столбцами являются `plan_guide_id`, `name` и `query_text`. Столбец `plan_guide_id` задает уникальный идентификатор структуры плана, столбец `name` определяет ее имя, столбец `query_text` задает текст запроса, для которого создана структура плана.

Существует также много представлений (и функций) динамического управления, связанных с оптимизатором запросов.

Представления динамического управления или DMV (Dynamic Management View):

- `sys.dm_exec_query_optimizer_info`;
- `sys.dm_exec_query_plan`;
- `sys.dm_exec_query_stats`;
- `sys.dm_exec_sql_text`;
- `sys.dm_exec_text_query_plan`;
- `sys.dm_exec_procedure_stats` (начиная с SQL Server 2008).

Представление `sys.dm_exec_query_optimizer_info` является наиболее важным DMV в плане работы оптимизатора запросов, потому что оно возвращает детальную статистику об операции оптимизатора. Можно использовать это представление при настройке рабочей нагрузки для

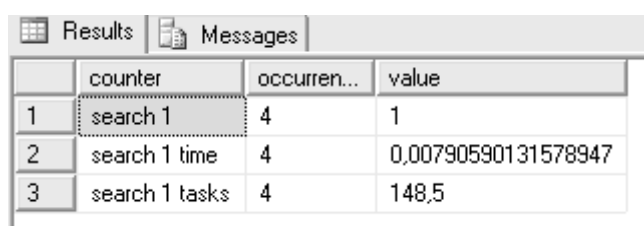
определения проблем оптимизации запросов или для повышения производительности.

Представление `sys.dm__exec_query_optimizer_info` содержит три столбца: `counter`, `occurrence` и `value`. Столбец `counter` содержит имя события оптимизатора, столбец `occurrence` отображает накапливаемое количество появлений этих событий. Значение столбца `value` содержит дополнительную информацию, связанную с событием (не все события выдают значение `value`.)

Используя это представление, можно, например, отобразить общее количество оптимизаций, общее затраченное время и окончательное значение стоимости для сравнения оптимизации запросов для текущей нагрузки и любых изменений, обнаруженных в процессе настроек.

На рисунке 9 приводится результат обращения к представлению `sys.dm_exec_query_optimizer_info` (исследуется, как много раз выполняется оптимизация фазы 1)

```
SELECT counter, occurrence, value  
  
FROM sys.dm_exec_query_optimizer_info  
  
WHERE value IS NOT NULL  
  
AND counter LIKE 'search 1%';
```



	counter	occurren...	value
1	search 1	4	1
2	search 1 time	4	0,00790590131578947
3	search 1 tasks	4	148,5

Рисунок 9. Результат обращения к представлению `sys.dm__exec_query_optimizer_info`.

Видно, что процесс разбивается на три фазы. Существуют три фазы оптимизации: фаза 0, фаза 1 и фаза 2, которые задаются в значениях `search 0`, `search 1` и `search 2` соответственно. Первая (фаза 0) рассматривает только непараллельные планы выполнения. Если стоимость фазы 0 не является оптимальной, то выполняется фаза 1, в которой рассматриваются

непараллельные планы и параллельные планы. Фаза 2 принимает во внимание только параллельные планы.

Представление `sys.dm_exec_query_stats` возвращает общие статистические данные для кэшированных планов запросов. Это представление содержит одну строку на один оператор запроса вместе с кэшированным планом, а время жизни строк связано с самим планом.

Использование приведенных представлений можно посмотреть справочной системе.

*Представление планов запросов в MS SQL Server.* Среда SQL Server Management Studio является интерактивным графическим инструментом, который позволяет анализировать план запросов, а также предоставляет помощь, необходимую для улучшения производительности запросов.

Графическое отображение очень полезно для понимания показателей производительности запроса. Среда SQL Server Management Studio показывает, какой статистики не хватает, тем самым побуждая к оптимизации запросов за счет оценок качества выборки, а также позволяет создать необходимую недостающую статистику.

Параметры плана выполнения графически отображают способы получения данных, выбранные с помощью оптимизатора запросов SQL Server. Для графического отображения плана выполнения используются специальные значки, представляющие выполнение определенных инструкций и запросов SQL Server.

Для отображения плана запроса нужно выполнить следующую последовательность действий:

1. Ввод запроса, для которого необходимо отобразить план выполнения.

В меню Запрос пункты Меню, которые связаны с оптимизацией - Включить действительный план, Включить статистику клиента, Показать предполагаемый план выполнения (Рисунок 10).

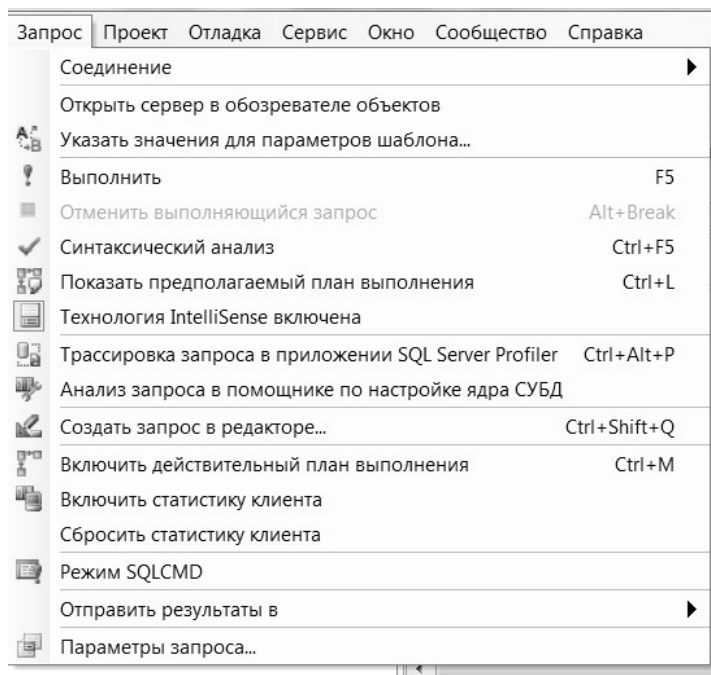


Рисунок 10. Меню Запрос.

2. При выборе пункта меню «Показать предполагаемый план выполнения», план, используемый оптимизатором запросов, отображается на вкладке План выполнения на панели результатов (Рисунок 11).

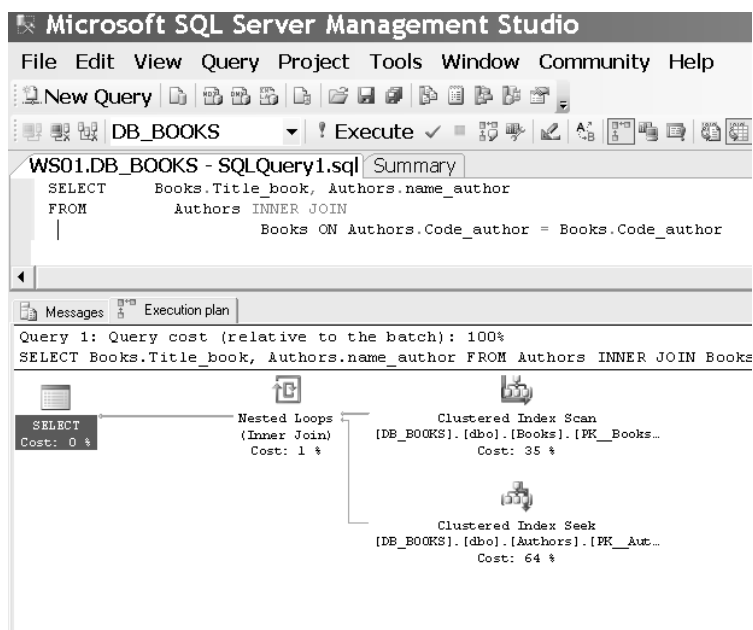


Рисунок 11. Запрос и план его выполнения.

Графическое представление плана выполнения в среде SQL Server Management Studio отображается в виде дерева, которое было рассмотрено в разделе «Процесс выполнения операторов SQL», и читается справа налево и сверху вниз. Каждый проанализированный запрос в пакете отображается

вместе с затратами на его выполнение в процентном отношении от общих затрат на выполнение всего пакета.

При интерпретации результатов графического отображения плана выполнения в среде Management Studio необходимо учитывать следующее:

- Каждый узел древовидной структуры представлен в виде значка, указывающего логический и физический оператор, используемый для выполнения этой части запроса или инструкции.

- Каждый узел связан со своим родительским узлом. Дочерние узлы одного родительского узла отображаются в одном столбце. Однако все узлы в одном столбце не обязательно имеют общий родительский узел. Правила со стрелками на конце соединяют каждый узел с его родителем.

- Операторы показаны в виде символов, связанных с определенным родительским узлом.

- Ширина стрелки пропорциональна количеству строк. Если имеются данные о фактическом количестве строк, используются эти данные. В противном случае используется ориентировочное количество строк.

- Если запрос содержит несколько инструкций, показывается несколько планов выполнения запроса.

- Части древовидных структур определены типом выполняемой инструкции.

- Для параллельных запросов, которые задействуют несколько процессоров, пункт меню Свойства для каждого узла в графическом плане выполнения отображает сведения об используемых потоках операционной системы.

План выполнения запроса на рисунке 25 состоит из следующей последовательности операций: Сканирование таблицы Books с использованием кластеризованного индекса (35%), Сканирование таблицы Authors с использованием кластеризованного индекса (64%), операция соединения таблиц с использованием метода Вложенный цикл (1%).

Ширина стрелки от таблицы Books больше, что означает большее количество передаваемых строк, чем от таблицы Authors.

Особенности представления дерева плана для разных типов инструкций приведены в таблице 1.










Тип инструкции	Элемент древовидной структуры
Язык Transact-SQL и хранимые процедуры	Если инструкция является хранимой процедурой или инструкцией на языке Transact-SQL, она становится корнем древовидной структуры графического представления плана выполнения. Хранимая процедура может иметь несколько дочерних элементов, которые представляют собой инструкции, вызываемые хранимой процедурой. Каждый дочерний элемент — это узел или ветвь дерева.
Язык обработки данных (DML)	Если анализируемая с помощью оптимизатора запросов SQL Server инструкция является DML-инструкцией, такой как SELECT, INSERT, DELETE или UPDATE, она является корневым элементом дерева. DML-инструкции могут содержать до двух дочерних элементов. Первый дочерний элемент — это план выполнения этой DML-инструкции. Второй дочерний элемент представляет собой триггер, если триггер используется в инструкции.
Условные	Графический план выполнения делит условные инструкции, такие как IF...ELSE (при выполнении условия делать одно, в противном случае — что-то другое), на три дочерних элемента. Инструкция IF...ELSE является корнем дерева. Условие IF становится дочерним узлом. Условия THEN и ELSE представлены в виде блоков инструкций. Инструкции WHILE и DO-UNTIL представлены аналогичным образом. Инструкции IF и WHILE имеют собственные значки.
Реляционные операторы	Операции, выполняемые с помощью ядра запросов, такие как сканирование таблиц, соединения и статистические выражения, представлены в виде узлов дерева.








DECLARE CURSOR	Инструкция DECLARE CURSOR является корневым элементом дерева графического плана выполнения; в качестве дочернего элемента выступает связанное с ним выражение.
----------------	--

Таблица 1. Особенности представления дерева плана для разных типов инструкций.

Существует большое количество пиктограмм, которые отображаются в дереве плана выполнения запроса. В таблице 2 приведены некоторые пиктограммы и их описание. Полный список пиктограмм и их описание можно найти в Справочной системе.

Значок	Оператор
	Clustered Index Scan (Сканирование кластеризованного индекса)
	Clustered Index Seek (Поиск в кластеризованном индексе)
	Clustered Index Update (Изменение кластеризованного индекса)
	Concatenation (Сцепление, объединение)
	Delete (Удаление)
	Filter (Выбор)
	Hash Match (Хеширование соединения)
	Insert
	Lazy Spool (сохранение всех строк входных данных в скрытом временном объекте, который хранится в базе данных tempdb. Если оператор сбрасывается на начало (например, оператором Nested Loops), но при этом не требуется повторная привязка, то вместо повторного сканирования ввода используются буферизованные данные)

	Log Row Scan (просмотр журнала транзакций)
	Merge Join (соединение слиянием)
	Nested Loops (соединение вложенным циклом)
	Nonclustered Index Scan (проход по некластеризованному индексу)
	Nonclustered Index Seek (поиск в некластеризованном индексе)
	RID Lookup (поиск RID)
	Row Count Spool (записывать в буферный файл количество строк)
	Segment (сегментация)
	Sort (сортировка)
	Split (разбиение)
	Spool (записывать в буферный файл)
	Table Scan (сканирование таблицы)
	Top

Таблицы 2. Примеры пиктограмм дерева плана.

Каждый узел отображает во всплывающей подсказке, возникающей при наведении указателя мыши на этот узел, соответствующие сведения. Пример параметров узла дерева запроса приведен на рисунке 12.

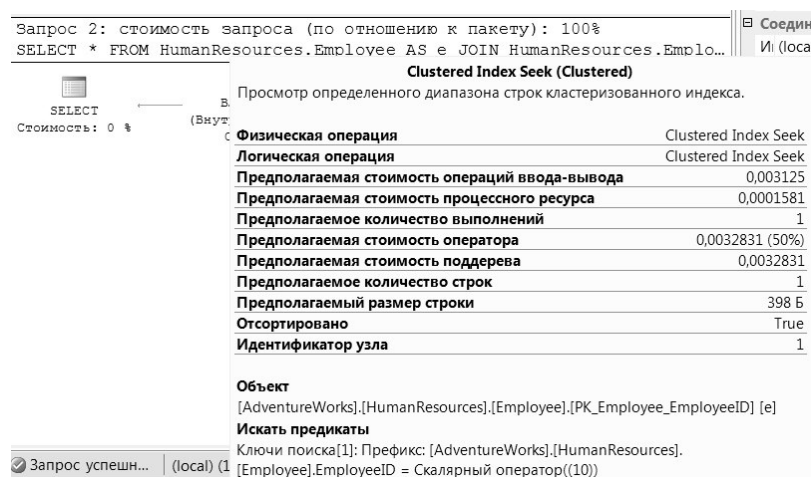


Рисунок 12. Пример параметров узла дерева запроса.

В таблице 3 приведено описание параметров узла дерева запроса.

Элемент	Описание
Физическая операция	Используемый оператор, например Hash Join или Nested Loops. Физические операторы, отображаемые красным цветом, показывают, что оптимизатор запросов выдал предупреждение, например предупреждение о нехватке статистики столбцов или отсутствии предикатов соединения. Это может привести к выбору оптимизатором запросов плана менее эффективного, чем тот, который ожидался. Если графический план выполнения предлагает создание или обновление статистики либо создание индекса, отсутствующая статистика или индексы могут быть немедленно созданы или изменены с использованием контекстных меню обозревателя объектов среды SQL Server Management Studio.
Логическая операция	Логический оператор, который соответствует физическому оператору, например, оператору Inner Join. Имя логического оператора приводится после физического оператора сверху всплывающей подсказки.
Предполагаемый размер строки	Предполагаемый размер строки, получаемой на выходе оператора (в байтах).

Предполагаемая стоимость операций ввода-вывода	Приблизительные затраты на выполнение действий ввода-вывода для данной операции. Это значение должно быть минимально возможным.
Предполагаемая стоимость процессного ресурса	Приблизительные затраты на произведение ЦП всех вычислений для данной операции.
Предполагаемая стоимость оператора	Затраты оптимизатора запросов на выполнение этой операции. Затраты на выполнение этой операции в процентном отношении к общим затратам на выполнение запроса отображаются в скобках. Так как ядро запросов выбирает наиболее эффективную операцию для выполнения запроса или инструкции, это значение должно быть минимально возможным.
Предполагаемая стоимость поддерева	Общие затраты оптимизатора запросов на выполнение этой и всех предшествующих операций в данной ветви дерева.
Предполагаемое количество строк	Количество строк, выдаваемых оператором.

Таблица 3. Описание параметров узла дерева запроса.

Не все узлы в графическом плане выполнения имеют всплывающие подсказки, которые приведены в таблице.

*Примеры оптимизации планов выполнения запроса.* Рассмотрим некоторые примеры оптимизации запросов и представление соответствующих деревьев их плана выполнения. На рисунке 13 приведен запрос и план его выполнения. В данном примере таблица Клиент1 не имеет индекса по полю Имя. Для выбора информации по условию Имя = 'Петров' используется просмотр строк таблицы (т.е. последовательно читается вся таблица без использования индекса).

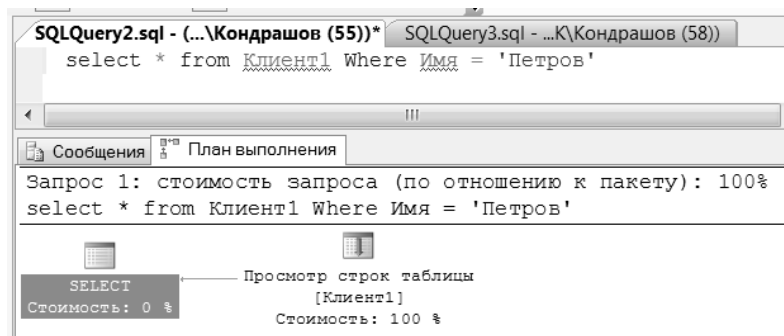


Рисунок 13. Пример плана выполнения запроса.

На рисунке 14 приведен другой запрос к этой же таблице и план его выполнения. В этом примере таблица Клиент1 имеет некластеризованный индекс по полю Кодклиента. Для выбора информации по условию Кодклиента =2 сначала используется поиск в индексе и по найденному значению RID производится обращение к соответствующей странице данных таблицы Клиент1. Таким образом, запрос с условием для индексированного поля может использовать оптимальный план с применением индекса. Для запроса с условием для неиндексированного поля индекс в плане использоваться не может.

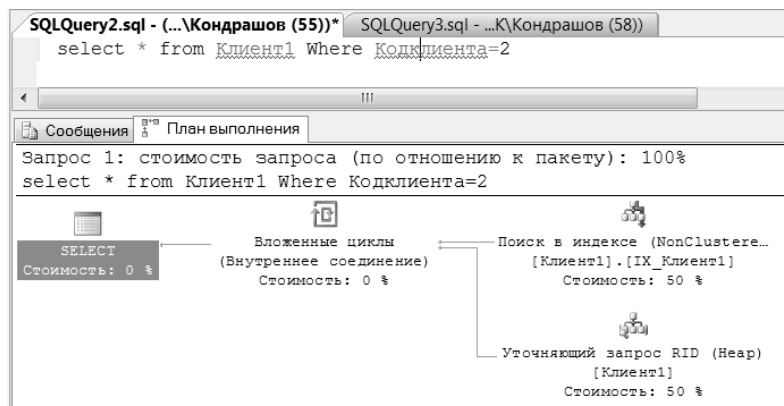


Рисунок 14. Пример плана выполнения запроса.

На рисунках 15, 16 приведены планы выполнения для запросов к таблице Клиент, которая имеет кластеризованный индекс. На рисунках видно, что независимо от того, по какому полю задано условие в запросе, используется Кластеризованный индекс (в разделе Организация хранения данных в MS SQL Server было показано, что кластеризованный индекс объединяет дерево индекса и данные). Отличие состоит в том, что если

условие в запросе задано по индексированному полю (в данном примере поле Кодклиента), то используется поиск в индексном дереве (Clustered Index Seek, рисунок 16). В противном случае используется последовательное сканирование индексного дерева (Clustered Index Scan, рисунок 15).

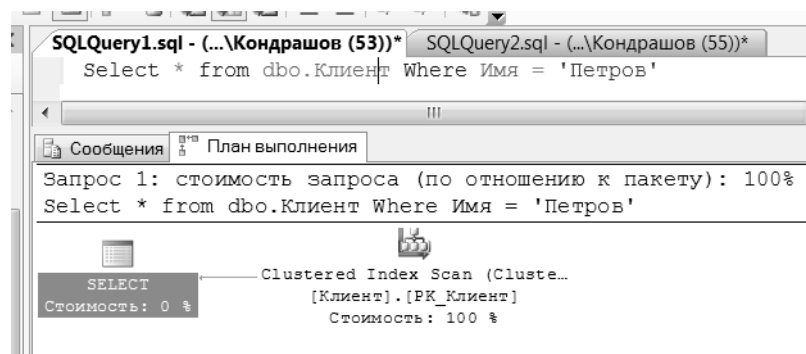


Рисунок 15. Пример плана выполнения запроса.

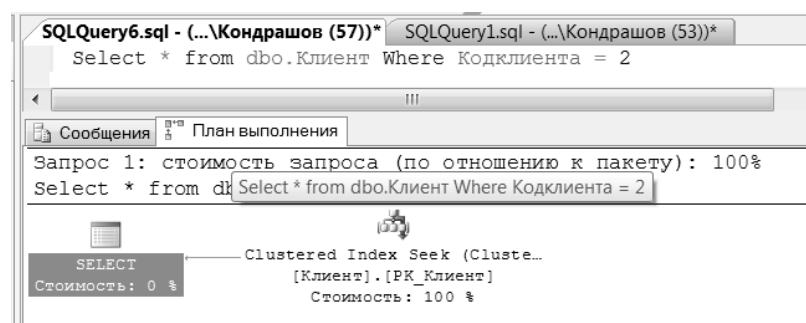


Рисунок 16. Пример плана выполнения запроса.

На рисунках 17, 18 приведены планы выполнения для запросов к таблице new\_addresses, которая имеет некластеризованный индекс для столбца StateProvinceID. Первый запрос по условию a.StateProvinceID = 32 является высоко селективным (выбирается одна строка из 19814):

```
SELECT * FROM new_addresses AS a
WHERE a.StateProvinceID = 32;
```

В этом случае из структуры плана видно (рисунок 17), что оптимизатор использует поиск в индексе для столбца StateProvinceID с последующим побращением к странице данных таблицы по найденному RID. Оператор Nested Loops (вложенный цикл) отображается для запроса к одной таблице (внешний цикл поиск в индексе, внутренний – обращение к таблице).

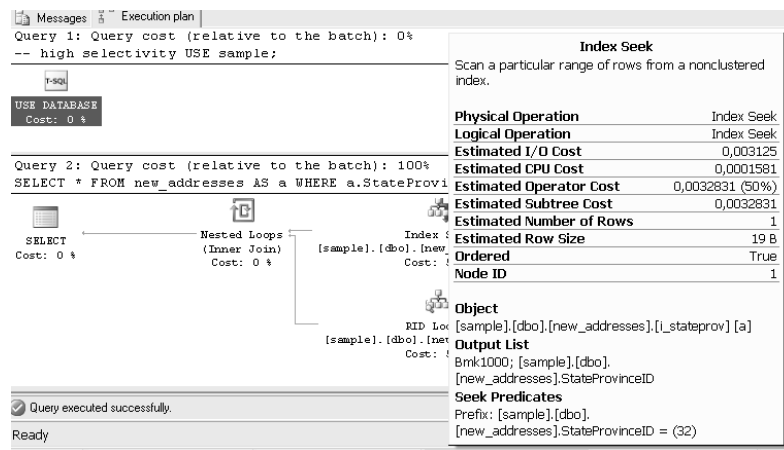


Рисунок 17. Пример плана выполнения запроса с высокой селективностью.

Второй запрос по условию `a.StateProvinceID = 9` является низко селективным (отношение количества строк, удовлетворяющему условию, к общему количеству строк равно 23%):

```
SELECT * FROM new_addresses a
WHERE a.StateProvinceID =9;
```

В этом случае из структуры плана видно (рисунок 18), что оптимизатор не использует поиск в индексе для столбца `StateProvinceID`, а обращается непосредственно к страницам таблицы данных (индекс не используется, таблица сканируется).

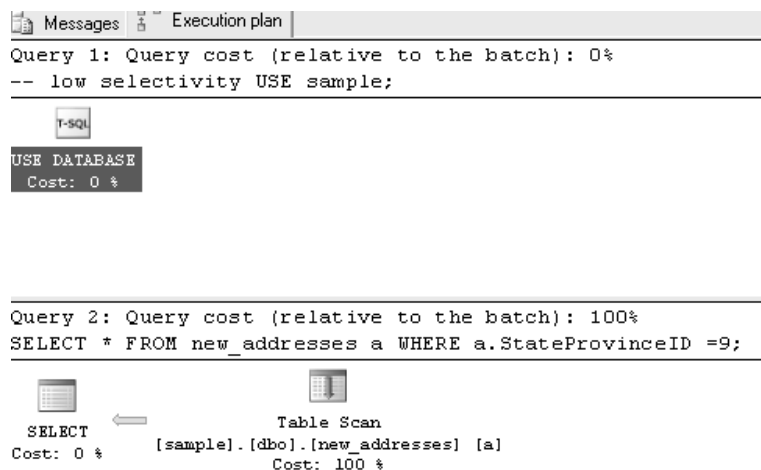


Рисунок 18. Пример плана выполнения запроса с низкой селективностью.

Рассмотрим пример использования подсказки оптимизатору `WITH(INDEX(0))`. Если не существует кластеризованный индекс для таблицы, то `INDEX (0)` требует сканирования таблицы. При существовании

кластеризованного индекса INDEX (0) определяет сканирование кластеризованного индекса.

Запрос:

*USE AdventureWorks;*

*SELECT \* FROM Person.Address AS a*  
*WHERE a.StateProvinceID = 32.*

Дерево плана выполнения запроса показано на рисунке 19.

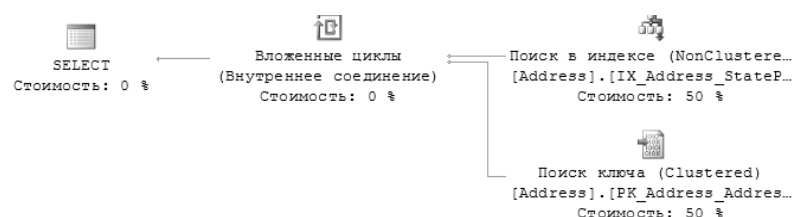


Рисунок 19. Пример плана выполнения запроса без подсказки оптимизатору.

Запрос:

*USE AdventureWorks;*

*SELECT \* FROM Person.Address AS a*  
*WITH(INDEX(0))*

*WHERE a.StateProvinceID = 32.*

Дерево плана выполнения запроса показано на рисунке 20. Видно, что с этой подсказкой не используется поиск в некластеризованном индексе.

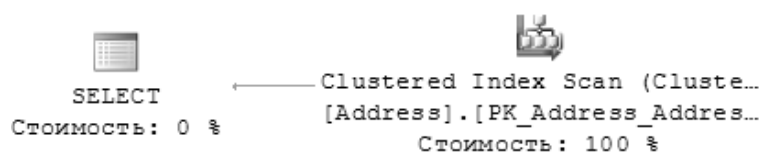


Рисунок 20. Пример плана выполнения запроса с подсказкой оптимизатору.

Рассмотрим примеры планов выполнения запросов с разными техниками обработки соединения таблиц. В соединяемых столбцах таблиц HumanResources.Employee и HumanResources.EmployeeAddress имеются первичные ключи и, соответственно, — кластеризованные индексы. На рисунке 21 показано, что в таблице HumanResources.EmployeeAddress



имеется составной кластеризованный индекс по полям EmployeeID и AddressID.

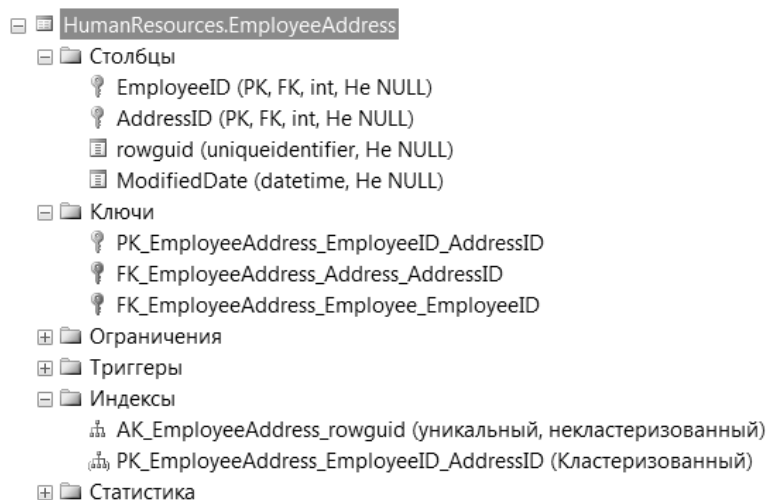


Рисунок 21. Таблица HumanResources.EmployeeAddress.

В первом запросе задано условие  $e.EmployeeID = 10$ :

*USE AdventureWorks;*

*SELECT \* FROM HumanResources.Employee AS e*

*JOIN HumanResources.EmployeeAddress AS a*

*ON e.EmployeeID = a.EmployeeID*

*AND e.EmployeeID = 10;*

Дерево плана запроса показано на рисунке 22. Видно, что используется техника обработки соединения Вложенные циклы. Оптимизатор принимает решение использовать вложенные циклы, так как существует дополнительный фильтр ( $e.EmployeeID = 10$ ), сокращающий результирующий набор до одной строки. Сначала производится поиск по индексу в таблице HumanResources.Employee соответствующей условию строки. Далее поиск в таблице HumanResources.EmployeeAddress. Оператор Clustered Index Seek использует поисковые возможности индексов для получения строк из кластеризованного индекса.

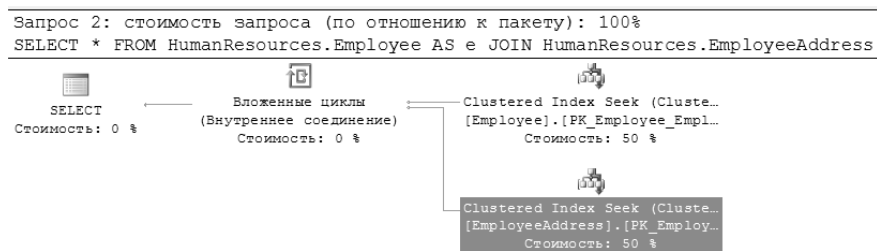


Рисунок 22. Пример плана выполнения запроса.

В данном примере последовательно производится обращение к таблицам HumanResources.Employee и HumanResources.EmployeeAddress. В такой последовательности эти таблицы указаны и в запросе (FROM HumanResources.Employee AS e JOIN HumanResources.EmployeeAddress AS a). Поменяем последовательность таблиц в запросе:

*USE AdventureWorks;*

```
SELECT * FROM HumanResources.EmployeeAddress AS a  
JOIN HumanResources.Employee AS e  
ON e.EmployeeID = a.EmployeeID  
AND e.EmployeeID = 10;
```

Дерево плана останется неизменным. Т.е. изменение последовательности таблиц в запросе не повлияло на последовательность их обработки. Это происходит потому, что оптимизатор определяет первоначальный выбор из таблицы HumanResources.Employee наилучшим (по одной найденной строке производится одно обращение к таблице HumanResources.EmployeeAddress по найденному значению). Если использовать другую последовательность обработки таблиц, то для каждой строки таблицы HumanResources.EmployeeAddress нужно много раз обращаться к таблице HumanResources.Employee и стоимость выполнения запроса будет выше.

Если из запроса удалить условие:

*USE AdventureWorks;*

```
SELECT * FROM HumanResources.Employee AS e  
JOIN HumanResources.EmployeeAddress AS a  
ON e.EmployeeID = a.EmployeeID,
```

то техника обработки соединения изменится. Дерево плана выполнения запроса показано на рисунке 23.

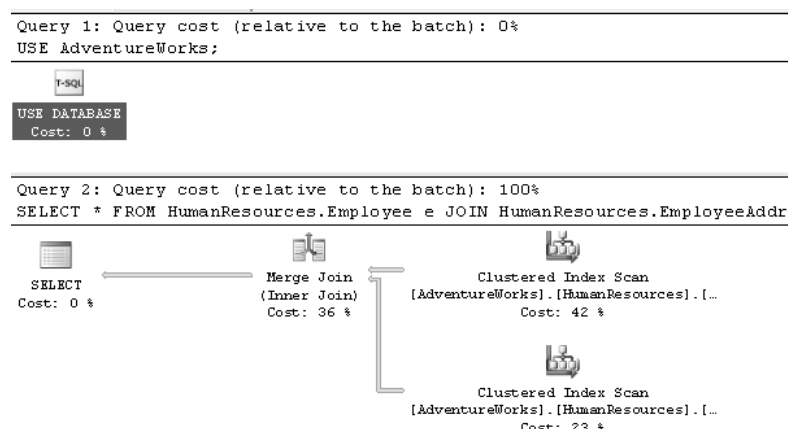


Рисунок 23. Пример плана выполнения запроса.

В этом запросе отсутствует условие `e.EmployeeID = 10` и результирующий набор содержит много строк, но они упорядочены по индексу. В этом случае оптимизатор выбирает метод слияния соединения. Merge Join (Слияние соединения).

Рассмотрим план выполнения запроса:

*USE AdventureWorks;*

*SELECT \* FROM Person.Address a JOIN Person.StateProvince s*  
*ON a.StateProvinceID = s.StateProvinceID*

Таблица Person.Address, ее ключи и индексы приведены на рисунке 24.

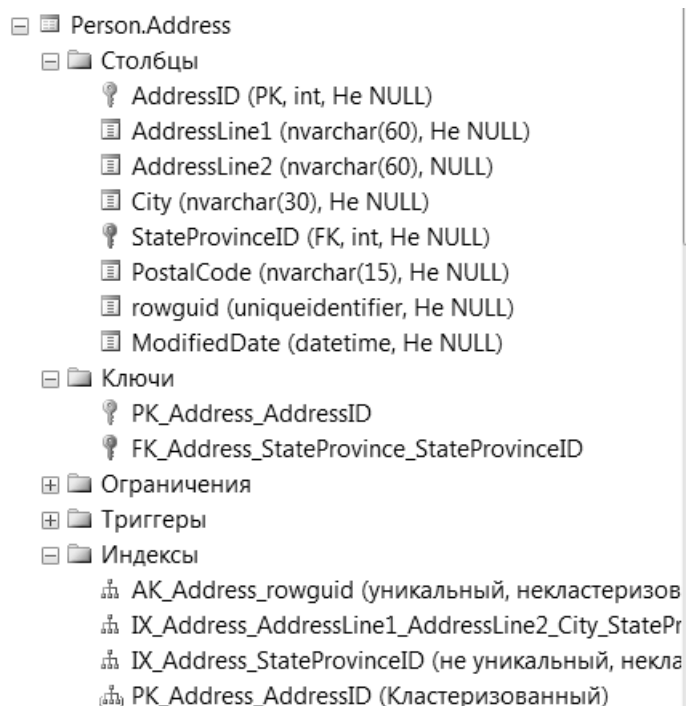


Рисунок 24. Таблица Person.Address.

Таблица Person.StateProvince, ее ключи и индексы приведены на рисунке 25.

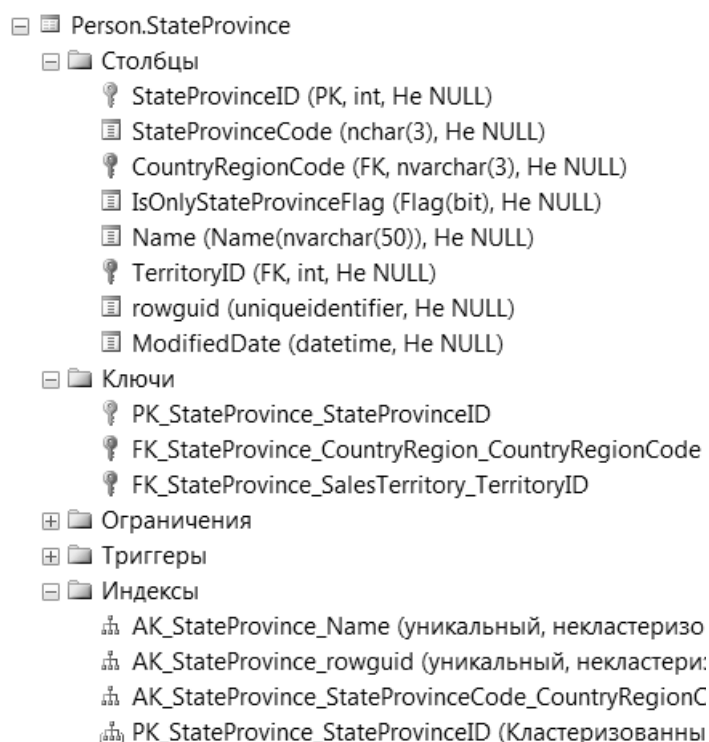


Рисунок 25. Таблица Person.StateProvince.

Дерево плана выполнения запроса показано на рисунке 26.

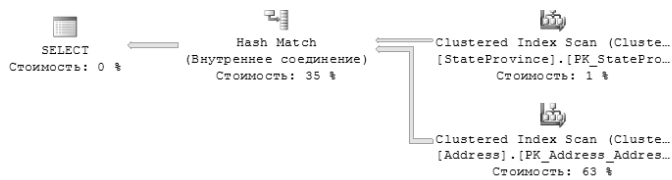


Рисунок 26. Пример дерева плана выполнения запроса.

Видно (рисунки 24, 25), что таблицы, участвующие в соединении, имеют кластеризованные индексы и дерево плана использует соответственно сканирование кластеризованного индекса для каждой из таблиц. Техника обработки соединения – хеширование соединения. Это происходит потому, что записи таблицы `Person.Address` не отсортированы по полю `StateProvinceID`.

Приведем пример позказки соединения. Выше был приведен запрос:

*USE AdventureWorks;*

*SELECT \* FROM Person.Address a JOIN Person.StateProvince s*  
*ON a.StateProvinceID = s.StateProvinceID*

и план его выполнения, в котором использовалась техника обработки соединения – хеширование соединения (Hash Match).

При добавлении в запрос подсказки

*USE AdventureWorks;*

*SELECT \* FROM Person.Address a JOIN Person.StateProvince s*  
*ON a.StateProvinceID = s.StateProvinceID*  
*OPTION (MERGE JOIN),*

техника обработки соединения меняется (дерево плана показано на рисунке 27). Используется техника обработки соединения – слиянием. На рисунке видно, что предварительно выполняется операция сортировки (Sort).

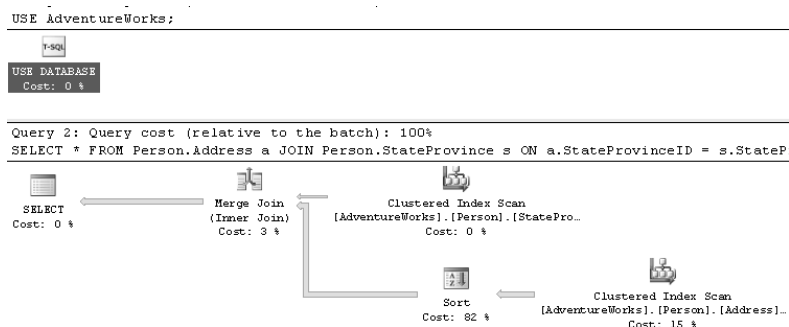


Рисунок 27. Пример плана выполнения запроса с подсказкой соединения.

## 26. Точечные оценки и их свойства (несмещенность, состоятельность и эффективность). Метод максимального правдоподобия и метод моментов.

В теории вероятностей неявно предполагалось, что генеральное распределение известно. Однако в математической статистике такая ситуация встречается достаточно редко. Напротив, типичной является ситуация, при которой генеральное распределение описывается неоднозначно. Далее распределения, подпадающие под такое описание, будем называть *допустимыми*. Пусть  $\mathcal{L}$  — множество всех допустимых распределений.

Тот факт, что распределение случайной величины  $X$  является допустимым, означает, очевидно, что

$$Law(X) \in \mathcal{L}.$$

Достаточно часто будем предполагать, что  $\mathcal{L}$  — параметрическое семейство распределений. Это означает, что  $\mathcal{L}$  можно представить в виде

$$\mathcal{L} = \{L_{\theta} : \theta \in \Theta\},$$

где  $L_{\theta}$  — распределение, однозначно определяемое параметром  $\theta$ , а  $\Theta \subset \mathbb{R}^d$  — множество значений параметра  $\theta$ , задающих допустимые распределения. Если параметр  $\theta$  записан в виде вектора

$$\theta = (\theta_1, \dots, \theta_d),$$

то говорят, что распределение  $L_{\theta}$  зависит от  $d$  (скалярных) параметров:  $\theta_1, \dots, \theta_d$ .

Пусть  $X_1, \dots, X_n$  — случайная выборка объема  $n$  из распределения  $Law(X)$ . Набор  $X_1, \dots, X_n$  называется *выборочными данными*, а любая вычисляемая по выборочным данным случайная величина  $h(X_1, \dots, X_n)$  называется *статистикой*.

В качестве примеров статистик можно привести сами случайные величины  $X_1, \dots, X_n$  и эмпирические характеристики выборки:  $\bar{X}$ ,  $\widehat{D}(X)$ , ...

Из определения статистики вытекает, что параметры распределения (генеральные характеристики) статистиками *не являются*. Не являются статистиками и комбинации статистик с параметрами распределения. Например,  $\bar{X} - E(X)$  не является статистикой, если  $E(X) \neq const$  для допустимых распределений.

Также заметим, что определение статистики не предполагает скалярного характера статистики: статистика может быть случайным вектором.

Примером двумерной статистики может служить случайный вектор  $(\bar{X}, \widehat{D}(X))$ , примером многомерной статистики является вектор  $\mathbf{X} = (X_1, \dots, X_n)$ .

Пусть  $\theta$  — параметр или иная характеристика генерального распределения. Статистика  $\hat{\theta} = \hat{\theta}_n = h(X_1, \dots, X_n)$ , предназначенная для приближенного вычисления  $\theta$ , называется *статистической оценкой*  $\theta$ .

*Определение.* Скалярная или векторная статистическая оценка  $\hat{\theta}$  называется *несмещенной*, если для всех допустимых генеральных распределений выполняется равенство

$$E(\hat{\theta}) = \theta.$$

Рассмотрим генеральный и выборочный начальные моменты порядка  $k$ :

$$\begin{aligned} v_k &= v_k(X) = E(X^k), \\ \hat{v}_k &= \hat{v}_k(X) = \frac{1}{n} \sum_{i=1}^n X_i^k. \end{aligned}$$

Выборочный момент  $\hat{\nu}_k$  часто используется в качестве замены неизвестного генерального момента  $\nu_k$ , другими словами,  $\hat{\nu}_k$  играет роль статистической оценки для  $\nu_k$ . Предположив, что для любого допустимого распределения момент  $\nu_k$  существует, покажем, что  $\hat{\nu}_k$  является его несмещенной оценкой:

$$E(\hat{\nu}_k) = \frac{1}{n} \sum_{i=1}^n E(X_i^k) = \frac{n}{n} E(X^k) = \nu_k.$$

Центральные выборочные моменты  $\hat{\mu}_k$  также иногда используются в качестве замены неизвестных генеральных моментов  $\mu_k$ . По аналогии с начальными моментами можно было бы предположить, что и  $E(\hat{\mu}_k) = \mu_k$ . Однако это неверно.

*Теорема.*  $E(\hat{\mu}_2) = \frac{n-1}{n} \mu_2.$

*Доказательство.* Используя тождества  $\hat{\mu}_2 = \hat{\nu}_2 - \hat{\nu}_1^2$  и  $\mu_2 = \nu_2 - \nu_1^2$ , с учетом несмещенности оценок  $\hat{\nu}_1$  и  $\hat{\nu}_2$  находим

$$\begin{aligned} E(\hat{\mu}_2) &= E(\hat{\nu}_2) - E(\hat{\nu}_1^2) = \\ &= \nu_2 - \{D(\hat{\nu}_1) + (E(\hat{\nu}_1))^2\} = \\ &= \nu_2 - \left(\frac{1}{n} \mu_2 + \nu_1^2\right) = \frac{n-1}{n} \mu_2. \end{aligned}$$

*Определение.* Исправленная эмпирическая дисперсия задается формулой

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Из предыдущей теоремы вытекает несмещенность исправленной дисперсии как статистической оценки генеральной дисперсии:

$$\begin{aligned} E(s^2) &= E\left(\frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \\ &= \frac{n}{n-1} E(\hat{\mu}_2) = \frac{n}{n-1} \cdot \frac{n-1}{n} \mu_2 = \mu_2. \end{aligned}$$



Предположим, что все допустимые генеральные распределения  $Law(X) \in \mathcal{L}$  обладают общим известным средним  $a = E(X)$ . В этом случае для оценки генеральной дисперсии используется другая статистика

$$s_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2.$$

Эта статистика также является несмещенной оценкой дисперсии:

$$\begin{aligned} E(s_0^2) &= \frac{1}{n} E \left( \sum_{i=1}^n (X_i - a)^2 \right) = \frac{1}{n} \sum_{i=1}^n E \left( (X_i - E(X))^2 \right) = \\ &= \frac{1}{n} \sum_{i=1}^n D(X_i) = \mu_2. \end{aligned}$$

*Определение.* Статистическая оценка  $\hat{\theta} = \hat{\theta}_n$  называется состоятельной, если при любом допустимом генеральном распределении

$$\text{plim}_{n \rightarrow \infty} \hat{\theta}_n = \theta.$$

Более подробно: пусть  $\hat{\theta}_n = h_n(X_1, \dots, X_n)$  — функция от выборочных данных, тогда состоятельность  $\hat{\theta}_n$  означает, что для всякого положительного  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|h_n(X_1, \dots, X_n) - \theta| \geq \varepsilon) = 0.$$

*Пример.* Будем считать допустимыми все распределения  $Law(X)$ , для которых существует математическое ожидание  $E(X) = \mu$  и конечная дисперсия,  $D(X) < \infty$ . Рассмотрим выборку

$$X_1, \dots, X_n \sim iid \text{ } Law(X).$$

По теореме Чебышева (закон больших чисел) имеем:

$$\text{plim}_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu.$$

Следовательно, выборочное среднее  $\bar{X}$  — состоятельная оценка генерального среднего  $\mu$ .

*Пример.* Рассмотрим выборку из распределения Бернулли:

$$X_1, \dots, X_n \sim iid \text{ } Bin(1, p).$$

Выборочную долю значения  $X = 1$ , как обычно, обозначим  $\hat{p}$ . По теореме Бернулли  $\text{plim } \hat{p} = p$ . Следовательно, *выборочная доля  $\hat{p}$  — состоятельная оценка генеральной доли  $p$ .*

*Метод максимального правдоподобия и метод моментов*

*Метод моментов построения статистических оценок*

Предположим, что генеральное распределение  $L$  зависит от  $k$  параметров:  $\theta_1, \dots, \theta_k$ . Все эти параметры могут объединяться в один вектор

$$\theta = (\theta_1, \dots, \theta_k).$$

Соответствующее распределение обозначается  $L_{\theta_1, \dots, \theta_k}$  или просто  $L_\theta$ . Поскольку не любая комбинация параметров задает допустимое распределение, вводится понятие *множества допустимых значений параметров*  $\Theta \subset \mathbb{R}^k$ . Распределение  $L_\theta$  является, по определению, *допустимым*, если  $\theta \in \Theta$ . Таким образом, семейство всех допустимых генеральных распределений является параметрическим и может быть записано в виде

$$\{L_\theta: \theta \in \Theta\}.$$

Поскольку генеральное распределение зависит от параметров  $\theta_1, \dots, \theta_k$ , то и любая числовая характеристика генерального распределения также зависит от  $\theta_1, \dots, \theta_k$ . Покажем это на примере начальных моментов. Предположим, что дискретное генеральное распределение  $L_{\theta_1, \dots, \theta_k}$  задается таблицей

$X$	$x_1(\theta_1, \dots, \theta_k)$	$x_2(\theta_1, \dots, \theta_k)$	...
$P$	$p_1(\theta_1, \dots, \theta_k)$	$p_2(\theta_1, \dots, \theta_k)$	...

Тогда начальный момент порядка  $m$  является функцией  $v_m = g_m(\theta_1, \dots, \theta_k)$  вида

$$g_m(\theta_1, \dots, \theta_k) = \sum_{i=1}^{\infty} x_i^m(\theta_1, \dots, \theta_k) p_i(\theta_1, \dots, \theta_k).$$

В случае абсолютно непрерывного генерального распределения с плотностью  $f(x; \theta_1, \dots, \theta_k)$  зависимость  $v_m$  от параметров задается функцией вида

$$g_m(\theta_1, \dots, \theta_k) = \int_{-\infty}^{\infty} x^m f(x; \theta_1, \dots, \theta_k) dx.$$

На первом шаге метода моментов составляется система уравнений вида:

$$(I) \quad \begin{cases} v_1 = g_1(\theta_1, \dots, \theta_k); \\ \dots \\ v_k = g_k(\theta_1, \dots, \theta_k). \end{cases}$$

Предположим, что систему (I) можно разрешить относительно параметров в виде системы

$$(II) \quad \begin{cases} \theta_1 = h_1(v_1, \dots, v_k); \\ \dots \\ \theta_k = h_k(v_1, \dots, v_k); \end{cases}$$

где  $h_1(v_1, \dots, v_k), \dots, h_k(v_1, \dots, v_k)$  — непрерывные функции от  $v_1, \dots, v_k$ .

*Определение.* Оценки параметров распределения вида

$$\hat{\theta}_j = h_j(\hat{v}_1, \dots, \hat{v}_k),$$

где  $j = 1, \dots, k$ , а  $\hat{v}_1, \dots, \hat{v}_k$  — выборочные моменты, называются **оценками метода моментов**.

*Теорема.* Если генеральное распределение зависит от  $k$  параметров и при любом допустимом их наборе существует начальный момент порядка  $k$ , то оценки метода моментов являются состоятельными.

*Доказательство.* Поскольку существует генеральный момент порядка  $k$ , выборочные начальные моменты — состоятельные оценки генеральных:

$$\text{plim } \hat{v}_j = v_j, \quad j = 1, \dots, k.$$

Используя теорему Slutsky, находим

$$\begin{aligned} \text{plim } \hat{\theta}_j &= \text{plim } h_j(\hat{v}_1, \dots, \hat{v}_k) = \\ &= h_j(\text{plim } \hat{v}_1, \dots, \text{plim } \hat{v}_k) = h_j(v_1, \dots, v_k) = \theta_j, \quad j = 1, \dots, k. \end{aligned}$$

*Пример.* Пусть  $N(\mu, \sigma^2)$  — генеральное распределение. Найдем оценки параметров  $\mu, \sigma^2$  методом моментов. Составляем сначала систему (I):

$$(I) \quad \begin{cases} v_1 = \mu; \\ v_2 = \sigma^2 + \mu^2. \end{cases}$$

Затем составляем систему (II)

$$(II) \quad \begin{cases} \mu = v_1; \\ \sigma^2 = v_2 - v_1^2. \end{cases}$$

Заменяя генеральные моменты выборочными, получим оценки метода моментов:

$$\hat{\mu} = \hat{v}_1 = \bar{X};$$

$$\hat{\sigma}^2 = \hat{v}_2 - \hat{v}_1^2 = \hat{\mu}_2 = \hat{D}(X).$$

*Замечание.* Поскольку выборочные центральные моменты являются состоятельными оценками генеральных центральных моментов при тех же условиях, что и начальные, часть начальных моментов в методе моментов можно заменить на центральные.

*Пример.* Генеральное распределение — это равномерное распределение на отрезке  $[a, b]$ . Требуется найти методом моментов оценки параметров  $a$  и  $b$ .

*Решение.* Составляем системы (I) и (II):

$$(I) \quad \begin{cases} v_1 = \frac{a+b}{2}; \\ \mu_2 = \frac{(b-a)^2}{12}. \end{cases} \Leftrightarrow \begin{cases} v_1 = \frac{a+b}{2}; \\ \sqrt{3\mu_2} = \frac{b-a}{2}. \end{cases}$$

$$(II) \quad \begin{cases} a = v_1 - \sqrt{3\mu_2}; \\ b = v_1 + \sqrt{3\mu_2}. \end{cases}$$

Для получения оценок метода моментов осталось заменить генеральные моменты выборочными:

$$\hat{a} = \hat{v}_1 - \sqrt{3\hat{\mu}_2};$$

$$\hat{b} = \hat{v}_1 + \sqrt{3\hat{\mu}_2}.$$

### *Метод максимального правдоподобия*

Как и в методе моментов предположим, что семейство  $\mathcal{P}_X$  допустимых генеральных распределений является параметрическим

$$\mathcal{P}_X = \{P_{X,\theta} : \theta \in \Theta\},$$

где  $\Theta \subset \mathbb{R}^k$  — множество допустимых значений векторного параметра  $\theta = (\theta_1, \dots, \theta_k)$ .

*Метод максимального правдоподобия* применяется в следующих двух случаях.

*Случай 1.* При любом  $\theta \in \Theta$  генеральное распределение  $P_{X,\theta}$  имеет плотность  $f(x; \theta)$ .

*Случай 2.* Существует конечное или счетное множество  $\mathcal{S} = \{x_1, x_2, \dots\}$ , такое, что  $P_{X,\theta}(\mathcal{S}) = 1$  при любом  $\theta \in \Theta$ . Другими словами, генеральное распределение  $P_{X,\theta}$  является дискретным, причем возможные значения не зависят от  $\theta$ .

С целью унификации обозначений в случае 2 зададим функцию  $f(., \theta) : \mathcal{S} \rightarrow [0,1]$  формулой

$$f(x; \theta) = P_{X,\theta}(\{x\}).$$

Поскольку распределение  $P_{X,\theta}$  — это закон распределения случайной величины  $X$ , заданной на некотором вероятностном пространстве  $(\Omega, \mathcal{F}, P_\theta)$  с вероятностной мерой  $P_\theta$ , функцию  $f(x; \theta)$  можно записать в виде

$$f(x; \theta) = P_\theta(X = x).$$

Таким образом, функция  $f(x; \theta)$  в случае 2 имеет смысл вероятности попадания в точку  $x$  (в отличие от случая 1, в котором  $f(x; \theta)$  имеет смысл плотности вероятности).

*Определение.* В случаях 1 и 2 функция правдоподобия задается формулой

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

а логарифмическая функция правдоподобия — формулой

$$\ln L(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \ln f(x_i; \theta).$$

Пусть  $\mathbf{X} = (X_1, \dots, X_n)$  — случайная выборка из распределения  $P_{X, \theta}$ , а  $\mathbf{x} = (x_1, \dots, x_n)$  — некоторая ее реализация. Поскольку  $X_1, \dots, X_n$  — независимые случайные величины, плотность их совместного распределения есть произведение плотностей  $f(x_i; \theta), i = 1, \dots, n$ . Таким образом, в случае 1 функция правдоподобия — это не что иное, как *плотность распределения* случайного вектора  $\mathbf{X}$  в точке  $\mathbf{x}$ . В случае 2 ситуация аналогичная: функция правдоподобия — это *вероятность* попадания случайного вектора  $\mathbf{X}$  в точку  $\mathbf{x}$ .

*Определение.* Пусть  $X_1, \dots, X_n$  — случайная выборка из распределения  $P_{X, \theta}$ , зависящего от параметра  $\theta \in \Theta$ . Оценкой максимального правдоподобия параметра  $\theta$  называется статистика вида

$$\hat{\theta} = h(X_1, \dots, X_n),$$

где  $h(x_1, \dots, x_n)$  — это единственная точка глобального максимума функции правдоподобия  $L(x_1, \dots, x_n; \theta)$  на множестве  $\Theta$  при фиксированных  $x_1, \dots, x_n$ .

*Пример.* Пусть  $X_1, \dots, X_n$  — выборка из распределения из распределения Бернулли  $Bin(1, \theta)$ . Требуется найти методом максимального правдоподобия оценку  $\hat{\theta}$ .

*Решение.* Каждое выборочное значение  $X_i$  распределено по закону

$X_i$	0	1
$P$	$1 - \theta$	$\theta$

Следовательно, в точках  $x = 0$  и  $x = 1$  плотность вероятности  $f(x; \theta)$  имеет вид

$$f(0; \theta) = 1 - \theta, \quad f(1; \theta) = \theta.$$

Отсюда получаем функцию правдоподобия

$$L(x_1, \dots, x_n; \theta) = \theta^k (1 - \theta)^{n-k},$$

где  $k$  — число единиц среди  $x_1, \dots, x_n$ .

Находим производную

$$\begin{aligned}\frac{\partial}{\partial \theta} L(x_1, \dots, x_n; \theta) &= k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1} = \\ &= \theta^{k-1}(1-\theta)^{n-k-1}\{k(1-\theta) - (n-k)\theta\} = \\ &= \theta^{k-1}(1-\theta)^{n-k-1}\{k - n\theta\}.\end{aligned}$$

Точки  $\theta = 0$  и  $\theta = 1$  не могут быть точками максимума, так как в них функция равна нулю.

Поэтому точка глобального максимума (ее существование вытекает из непрерывности  $L$ ) является внутренней точкой отрезка  $[0,1]$  вследствие чего она задается уравнением  $k - n\theta = 0$ .

Это уравнение имеет единственное решение, которое и является искомой оценкой:

$$\hat{\theta} = \frac{k}{n} = \bar{X}.$$

Точки глобального максимума функций  $L(x_1, \dots, x_n; \theta)$  и  $\ln L(x_1, \dots, x_n; \theta)$  на  $\theta$  при фиксированных  $x_1, \dots, x_n$ , очевидно, совпадают.

Найдем теперь оценку  $\hat{\theta}$ , используя логарифмическую функцию правдоподобия.

$$\begin{aligned}\frac{d}{d\theta} \ln L(x_1, \dots, x_n; \theta) &= \frac{d}{d\theta} \{k \ln \theta + (n-k) \ln(1-\theta)\} = \\ &= \frac{k}{\theta} - \frac{n-k}{1-\theta} = \frac{k - n\theta}{\theta(1-\theta)}.\end{aligned}$$

Приравнивая к нулю производную, вновь получаем уравнение  $k - n\theta = 0$ .

На данном примере видно, что применение логарифмической функции правдоподобия упрощает получение оценки максимального правдоподобия.

В других случаях сокращение вычислений может быть еще более существенным.

**27. Проверка гипотез об определенных значениях параметров нормальных распределений. Проверка гипотезы о равенстве средних при известной дисперсии. Проверка гипотезы о равенстве средних при неизвестной дисперсии. Р-значение критерия. Определение и способ его вычисления.**

*Проверка гипотезы об определенном значении параметра нормального распределения.* Пусть  $X_1, \dots, X_n$  – выборка из распределения  $N(\mu, \sigma^2)$ . Гипотеза об определенном значении генерального среднего состоит в том, что эта выборка получена при фиксированном значении  $\mu = \mu_0$ . Структура критерия по проверке данной гипотезы зависит от того, известна или нет генеральная дисперсия  $\sigma^2$ , а также от вида альтернативной гипотезы. Аналогично гипотеза об определенном значении дисперсии  $\sigma^2$  состоит в том, что выборка  $X_1, \dots, X_n$  получена при фиксированном значении  $\sigma^2 = \sigma_0^2$ , а способ ее проверки зависит от того, известно или нет генеральное среднее  $\mu$  и, конечно, от вида альтернативной гипотезы. Отметим, что во всех случаях прослеживается явное сходство с построением доверительных интервалов для соответствующих параметров нормального распределения.

*Проверка гипотезы об определенном значении генерального среднего при известной дисперсии.* Для проверки гипотезы  $H_0: \mu = \mu_0$  против любой из трех альтернативных гипотез  $H_1$ : 1)  $\mu > \mu_0$ , 2)  $\mu < \mu_0$  или 3)  $\mu \neq \mu_0$  используется статистика

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Критическая область  $K$  выбирается в соответствии с таблицей

$H_1$	$K$
$\mu > \mu_0$	$Z > Z_\alpha$
$\mu < \mu_0$	$Z < -Z_\alpha$
$\mu \neq \mu_0$	$ Z  > Z_{\alpha/2}$

Вероятность ошибки первого рода равна

$$P_{H_0}(|Z| > Z_{\alpha/2}) = P_{H_0}(Z > Z_{\alpha/2}) + P_{H_0}(Z < -Z_{\alpha/2}) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.$$



*Проверка гипотезы об определенном значении генерального среднего при неизвестной дисперсии.* Если генеральная дисперсия  $\sigma^2$  неизвестна, в качестве статистики критерия используется статистика

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}},$$

где

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

В соответствии с видом альтернативы  $H_1$  критическая область  $K$  выбирается по таблице

$H_1$	$K$
$\mu > \mu_0$	$t > t_\alpha(n-1)$
$\mu < \mu_0$	$t < -t_\alpha(n-1)$
$\mu \neq \mu_0$	$ t  > t_{\alpha/2}(n-1)$

в которой  $t_\alpha(n-1)$  –  $100\alpha$ -процентная точка распределения Стьюдента с  $n-1$  степенями свободы.

*Пример.* Цена акции  $S_t$  в момент времени  $t = 0, 1, \dots, T$  задается моделью случайного блуждания:

$$\ln S_t = \ln S_0 + \sum_{i=1}^t X_i,$$

где

$$X_i = \ln \frac{S_i}{S_{i-1}}$$

– логарифмическая доходность акции за период  $[i-1, i]$ , а  $X_1, \dots, X_T$  в совокупности образуют выборку из  $N(\mu, \sigma^2)$ . Предположим, что за период времени  $T = 10$  средняя логарифмическая доходность составила  $\bar{x} = 0,01$  при дисперсии  $s_x^2 = 0,02^2$ . Требуется при 5%-ном уровне значимости проверить гипотезу  $H_0: \mu = 0$  против гипотезы  $H_1: \mu > 0$ .

*Решение.* Находим значение статистики

$$t = \frac{0,01 - 0}{0,03/3} = 1,5.$$

Затем по табл. процентных точек распределения Стьюдента находим критическое значение  $t_{0,05}(9) = 1,833$ . Поскольку  $t < t_{0,05}(9)$ , наблюдаемое значение статистики не принадлежит критической области. Следовательно, гипотеза  $\mu = 0$  принимается.

*Проверка гипотезы об определенном значении генеральной дисперсии.*  
Способ проверки гипотезы  $H_0: \sigma = \sigma_0$  зависит от того, известно или нет генеральное среднее  $\mu$ . Если  $\mu$  известно, то используется статистика

$$\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{ns_0^2}{\sigma_0^2},$$

а критическая область  $K$  выбирается по следующей таблице.

$H_1$	$K$
$\sigma > \sigma_0$	$\chi^2 > \chi_{\alpha}^2(m)$
$\sigma < \sigma_0$	$\chi^2 < \chi_{1-\alpha}^2(m)$
$\sigma \neq \sigma_0$	$\{\chi^2 < \chi_{1-\alpha/2}^2(m)\} \cup \{\chi^2 > \chi_{\alpha/2}^2(m)\}$

в которой  $m = n - 1$  – объем выборки,  $\chi_{\alpha}^2(m)$  –  $100\alpha$ -процентная точка распределения  $\chi^2$  с  $m$  степенями свободы.

Если же генеральное среднее неизвестно, используется статистика

$$\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)s^2}{\sigma_0^2}.$$

При этом критическая область определяется с другим числом степеней свободы  $m = n - 1$ .

*Пример.* Используя данные из предыдущего примера, проверить при уровне значимости  $\alpha = 0,05$  гипотезу  $H_0: \sigma = 0,01$  против гипотезы  $H_1: \sigma > 0,01$ .

*Решение.* Вычисляем наблюдаемое значение статистики критерия:

$$\chi^2 = \frac{9 \cdot 0,02^2}{0,012} = 36.$$

Затем по таблице процентных точек распределения  $\chi^2$  находим критическое значение  $\chi_{0,05}^2(9) = 16,92$ . Поскольку значение статистики  $\chi^2$  принадлежит критической области,  $\chi^2 > \chi_{0,05}^2(9)$ , основная гипотеза отвергается в пользу гипотезы  $\sigma > 0,01$ .

### *Сравнение генеральных средних нормальных распределений*

Пусть  $\vec{X} = (X_1, \dots, X_m)$  – выборка из  $N(\mu_x, \sigma_x^2)$ , а  $\vec{Y} = (Y_1, \dots, Y_n)$  – выборка из нормального распределения  $N(\mu_y, \sigma_y^2)$ . Далее считаем выборки  $\vec{X}$  и  $\vec{Y}$  независимыми, что означает независимость в совокупности  $m + n$  случайных величин  $X_1, \dots, X_m, Y_1, \dots, Y_n$ .

*Сравнение генеральных средних при известной дисперсии.* Дисперсии  $\sigma_x^2$  и  $\sigma_y^2$  известны, генеральные средние  $\mu_x$  и  $\mu_y$  неизвестны. Рассматривается основная гипотеза  $H_0: \mu_x = \mu_y$ . Альтернативная гипотеза чаще всего имеет вид:

$$1) H_1: \mu_x > \mu_y;$$

$$2) H_1: \mu_x < \mu_y;$$

$$3) H_1: \mu_x \neq \mu_y.$$

При проверке  $H_0$  против  $H_1$  вида 1, 2 или 3 используется одна и та же статистика:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}}$$

*Утверждение.* Если верна  $H_0$ , то  $Z \sim N(0,1)$ .

*Доказательство.* Статистика  $Z$  является линейной функцией от нормально распределенных, независимых случайных величин  $X_1, \dots, X_m, Y_1, \dots, Y_n$ . Следовательно,  $Z \sim N(\mu_z, \sigma_z^2)$ . Нетрудно видеть, что  $\mu_z =$

$C^{-1}(E(\bar{X}) - E(\bar{Y}))$ , где  $C = \sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}$ . При верной  $H_0$  имеем

$$E(\bar{X}) - E(\bar{Y}) = \mu_x - \mu_y = 0.$$

Следовательно,  $\mu_z = 0$ . Далее находим

$$D(Z) = C^{-2} \left( \frac{1}{m} \sigma_x^2 + \frac{1}{n} \sigma_y^2 \right) = 1,$$

что и завершает доказательство.

Пусть  $Z_\alpha, Z_{\alpha/2}$  – соответствующие процентные точки стандартного нормального распределения  $N(0,1)$ . При проверке  $H_0$  против  $H_1$  применяется критерий с критической областью, определяемой по следующей таблице:

$H_1$	$K$
1) $\mu_x > \mu_y$	$Z > Z_\alpha$
2) $\mu_x < \mu_y$	$Z < -Z_\alpha$
3) $\mu_x \neq \mu_y$	$ Z  > Z_{\alpha/2}$

*Утверждение 2.* Критерий с критической областью 1 – 3 имеет уровень значимости  $\alpha$ .

*Доказательство.* Утверждение для критической области 1 следует из утверждения 1 и определения процентной точки.

Если верна  $H_0$ , то  $Z \sim -Z$  и

$$P_{H_0}(Z < -Z_\alpha) = P_{H_0}(-Z > Z_\alpha) = P_{H_0}(Z > Z_\alpha) = \alpha,$$

что доказывает утверждение для критической области 2.

Рассмотрим теперь случай критической области 3. Событие  $|Z| > Z_{\alpha/2}$  является суммой несовместных событий  $Z > Z_{\alpha/2}$  и  $Z < -Z_{\alpha/2}$ . Отсюда следует, что

$$P_{H_0}(|Z| > Z_{\alpha/2}) = P_{H_0}(Z > Z_{\alpha/2}) + P_{H_0}(Z < -Z_{\alpha/2}) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.$$

Можно доказать, что во всех трех случаях при фиксированных  $\mu_x$  и  $\mu_y$

$$\lim_{\substack{m \rightarrow \infty \\ n \rightarrow \infty}} P_{H_1}(\text{отвергается } H_0) = 1.$$

Таким образом, данные критерии имеют высокую мощность, по крайней мере, для выборок достаточно большого объема.

*Сравнение генеральных средних при неизвестной дисперсии.* Теперь предположим, что обе генеральные дисперсии неизвестны, но одинаковы,  $\sigma_x^2 = \sigma_y^2 = \sigma^2$ . С учетом этого равенства имеем

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}} = \frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

Поскольку значение  $\sigma$  неизвестно, попробуем заменить  $\sigma$  на корень из несмещенной оценки дисперсии. В данном случае имеется, по крайней мере, две такие оценки:

$$s_x^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$$

и

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Оказывается, что вместо того, чтобы использовать одну из них, лучше воспользоваться линейной комбинацией

$$s^2 = \frac{m-1}{m+n-2} s_x^2 + \frac{n-1}{m+n-2} s_y^2.$$

Очевидно проверяется, что  $s^2$  – несмещенная оценка  $\sigma^2$ . Кроме того, можно доказать, что

$$\frac{(m-1)s_x^2}{\sigma^2} = \frac{(X_1 - \bar{X})^2 + \dots + (X_m - \bar{X})^2}{\sigma^2} \sim \chi^2(m-1)$$

и

$$\frac{(n-1)s_y^2}{\sigma^2} = \frac{(Y_1 - \bar{Y})^2 + \dots + (Y_n - \bar{Y})^2}{\sigma^2} \sim \chi^2(n-1).$$

Отсюда с учетом независимости  $\vec{X}$  и  $\vec{Y}$  следует, что

$$\frac{(m-1)s_x^2 + (n-1)s_y^2}{\sigma^2} \sim \chi^2(m+n-2).$$

Заменив  $\sigma$  на  $s$  в записи статистики, получим новую статистику

$$T = \frac{\bar{X} - \bar{Y}}{s \sqrt{\frac{1}{m} + \frac{1}{n}}}.$$

Примем без доказательства следующую теорему.

*Теорема.* Если верна  $H_0$ , то  $T \sim t(m + n - 2)$ , где  $t(m + n - 2)$  – распределение Стьюдента с  $m + n - 2$  степенями свободы.

В силу данной теоремы статистику  $T$  можно использовать для построения критерия по проверке  $H_0$ . Как и раньше, рассмотрим три основных случая альтернативной гипотезы  $H_1$ , в соответствии с которыми выбирается критическая область  $K$  того или иного вида.

	$H_1$	$K$
1)	$\mu_x > \mu_y$	$T > t_\alpha(m + n - 2)$
2)	$\mu_x < \mu_y$	$T < -t_\alpha(m + n - 2)$
3)	$\mu_x \neq \mu_y$	$ T  > t_{\alpha/2}(m + n - 2)$

*Утверждение 3.* Статистический критерий с критической областью в любом из случаев 1 – 3 имеет уровень значимости  $\alpha$ .

**P-значение критерия.** Описанная выше процедура проверки статистической гипотезы позволяет принять или отклонить основную гипотезу путем сравнения наблюдаемого значения статистики критерия с критическим значением, связанным с данным уровнем значимости, однако если уровень значимости будет другим, то придется вновь вычислить соответствующее критическое значение. Вводимое ниже понятие, ставшее популярным в связи с широким распространением статистических программ, позволяет решить вопрос о принятии или отклонении основной гипотезы одновременно для всех уровней значимости без вычисления критических значений.

*Определение.* Для фиксированной реализации  $\vec{x}$  случайной выборки  $\vec{X} = (X_1, \dots, X_n)$  **P-значением (P-value)** статистического критерия называется такое число  $PV(\vec{x})$ , что  $PV(\vec{x}) \geq \alpha$  для любого уровня

значимости  $\alpha$ , при котором гипотеза  $H_0$  принимается, и  $PV(\vec{x}) \leq \alpha$  – для любого уровня значимости  $\alpha$ , при котором гипотеза  $H_0$  отвергается.

Предположим, что  $P$ -значение  $PV(\vec{x})$  уже каким-либо способом найдено. Тогда решение о принятии (отклонении)  $H_0$  для заданного  $\alpha$  осуществляется на основе следующего простого правила: *если  $\alpha > PV(\vec{x})$ , то гипотеза  $H_0$  отвергается, а если  $\alpha < PV(\vec{x})$  – гипотеза  $H_0$  принимается.*

Рассмотрим отдельно случай  $\alpha = PV(\vec{x})$ . Как правило, критическую область можно представить в виде

$$K_\alpha = \{t(\vec{x}) > c(\alpha)\},$$

где  $c(\alpha)$  – непрерывная убывающая функция. Как нетрудно видеть, в этом случае  $PV(\vec{x}) = c^{-1}(t(\vec{x}))$  и для  $\alpha = PV(\vec{x})$  имеет место равенство

$$t(\vec{x}) = c(\alpha),$$

означающее, что  $H_0$  принимается. Отсюда уже легко получить широко применяемую формулу:

$$PV(\vec{x}) = P_{H_0}(t(\vec{X}) > t(\vec{x})).$$

Совершенно аналогично в случае

$$K_\alpha = \{t(\vec{x}) < c(\alpha)\},$$

где  $c(\alpha)$  – непрерывная возрастающая функция,  $P$ -значение удовлетворяет соотношению  $PV(\vec{x}) = P_{H_0}(t(\vec{X}) < t(\vec{x}))$ .

*Пример.* Пусть  $\vec{x}$  – реализация случайной выборки  $X_1, \dots, X_{100}$  из распределения  $N(\mu, 5^2)$ , такая, что  $\bar{x} = 3$ . Требуется при уровне значимости  $\alpha = 0,05$  проверить гипотезу  $H_0: \mu = 0$  против альтернативы  $H_1: \mu > 0$ . Необходимо также вычислить  $P$ -значение критерия.

*Решение.* Для проверки гипотезы используем критическую область вида  $K = \{Z > Z_\alpha\}$ , где  $Z = \bar{x}/5$ ,  $Z_{0,05} = \Phi^{-1}(0,45) = 1,645$ . Поскольку наблюдаемое значение статистики  $Z_{\text{набл}} = 1,8 > 1,645$ , гипотеза  $H_0$  отвергается.

Прежде чем найти  $P$ -значение, заметим, что в случае  $Z$ -критерия формулу для нахождения  $P$ -значения можно представить в виде

$$P - value = P(Z > Z_{\text{набл}}) = \frac{1}{2} - \Phi(Z_{\text{набл}}).$$

Под  $Z$ -критерием здесь понимается любой критерий, статистика которого  $Z \sim N(0,1)$  при верной  $H_0$ . С учетом предыдущего выражения находим

$$P - value = 0,5 - \Phi(1,8) = 0,0359.$$

Следовательно, при любом уровне значимости  $\alpha > 0,0359$  нулевая гипотеза отвергается, что вполне согласуется с приведенной раньше проверкой при заданном  $\alpha = 0,05$ .

На этом примере видно, что проверка гипотезы при помощи  $P$ -значения более информативна, нежели традиционная проверка с помощью критического значения. Тем не менее выбор того или иного способа проверки зависит прежде всего от наличия соответствующих таблиц (или компьютерных программ).

Может сложиться впечатление, что  $P$ -значение является своеобразной характеристикой адекватности  $H_0$ , однако это не совсем так. Нетрудно доказать, что при верной основной гипотезе  $P$ -значение равномерно распределено на отрезке  $[0,1]$ . Поэтому вероятность получить малое  $P$ -значение, скажем, меньше  $\alpha$ , равно вероятности получить соответствующее большое значение (больше  $1 - \alpha$ ). Тем не менее если  $H_0$  не верна, наблюдаемые  $P$ -значения (при достаточно высокой мощности критерия) концентрируются около нуля.

**28. Критерии согласия. Проверка гипотезы о соответствии эмпирических данных теоретическому закону с данной функцией распределения по критерию Пирсона. Проверка гипотезы об однородности нескольких выборок по критериям Пирсона и Колмогорова – Смирнова.**

*Проверка простой гипотезы по критерию Пирсона (общая схема)*



Рассматривается сложный опыт, выполняемый по схеме повторных независимых испытаний. Пусть  $n$  — число испытаний,  $(\Omega, \mathcal{F}, P)$  — вероятностное пространство, соответствующее одному испытанию. Тогда сложному опыту соответствует вероятностное пространство  $(\Omega^n, \mathcal{F}^n, P^n)$ , где:

- $\Omega^n = \underbrace{\Omega \times \dots \times \Omega}_n$  — декартова степень  $\Omega$ ;
- $\mathcal{F}^n$  — минимальная  $\sigma$ -алгебра, содержащая все события  $A_1 \times \dots \times A_n$ ,  $A_i \in \mathcal{F}$ ;
- $P^n$  — вероятностная мера на  $\mathcal{F}^n$ , такая, что  $P^n(A_1 \times \dots \times A_n) = P(A_1) \cdot \dots \cdot P(A_n)$ .

Пусть  $A_1, \dots, A_l \in \mathcal{F}$  — полная группа событий (эквивалентно, разбиение  $\Omega$ ). Фиксируем набор положительных чисел  $p_i > 0$ , таких, что  $p_1 + \dots + p_l = 1$ . В качестве основной гипотезы примем гипотезу, утверждающую, что события  $A_1, \dots, A_l$  имеют известные вероятности,

$$H_0: P(A_i) = p_i, i = 1, \dots, l.$$

Число испытаний, в которых событие  $A_i$  наступило, называется *частотой* этого события и обозначается, как правило,  $n_i, i = 1, \dots, l$ . Поскольку  $A_1, \dots, A_l$  — полная группа событий, имеем

$$n_1 + \dots + n_l = n.$$

*Относительной частотой* события  $A_i$  называется отношение

$$\hat{p}_i = \frac{n_i}{n}.$$

Фиксируем набор положительных чисел:  $b_1 > 0, \dots, b_l > 0$  и рассмотрим статистику вида

$$t(n_1, \dots, n_l) = \sum_{i=1}^l b_i (\hat{p}_i - p_i)^2.$$

Так как при верной  $H_0$  частота  $\hat{p}_i$  — состоятельная оценка  $p_i$ ,

$$\text{plim}_{n \rightarrow \infty} \hat{p}_i = p_i,$$

то при верной  $H_0$  статистика  $t(n_1, \dots, n_l)$  сходится по вероятности к нулю,

$$\text{plim}_{n \rightarrow \infty} t(n_1, \dots, n_l) = 0.$$

Следовательно, если  $t(n_1, \dots, n_l)$  принимает большое значение, то это можно расценивать как указание на ложность основной гипотезы. Рассмотрим критерий с критической областью вида

$$K = \{t(n_1, \dots, n_l) > c_\alpha\}.$$

Если бы распределение случайной величины было известно, то взяв в качестве  $c_\alpha$  квантиль этого распределения уровня  $(1 - \alpha)$ , получили бы критерий с уровнем значимости  $\alpha$ . Однако точное распределение статистики  $t(n_1, \dots, n_l)$  зависит от неизвестных вероятностей  $p_i$  и построить критерий с точным уровнем значимости  $\alpha$  таким образом не удастся. Вместе с тем, как заметил К. Пирсон, при подходящем выборе коэффициентов  $b_1, \dots, b_l$  предельное распределение статистики  $t(n_1, \dots, n_l)$  существует и не зависит от  $p_1, \dots, p_l$ .

*Теорема Пирсона.* Если  $H_0$  верна и  $n \rightarrow \infty$ , то

$$Law \left( \sum_{i=1}^l \frac{n}{p_i} (\hat{p}_i - p_i)^2 \right) \rightarrow \chi^2(l-1).$$

*Статистика хи-квадрат* задается формулой

$$\chi^2 = \sum_{i=1}^l \frac{n}{p_i} (\hat{p}_i - p_i)^2 = \sum_{i=1}^l \frac{n}{p_i} \left( \frac{n_i}{n} - \frac{np_i}{n} \right)^2 = \sum_{i=1}^l \frac{(n_i - np_i)^2}{np_i}.$$

Таким образом, вследствие теоремы Пирсона критическая область

$$\{\chi^2 > \chi_\alpha^2(l-1)\}$$

задает статистический критерий с асимптотическим уровнем значимости  $\alpha$ , который и называется *критерием согласия (хи-квадрат) Пирсона*.

Предположим, что гипотеза  $H_0$  верна. Тогда частота  $n_i$  распределена по биномиальному закону,

$$n_i \sim \text{Bin}(n, p_i).$$

Таким образом,  $E(n_i) = np_i$ , вследствие чего  $np_i$  называется *ожидаемой частотой* события  $A_i$ .

Считается, что если все ожидаемые частоты  $np_i \geq 5$  и  $n \geq 50$ , то точный уровень значимости критерия Пирсона достаточно близок к  $\alpha$ . Именно это условие и приводится в большинстве указаний по применению критерия Пирсона.

*Проверка сложной параметрической гипотезы по критерию Пирсона.* Случай, когда предполагаемые вероятности  $p_1, \dots, p_l$  определены однозначно (т.е.  $H_0$  — простая гипотеза) встречается редко. Гораздо чаще вероятности  $p_1, \dots, p_l$  зависят от одного или нескольких неизвестных параметров  $\theta_1, \dots, \theta_r$ .

Таким образом, сложная гипотеза  $H_0$  утверждает, что вероятности событий  $A_1, \dots, A_l$  могут быть представлены в виде заданных функций:

$$H_0: P(A_i) = p_i(\theta_1, \dots, \theta_r).$$

Для проверки сложной гипотезы  $H_0$  по критерию Пирсона применяется следующий прием. Сначала находятся оценки параметров  $\hat{\theta}_1, \dots, \hat{\theta}_r$ , а через них — и оценки вероятностей:

$$\hat{p}_i = p_i(\hat{\theta}_1, \dots, \hat{\theta}_r), i = 1, \dots, l.$$

Затем определяется статистика хи-квадрат вида

$$\hat{\chi}^2 = \sum_{i=1}^l \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}.$$

Произведение  $n\hat{p}_i$  называется либо *теоретической частотой*, либо (по аналогии с  $np_i$ ) — *ожидаемой частотой* события  $A_i$ .

Можно доказать, что при  $n \rightarrow \infty$  и верной  $H_0$  распределение статистики  $\hat{\chi}^2$ , как и раньше, стремится к распределению хи-квадрат, но уже с другим числом степеней свободы,

$$\text{Law} \left( \sum_{i=1}^l \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} \right) \rightarrow \chi^2(l - r - 1).$$

Таким образом, критерий с критической областью

$$K = \{\hat{\chi}^2 > \chi_\alpha^2(l - r - 1)\}$$

при  $n \gg 1$  имеет уровень значимости, близкий к  $\alpha$ . Этот критерий также называется критерием Пирсона и применяется, когда все частоты  $n\hat{p}_i \geq 5$  и объем выборки  $n \geq 50$ .

*Пример проверки сложной гипотезы по критерию Пирсона.*

Допустимое распределение  $Law(X)$  задается таблицей

1	2	3	4	5
$p_1$	$p_2$	$p_3$	$p_4$	$p_5$

Здесь  $p_1, \dots, p_5$  — неизвестные вероятности. Гипотеза  $H_0$  состоит в том, что распределение  $Law(X)$  симметрично относительно 3:

$$X \sim 6 - X.$$

Эту гипотезу можно записать в виде

$$H_0: p_1 = p_5, p_2 = p_4.$$

Информация о выборке представлена в виде таблицы частот

1	2	3	4	5
$n_1$	$n_2$	$n_3$	$n_4$	$n_5$

Проверим  $H_0$  по критерию Пирсона. Находим оценки:

$$\hat{p}_1 = \hat{p}_5 = \frac{n_1 + n_5}{2n},$$

$$\hat{p}_2 = \hat{p}_4 = \frac{n_2 + n_4}{2n},$$

$$\hat{p}_3 = 1 - 2\hat{p}_1 - 2\hat{p}_2,$$

где

$$n = n_1 + \dots + n_5.$$

Затем находим статистику

$$\hat{\chi}^2 = \sum_{i=1}^5 \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = \frac{(n_1 - n_5)^2}{2n\hat{p}_1} + \frac{(n_2 - n_4)^2}{2n\hat{p}_2}.$$

Критическая область имеет вид

$$\hat{\chi}^2 > \chi_\alpha^2(2).$$

*Проверка однородности по критериям Смирнова и Колмогорова–Смирнова.* Предположим, что имеются только две выборки  $\vec{X}_1^m$  и  $\vec{X}_2^n$ , объемы которых равны соответственно  $m$  и  $n$ . Пусть  $\vec{x}_1^m$  – реализация  $\vec{X}_1^m$  и  $\vec{x}_2^n$  – реализация  $\vec{X}_2^n$ . Эмпирические функции распределения обозначим  $\hat{F}_1(x; \vec{X}_1^m)$ ,  $\hat{F}_2(x; \vec{X}_2^n)$  для случайных выборок  $\vec{X}_1^m$ ,  $\vec{X}_2^n$  и  $\hat{F}_1(x; \vec{x}_1^m)$ ,  $\hat{F}_2(x; \vec{x}_2^n)$  – для их реализаций. Определим расстояние между функциями  $\hat{F}_1(x) = \hat{F}_1(x; \vec{x}_1^m)$  и  $\hat{F}_2(x) = \hat{F}_2(x; \vec{x}_2^n)$  формулой

$$d(\vec{x}_1^m, \vec{x}_2^n) = \sup_{x \in R} |\hat{F}_1(x) - \hat{F}_2(x)|.$$

Пусть  $X$  – конечное множество, образованное всеми различными компонентами выборок  $\vec{x}_1^m$  и  $\vec{x}_2^n$ . Нетрудно видеть, что

$$d(\vec{x}_1^m, \vec{x}_2^n) = \max_{x \in X} |\hat{F}_1(x) - \hat{F}_2(x)|,$$

поэтому точная верхняя грань в формуле всегда достигается.

Для случайных выборок расстояние  $d(\vec{X}_1^m, \vec{X}_2^n)$  между  $\hat{F}_1(x) = \hat{F}_1(x; \vec{X}_1^m)$  и  $\hat{F}_2(x) = \hat{F}_2(x; \vec{X}_2^n)$  определяется аналогично.

Предположим, что выборки  $\vec{X}_1^m$  и  $\vec{X}_2^n$  однородны и, кроме того, извлечены из распределения с непрерывной функцией  $F(x)$ . Оказывается, что в этом случае распределение случайной величины  $d(\vec{X}_1^m, \vec{X}_2^n)$  не зависит от  $F(x)$ .

Процентные точки  $d_\alpha(m, n)$  распределения  $d(\vec{Y}_1^m, \vec{Y}_2^n)$ , как обычно, определяются соотношением  $P(d(\vec{Y}_1^m, \vec{Y}_2^n) > d_\alpha(m, n)) = \alpha$ . Для проверки гипотезы однородности двух выборок, полученных из непрерывных распределений, применяется *критерий Н.В. Смирнова*, критическая область которого задается неравенством  $d(\vec{x}_1^m, \vec{x}_2^n) > d_\alpha(m, n)$ .

Таблицы критических значений  $d_\alpha(m, n)$  при небольших  $m$  и  $n$  можно найти в некоторых справочниках.

При больших  $m$  и  $n$  можно воспользоваться результатом *Н. В. Смирнова*, который доказал, что при  $m, n \rightarrow \infty$  (и непрерывной  $F(x)$ )

$$P \left\{ \sqrt{\frac{mn}{m+n}} d(\vec{X}_1^m, \vec{X}_2^n) < u \right\} \rightarrow K(u),$$

где  $K(u)$  – функция предельного распределения Колмогорова. Соответствующий критерий называется *критерием Колмогорова–Смирнова* и задается критической областью

$$\sqrt{\frac{mn}{m+n}} d(\vec{x}_1^m, \vec{x}_2^n) > u_\alpha,$$

где  $u_\alpha$  – корень уравнения  $K(u) = 1 - \alpha$ . Критерий Колмогорова–Смирнова имеет асимптотический уровень значимости  $\alpha$ , т.е. его фактическое значение достигает  $\alpha$  лишь в пределе при  $m, n \rightarrow \infty$ . На практике критерий Колмогорова–Смирнова применяется, если объемы обеих выборок не меньше 50.

## 29. Сравнение нескольких средних методом дисперсионного анализа.

Пусть  $\vec{X}_i = (X_{i1}, \dots, X_{in_i})$  – выборка объема  $n_i$  из  $N(\mu_i, \sigma^2)$ , где  $i = 1, \dots, k$ . Предположим также, что  $n = n_1 + \dots + n_k$  случайных величин

$$X_{11}, \dots, X_{1n_1}; X_{21}, \dots, X_{2n_2}; \dots; X_{k1}, \dots, X_{kn_k}$$

независимы в совокупности. Таким образом, выборки  $\vec{X}_1, \dots, \vec{X}_k$  независимы и получены из нормальных распределений с одинаковой дисперсией  $\sigma^2$  и, возможно, различными средними  $\mu_1, \dots, \mu_k$ . Гипотеза о равенстве всех средних одновременно записывается как

$$H_0: \mu_1 = \dots = \mu_k,$$

а альтернативная гипотеза – как

$$H_1: (\exists ij) \mu_i \neq \mu_j.$$

Заметим, что при верной  $H_0$  выборки  $\vec{X}_1, \dots, \vec{X}_k$  являются однородными, поскольку в этом случае генеральные распределения совпадают:

$$N(\mu_1, \sigma^2) = \dots = N(\mu_k, \sigma^2).$$

Следовательно, гипотеза  $H_0$  при сделанных базисных предположениях может рассматриваться как параметрический аналог рассмотренных ранее непараметрических гипотез однородности.

Рассмотрим объединенную выборку объема  $n = n_1 + \dots + n_k$ :

$$\vec{X} = (X_{11}, \dots, X_{1n_1}; \dots; X_{k1}, \dots, X_{kn_k}).$$

Интерпретируя выборки  $\vec{X}_1, \dots, \vec{X}_k$  как группы, на которые разбита совокупность  $\vec{X}$ , введем обозначения:

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

– выборочное среднее в  $i$ -й совокупности;

$$\hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

– выборочная дисперсия в той же выборке;

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k \bar{X}_i n_i = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}$$

– выборочное среднее в объединенной выборке  $\vec{X}$ ;

$$\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^k \hat{\sigma}_i^2 n_i$$

– средняя групповая дисперсия;

$$\delta^2 = \frac{1}{n} \sum_{i=1}^k (\bar{X}_i - \bar{X})^2 n_i$$

– межгрупповая дисперсия;

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$

– выборочная дисперсия признака в объединенной выборке  $\vec{X}$ .

Выборочную дисперсию  $\hat{\sigma}^2$  можно представить в виде суммы  $\hat{\sigma}^2 = \bar{\sigma}^2 + \delta^2$ , где первое слагаемое  $\bar{\sigma}^2$  характеризует среднюю изменчивость

признака в каждой выборке  $\vec{X}_1, \dots, \vec{X}_k$ , а второе слагаемое  $\delta^2$  характеризует разброс выборочных средних  $\bar{X}_1, \dots, \bar{X}_k$ .

Критерий проверки  $H_0$  против  $H_1$  основан на следующей теореме.

*Теорема.* Предположим, что верна гипотеза  $H_0$ . Тогда:

1)  $\frac{n\delta^2}{\sigma^2} \sim \chi^2(k-1)$ ; 2)  $\frac{n\bar{\sigma}^2}{\sigma^2} \sim \chi^2(n-k)$ ; 3)  $\delta^2$  и  $\bar{\sigma}^2$  независимы.

Определим так называемое  $F$ -отношение:

$$F = \frac{\frac{1}{k-1} \sum_{i=1}^k (\bar{X}_i - \bar{X})^2 n_i}{\frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}.$$

Статистику  $F$  можно представить также в виде

$$F = \frac{(n\delta^2)/(k-1)}{(n\bar{\sigma}^2)/(n-k)} = \frac{(n\delta^2/\sigma^2)/(k-1)}{(n\bar{\sigma}^2/\sigma^2)/(n-k)}.$$

Тогда можно показать, что

$$F \sim F(k-1, n-k),$$

где  $F(k-1, n-k)$  – распределение Фишера с  $k-1$  и  $n-k$  степенями свободы. Для проверки  $H_0$  с уровнем значимости  $\alpha$  применяется критерий с критической областью  $F > F_\alpha(k-1, n-k)$ , где  $F_\alpha(k-1, n-k)$  –  $100\alpha$  процентная точка распределения  $F(k-1, n-k)$ .

**30. Денежные потоки и их числовые характеристики. Чистая приведённая стоимость. Ренты. Объединение и замена рент. Инвестиционные потоки. Внутренняя норма доходности.**

**Мгновенным финансовым событием** называется точка плоскости время – деньги, т.е. пара  $(t, C)$ , состоящая из момента времени  $t$  и значения денежной суммы  $C$ .

В качестве финансового события может выступать поступление или выплата суммы  $C$  в момент времени  $t$ . Величина  $C$  может быть как положительной, так и отрицательной. Например, поступления указываются «с плюсом», а выплаты – «с минусом».

Последовательность (конечная или бесконечная) финансовых событий

$$(t_0, C_0), (t_1, C_1), (t_2, C_2), \dots, (t_n, C_n), \dots,$$



называется **дискретным финансовым потоком**. Предполагается, что  $t_0 < t_1 < t_2 < \dots$ . В случае бесконечного потока мы предполагаем дополнительно, что  $t_k$  неограниченно возрастает с ростом  $k$ .

Финансовый поток можно охарактеризовать так называемой **платежной функцией**, которая каждому моменту времени  $t$  сопоставляет значение денежной суммы  $C(t)$  так, что  $C(t_k) = C_k$ , и  $C(t) = 0$ , если  $t$  не совпадает ни с одним из моментов  $t_k$ ,  $k = 0, 1, \dots$ .

С учетом этого естественным образом определяются операции *умножения потока на число и сложения потоков*. Если  $CF$  — финансовый поток с платежной функцией  $C(t)$ , а  $\lambda$  — произвольное число, то  $\lambda \cdot CF$  — это финансовый поток с платежной функцией  $\lambda \cdot C(t)$ . Таким образом, поток  $\lambda \cdot CF$  получается из потока  $CF$  умножением каждой суммы на число  $\lambda$ . Если  $CF^{(1)}$  и  $CF^{(2)}$  — финансовые потоки с платежными функциями соответственно  $C^{(1)}(t)$  и  $C^{(2)}(t)$ , то  $CF^{(1)} + CF^{(2)}$  — это финансовый поток с платежной функцией  $C^{(1)}(t) + C^{(2)}(t)$ . Относительно этих операций *множество финансовых потоков образует векторное пространство*.

Пусть

$$CF = \{(t_0, C_0), (t_1, C_1), (t_2, C_2), \dots, (t_n, C_n), \dots\} -$$

денежный поток. Сумма всех платежей денежного потока, приведенных к некоторому моменту времени  $t$ , называется *текущим*, или **приведенным значением** потока (в момент времени  $t$ ), и обозначается  $PV_t(CF, i)$ , или просто  $PV_t$ . Таким образом,

$$PV_t = \frac{C_0}{(1+i)^{t_0-t}} + \frac{C_1}{(1+i)^{t_1-t}} + \dots$$

В случае бесконечного потока текущее значение считается определенным лишь в том случае, когда ряд в правой части сходится. Если  $t_0 = 0$ , текущее значение потока в начальный момент времени обозначается просто  $PV$ :

$$PV = C_0 + \frac{C_1}{(1+i)^{t_1}} + \frac{C_2}{(1+i)^{t_2}} + \dots$$

Эту величину называют также **чистой приведенной стоимостью** денежного потока.

Под **рентой** понимают поток положительных платежей, разделенных равными интервалами времени. Промежуток времени между двумя последовательными платежами называют *периодом* ренты. Считается, что каждый платеж производится либо в конце соответствующего ему периода, либо в начале. В первом случае ренту называют *обыкновенной*, а также *подрасчетной*, или рентой *постнумерандо*, во втором – *авансированной*, или *пренумерандо*. Ренты с конечным числом платежей называют *конечными*. Промежуток времени между началом первого периода и окончанием последнего называется *сроком* конечной ренты. Ренты с бесконечным числом платежей называют *бесконечными*, или *вечными*. Если все платежи равны между собой, ренту называют *постоянной*. В случае, когда период постоянной ренты равен одному году (платежи производятся раз в год), ренту называют *годовой*, или **аннуитетом**. Более формально, аннуитет – это поток платежей вида

$$\{(0, 0), (1, R), (2, R), \dots, (n, R)\}.$$

Найдем текущую стоимость  $A$  этого потока платежей относительно процентной ставки  $i$ . В соответствии с определением

$$A = \frac{R}{1+i} + \frac{R}{(1+i)^2} + \dots + \frac{R}{(1+i)^n} = R \frac{1-(1+i)^{-n}}{i}.$$

Множитель при  $R$  в правой части **Ошибка! Источник ссылки не найден.** называют **коэффициентом приведения** годовой ренты. Объединение и замена рент выполняются путем приравнивания приведенной стоимости. Пусть, например, имеются два аннуитета с годовым платежом  $R_k$ , числом платежей  $n_k$  и ставкой дисконтирования  $i_k$ ,  $i = 1, 2$ . Чтобы заменить объединение этих рент аннуитетом аннуитета с

годовым платежом  $R_k$ , числом платежей  $n_k$  и ставкой дисконтирования  $i_k$ , нужно решить уравнение вида  $A_1 + A_2 = A$ , т.е. уравнение

$$R_1 \frac{1-(1+i_1)^{-n_1}}{i_1} + R_2 \frac{1-(1+i_2)^{-n_2}}{i_2} = R \frac{1-(1+i)^{-n}}{i}.$$

При объединении рент, как правило решают полученное уравнение относительно одной из трех величин  $R, i, n$  при известных остальных. Наиболее распространенные виды замены рент — выкуп ренты и рассрочка платежей — сводятся к решению указанного уравнения при равном нулю втором слагаемом в левой части.

Денежный поток

$$CF = \{(0, -K), (t_1, C_1), (t_2, C_2), \dots, (t_n, C_n)\}$$

такой, что  $K > 0$ ,  $C_k \geq 0$  при  $k = 1, 2, \dots, n$ ,  $0 < t_1 < \dots < t_n$  и  $C_n > 0$  характеризует типичный **инвестиционный процесс**. Величина  $C_k$  представляет собой баланс инвестиционных затрат и чистого дохода за период  $k$ , актуализированный на конец этого периода, величина  $K$  — начальные инвестиции. Величину

$$NPV = -K + \frac{C_1}{(1+i)^{t_1}} + \frac{C_2}{(1+i)^{t_2}} + \dots + \frac{C_n}{(1+i)^{t_n}}$$

в «инвестиционном» контексте называют **чистым приведенным доходом** (или **чистым дисконтированным доходом**).

Если чистый приведенный доход имеет отрицательное значение, это означает, что доходы не окупают затрат при принятой норме доходности  $i$ . Условие  $NPV > 0$  является **критерием эффективности инвестиционного проекта**.

При  $i > -1$  чистый приведенный доход  $NPV = NPV(i)$  является убывающей функцией ставки приведения  $i$ . При  $i$ , стремящемся к  $-1$  (справа), каждое слагаемое  $\frac{C_k}{(1+i)^{t_k}}$  с положительным  $C_k$  стремится к бесконечности, а, значит, и  $NPV$  стремится к бесконечности. Следовательно,  $NPV(i) > 0$  для  $i$  достаточно близких к  $-1$ . С другой стороны, при

неограниченном росте  $i$  все слагаемые  $\frac{C_k}{(1+i)^{t_k}}$  стремятся к нулю, а  $NPV$  стремится к  $-K$ . Значит,  $NPV(i) < 0$  для достаточно больших  $i$ . Таким образом, при  $i > -1$  непрерывная функция  $NPV(i)$ , убывая, меняет знак с плюса на минус. Следовательно,  $NPV(i)$  обращается в ноль для некоторого  $i = i_0$ . Значение  $i_0$  называют **внутренней нормой доходности** инвестиционного потока платежей. Внутренняя норма доходности служит границей процентных ставок, для которых проект имеет положительную и отрицательную приведенную стоимость: если  $i > i_0$ , то  $NPV(i) < 0$ ; если  $i < i_0$ , то  $NPV(i) > 0$ .

**31. Облигации и их числовые характеристики. Рыночная цена облигации и доходность к погашению. Курс облигации, премия и дисконт. Цена облигации как функция доходности к погашению и срока ее обращения.**

**Облигацией** называют долговое обязательство, выпускаемое *эмитентом* для заимствования денежных средств на открытом рынке. Выпущенная в обращение в момент времени  $t = 0$  облигация характеризуется следующими параметрами:

**дата погашения**  $t = T$ , где  $T$  – время обращения облигации с момента выпуска и **срок погашения**  $n = T - \tau$ , где  $\tau$  – текущая дата;

**номинальная стоимость**  $N$  – сумма денег, которая выплачивается владельцу облигации на дату погашения. Номинальная стоимость обычно указывается на самой облигации;

**купонный доход**  $C$  – постоянные платежи, которые выплачиваются владельцу ежегодно по **купонной ставке**  $c = \frac{C}{N}$ . Если выплаты по купонам не предусмотрены, то такую облигацию называют *бескупонной*.

С каждой облигацией связан поток платежей, состоящий из ежегодной выплаты купонного дохода и выплаты номинальной стоимости на дату погашения. Поэтому в момент  $t = 0$  имеет смысл говорить о

**текущей стоимости**  $P$  облигации. Пусть  $r$  – ставка рефинансирования (процентная ставка), а до погашения облигации осталось ровно  $n$  лет. Тогда имеем

$$P = \sum_{k=1}^n \frac{C}{(1+r)^k} + \frac{N}{(1+r)^n}.$$

Поскольку купонные платежи  $C = cN$  образуют простую ренту,

$$P = cN \frac{1 - (1+r)^{-n}}{r} + N(1+r)^{-n}.$$

После выпуска облигации она поступает на рынок, где свободно продается и покупается по **рыночной цене**  $V$ . Важным понятием, служащим заменой процентной ставки  $r$  в этой ситуации, является **доходность к погашению**  $\rho$ . А именно, если известны рыночная цена облигации  $V$ , ее номинальная стоимость  $N$ , срок погашения  $n$  и купонная ставка  $c$ , то доходность к погашению определяют как решение уравнения

$$V = cN \frac{1 - (1+\rho)^{-n}}{\rho} + N(1+\rho)^{-n}.$$

В общем случае для больших значений  $n$  точное решение этого уравнения затруднительно и для нахождения доходности к погашению используют приближенные формулы. Приведем одну из таких формул:

$$\rho \approx \frac{2(cn + 1 - v)}{v - 1 + n(1 + v)},$$

где  $v = \frac{V}{N}$  – курс облигации.

Цена облигации как функция доходности к погашению является выпуклой. Уменьшение доходности облигации приведет к росту ее рыночной цены на величину большую, чем соответствующее уменьшение рыночной цены при увеличении доходности на ту же величину.

**Теорема.** а) Рыночная цена облигации равна ее номинальной стоимости, тогда и только тогда, когда доходность к погашению равна

купонной ставке, т.е.  $V = N \Leftrightarrow \rho = c$ .

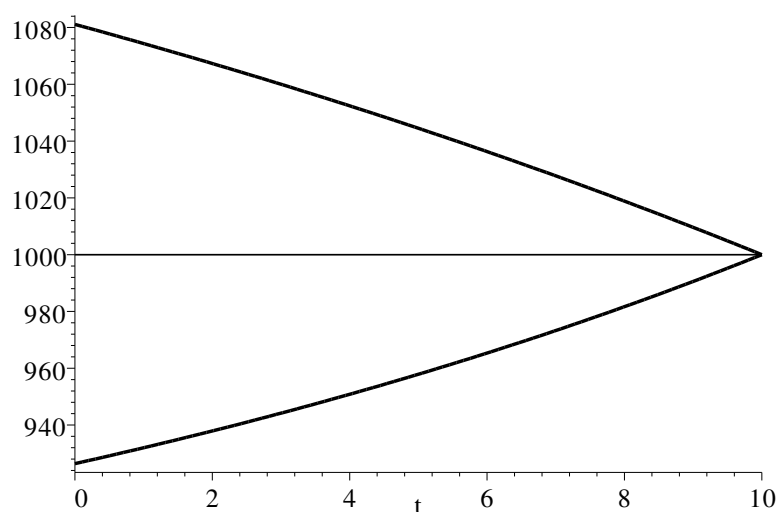
б) Рыночная цена облигации больше ее номинальной стоимости, тогда и только тогда, когда доходность к погашению меньше купонной ставки, т.е.  $V > N \Leftrightarrow \rho < c$ .

в) Рыночная цена облигации меньше ее номинальной стоимости, тогда и только тогда, когда доходность к погашению больше купонной ставки, т.е.  $V < N \Leftrightarrow \rho > c$ .

Говорят, что в случае а) облигация продается по номиналу, в случае б) с премией, в случае в) с дисконтом.

**Теорема.** Если доходность облигации  $\rho$  не меняется в течение времени ее обращения, то величина дисконта или премии уменьшается при уменьшении срока ее обращения  $n$ . Цена облигации с уменьшением срока ее обращения приближается к номинальной. Величины дисконта или премии будут уменьшаться тем быстрее, чем меньше срок обращения.

На рисунке приведены графики зависимости цены облигации (вертикальная ось) от времени ее обращения (горизонтальная ось) для двух десятилетних облигаций номиналом 1000 с купонной ставкой 5 %: с доходностью 6 % (с дисконтом) — нижний график, с доходностью 4 % (с премией) — верхний график.



**Зависимости цены облигации от времени ее обращения**

*Теорема.* Относительное изменение цены облигации (в процентах) в результате изменения доходности к погашению будет тем меньше, чем выше купонная ставка.

**32. Дюрация потока платежей. Дюрация облигации и ее свойства. Выпуклость облигаций. Хеджирование риска изменений процентной ставки. Теорема об иммунизации.**

Дюрацией потока платежей **Ошибка! Источник ссылки не найден.**

называют величину  $D = \sum_{k=1}^n w_k t_k$ , где  $y$  — ставка дисконтирования,

$$w_k = \frac{R_k (1+y)^{-t_k}}{P}, \text{ а } P = \sum_{k=1}^n R_k (1+y)^{-t_k}.$$

Свойства дюрации для положительных потоков платежей.

1. Если  $n = 1$ , то  $D = t_n$ . Если  $n > 1$ , то  $D < t_n$ .
2. Выполняется соотношение  $\frac{P'(y)}{P(y)} = -\frac{D}{1+y}$ .
3.  $D = D(y)$  — убывающая функция процентной ставки  $y$ .

В случае облигаций поток платежей относительно процентной ставки (доходности к погашению)  $y$  имеет вид  $\{(1, cN), (2, cN), \dots, (n, cN + N)\}$ , где  $c$  — купонная ставка,  $N$  — номинальная стоимость,  $n$  — срок погашения.

Для бескупонной облигации дюрация совпадает со сроком погашения:  
 $D = n$ .

Для относительного изменения цены облигации  $\frac{\Delta V}{V}$  при изменении доходности на величину  $\Delta y$  справедлива приближенная формула

$$\frac{\Delta V}{V} \approx -\frac{D}{1+y} \Delta y.$$

Если облигация продается по номиналу, то  $D = \frac{1+y}{y} (1 - (1+y)^{-n})$ .

*Теорема.* Если купонная ставка больше или равна доходности к

погашению ( $c \geq y$ ), то  $D = D(n)$  – возрастающая функция от  $n$ .

Если  $c < y$ , то функция  $D(n)$  имеет единственный максимум, приближенную оценку которого дает формула  $n_{\max} \approx \frac{1}{\ln(1+y)} + \frac{1+y}{y-c}$ .

**Выпуклостью** облигации  $W(y)$  при данной доходности  $y$  называют величину

$$W(y) = \frac{V''(y)}{V(y)}.$$

*Теорема.* Для относительного изменения цены облигации при изменении доходности на величину  $\Delta y$  справедлива приближенная формула

$$\frac{\Delta V}{V} \approx -\frac{D}{1+y} \Delta y + \frac{1}{2} W \cdot (\Delta y)^2.$$

*Теорема об иммунизации.* Предположим, что требуется выплатить долг в размере  $R$  в момент времени  $t$ . Покупка бескупонной облигации номинальной стоимостью  $N = R$  со сроком погашения  $t$  обеспечит выплату долга. Назовем ее облигацией I. Текущая стоимость облигации I равна  $P_1 = N / (1+r)^t$ .

Рассмотрим две бескупонные облигации с номинальными стоимостями  $N_1$  и  $N_2$  и сроками погашения соответственно  $t_1$  и  $t_2$ , такие, что выполняется неравенство  $t_1 < t < t_2$ .

Портфель, состоящий из этих облигаций, назовем облигацией II. Облигация II имеет текущую стоимость  $P_2 = \frac{N_1}{(1+r)^{t_1}} + \frac{N_2}{(1+r)^{t_2}}$ . Потребуем,

чтобы при  $r = r_0$  выполнялись условия  $\begin{cases} P_1(r_0) = P_2(r_0), \\ D_1(r_0) = D_2(r_0). \end{cases}$  Эти условия

обеспечивают эквивалентность двух денежных потоков, связанных с облигациями I и II, и равенство их дюраций при  $r = r_0$ . Графики функций



$P_1(r)$  и  $P_2(r)$  касаются в некоторой точке  $r = r_0$ . Для остальных значений  $r$  выполняется неравенство  $P_1(r) < P_2(r)$ .

Для того, чтобы обеспечить выполнение указанных выше условий, достаточно потребовать, чтобы веса платежей  $w_1, w_2$  удовлетворяли системе уравнений

$$\begin{cases} w_1 + w_2 = 1, \\ t_1 w_1 + t_2 w_2 = t. \end{cases}$$

В рассмотренной ситуации говорят, что облигация II *иммунизирует* облигацию I. При небольших изменениях доходности к погашению иммунизирующий пакет облигаций будет иметь большую стоимость, чем иммунизируемый. Тем самым, иммунизируя пакет облигаций, можно хеджировать риск изменения процентной доходности к погашению.

Идея иммунизации принадлежит лауреату Нобелевской премии П. Самюэльсону.

### **33. Портфель ценных бумаг, его доходность и риск. Эффективное множество портфелей. Оптимальный портфель при наличии безрискового актива.**

Фиксируем период времени, в течение которого инвестор владеет ценной бумагой. Стоимость этой ценной бумаги в начале периода обозначим  $p^0$ , в конце периода –  $p^1$ . **Доходностью** ценной бумаги за указанный период называется величина  $r = \frac{p^1 - p^0}{p^0}$ .

**Портфелем**, состоящим из  $n$  видов ценных бумаг, называют вектор  $X = (x_1, x_2, \dots, x_n)$ , где  $x_i$  – доля инвестиций в ценные бумаги вида  $i$ . **Доходностью портфеля**  $X$  называют величину  $r_X = \frac{p_X^1 - p_X^0}{p_X^0}$ , где  $p_X^0$  – стоимость портфеля в начале периода,  $p_X^1$  – в конце периода.

Доходность портфеля  $X$  можно вычислить по формуле  $r_X = x_1 r_1 + x_2 r_2 + \dots + x_n r_n$ , где  $r_1, r_2, \dots, r_n$  – доходности ценных бумаг, входящих в

портфель  $X$ .

Определяющую роль в теории портфельного анализа играют две характеристики рискового актива: **ожидаемая доходность** и **риск**. Ожидаемая доходность  $\mu$  – это математическое ожидание доходности  $R$  рискового актива  $\mu = E(R)$ . Для количественной оценки риска используется среднее квадратичное отклонение доходности. Если  $\sigma^2 = D(R)$  – дисперсия доходности рискового актива, то мерой риска служит величина  $\sigma$ .

Используя свойства математического ожидания, получим формулу для ожидаемой доходности портфеля

$$\mu = x_1\mu_1 + x_2\mu_2 + \dots + x_n\mu_n,$$

где  $\mu_1 = E(R_1)$ ,  $\mu_2 = E(R_2)$ , ...,  $\mu_n = E(R_n)$  – ожидаемые доходности составляющих портфель активов. Обозначим через  $\vec{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$  – вектор ожидаемых доходностей, тогда в матричных обозначениях имеем  $\mu = \vec{\mu}^T X$ , где  $X = (x_1, x_2, \dots, x_n)^T$  – вектор-столбец долей активов в портфеле.

Для вычисления риска воспользуемся формулой  $\sigma^2 = X^T V X$ , где  $V$  – ковариационная матрица случайных величин  $R_1, R_2, \dots, R_n$ .

Обозначим вектор, состоящий из одних единиц, через  $I = (1, 1, \dots, 1)^T$ .

Задача построения оптимального портфеля имеет следующий вид:

$\frac{1}{2}\sigma^2 = \frac{1}{2}X^T V X \rightarrow \min$ , при условиях  $\vec{\mu}^T X = \mu$ ,  $I^T X = 1$ , где  $I = (1, 1, \dots, 1)^T$  — вектор,

Основные предположения:

- 1) ковариационная матрица  $V$  положительно определена;
- 2) среди ожидаемых доходностей  $\mu_1, \mu_2, \dots, \mu_n$  нет одинаковых.

Положим:

$$\alpha = I^T V^{-1} I, \quad \beta = I^T V^{-1} \vec{\mu} = \vec{\mu}^T V^{-1} I, \quad \gamma = \vec{\mu}^T V^{-1} \vec{\mu}, \quad \delta = \alpha\gamma - \beta^2.$$

Задача нахождения оптимального портфеля имеет единственное

решение:

$$X = V^{-1}(\lambda I + \tau \bar{\mu}), \text{ где } \lambda = (\gamma - \beta\mu)/\delta, \tau = (\alpha\mu - \beta)/\delta.$$

Таким образом, для каждого значения ожидаемой доходности  $\mu$  имеется единственный портфель  $X$ , обеспечивающий минимальное значение риска  $\sigma = \sigma_{\min}$ . Сопоставляя допустимому значению  $\mu$  минимальное значение риска  $\sigma$ , получаем функцию  $\sigma = \sigma(\mu)$ . График этой функции называют **минимальной границей**. Уравнение минимальной границы имеет вид

$$\sigma^2 = \frac{\alpha\mu^2 - 2\beta\mu + \gamma}{\delta}.$$

Минимальная граница представляет собой ветвь гиперболы с асимптотами  $\sigma = \sqrt{\frac{\alpha}{\delta}} \left| \mu - \frac{\beta}{\alpha} \right|$  и абсолютным минимумом  $M\left(\frac{\beta}{\alpha}, \frac{1}{\sqrt{\alpha}}\right)$ .

Сопоставление  $X \rightarrow (\mu_X, \sigma_X)$  задает отображение множества допустимых портфелей на плоскость  $(\mu, \sigma)$ . На множестве портфелей можно задать отношение предпочтения, считая, что  $X \succ Y$ , если  $\mu_X \geq \mu_Y$ ,  $\sigma_X \leq \sigma_Y$  и хотя бы одно из неравенств строгое. Множество точек плоскости  $(\mu, \sigma)$ , соответствующих недоминируемым портфелям, называется **эффективным множеством** или **эффективным фронтом**. Это множество совпадает с частью минимальной границы, лежащей правее точки абсолютного минимума.

Предположим, что вместе с  $n$  рисковыми активами портфель инвестора включает безрисковую компоненту с детерминированной доходностью  $\mu_f = R_f$  и долей в портфеле, составляющей  $x_f$ . При этом задача построения оптимального портфеля будет выглядеть следующим образом:

$$\frac{1}{2}\sigma^2 = \frac{1}{2}X^T V X \rightarrow \min, \text{ при условиях } \mu_f x_f + \bar{\mu}^T X = \mu, x_f + I^T X = 1.$$

$$\text{Положим } d = \sqrt{(\bar{\mu} - \mu_f I)^T V^{-1} (\bar{\mu} - \mu_f I)}. \text{ Тогда } X = \frac{\mu - \mu_f}{d^2} V^{-1} (\bar{\mu} - \mu_f I) —$$

искомый вектор рискованных долей.

При  $\mu \geq \mu_f$  уравнение минимальной границы имеет вид  $\sigma = \frac{\mu - \mu_f}{d}$ .

Прямая, заданная этим уравнением является касательной к графику минимальной границы портфелей, составленных из рисковых активов. Точка касания соответствует портфелю рисковых активов, который называется **касательным портфелем**.

*Теорема.* Всякий портфель  $X_M$ , соответствующий точке  $M$ , лежащей на минимальной границе, представляется в виде линейной комбинации безрискового портфеля  $X_F$  и касательного портфеля  $X_T$ :  $X_M = \lambda X_F + (1 - \lambda) X_T$ .

### 34. Биномиальная модель ценообразования активов. Оценка опционов с использованием биномиальной модели. Модель Кокса-Росса-Рубинштейна.

Начнем с **однопериодной биномиальной модели**. В рамках этой модели предполагается, что на рынке имеются рисковый актив и безрисковый актив. За единичный промежуток времени текущая стоимость рискового актива  $S$  изменится и примет одно из двух возможных значений: верхнее значение  $S_u$  или нижнее значение  $S_d$ , так, что  $S_d < S_u$ . Положим  $u = \frac{S_u}{S}$ ,  $d = \frac{S_d}{S}$ . Обозначим через  $r$  безрисковую ставку на единичный промежуток времени. В соответствии с теоремой об арбитраже отсутствие арбитражных возможностей равносильно тому, что существует нейтральная к риску вероятность. В рассматриваемом случае — числа  $\tilde{p}$  и  $\tilde{q}$ , такие, что  $0 \leq \tilde{p}, \tilde{q} \leq 1$ ,  $\tilde{p} + \tilde{q} = 1$  и

$$\left(\frac{S_u}{1+r} - S\right)\tilde{p} + \left(\frac{S_d}{1+r} - S\right)\tilde{q} = 0.$$

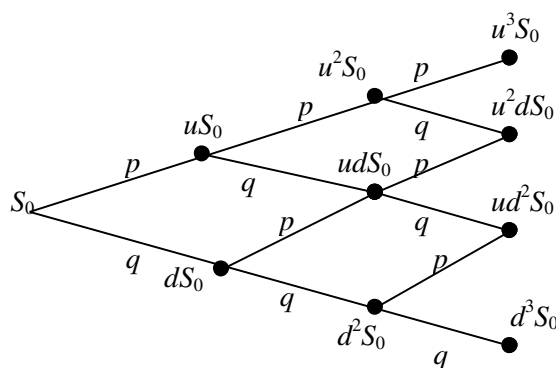
Отсюда

$$\tilde{p} = \frac{1+r-d}{u-d}, \quad \tilde{q} = \frac{u-(1+r)}{u-d}.$$

*Следствие.* В однопериодной биномиальной модели арбитражные возможности отсутствуют тогда и только тогда, когда  $d \leq 1 + r \leq u$ .

Обозначим через  $S(t)$  цену рискового актива в момент времени  $t$ .

Пусть  $h$  – некоторый фиксированный промежуток времени. Предполагается, что за время  $h$  цена актива может возрасти и стать равной  $uS(t)$ , где  $u$  – повышающий коэффициент, или понизиться и стать равной  $dS(t)$ , где  $d$  – понижающий коэффициент. Повышение происходит с вероятностью  $p$ , а понижение – с вероятностью  $q = 1 - p$ . Таким образом,  $S(t + h) = uS(t)$  с вероятностью  $p$ , и  $S(t + h) = dS(t)$  с вероятностью  $q$ . Если считать, что изменение цены актива происходит по этой же схеме на протяжении нескольких временных периодов длительности  $h$ , получается **многопериодная биномиальная модель**. На рисунке схематически представлен процесс изменения цен рискованного актива в трехпериодной модели.



Если за промежуток времени  $T = nh$  цена актива повышалась  $k$  раз и понижалась  $n - k$  раз, то она станет равной  $S_0 u^k d^{n-k}$ . В соответствии со схемой Бернулли можно определить вероятность этого события:

$$P(S(T) = S_0 u^k d^{n-k}) = C_n^k p^k q^{n-k}.$$

Пусть  $Bin(n, k, p)$  – вероятность того, что повышение произошло  $k$  и более раз, т. е.

$$Bin(n, k, p) = \sum_{i=k}^n C_n^i p^i (1-p)^{n-i}.$$

Если  $n$  достаточно велико, то, применяя интегральную приближенную формулу Лапласа, получаем:

$$Bin(n, k, p) \approx N\left(\frac{np - k}{\sqrt{np(1-p)}}\right),$$

где  $N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$  – функция стандартного нормального распределения

Производный инструмент европейского типа со сроком исполнения  $T$  определяется своей *платежной функцией*  $f_T(S_T)$ . Например,  $f_T(S_T) = (S_T - K)^+$  для опциона колл с ценой исполнения  $K$ ;  $f_T(S_T) = (K - S_T)^+$  для опциона пут с ценой исполнения  $K$ .

Пусть  $S(t)$  – стоимость рискового актива в момент времени  $t$ . Стоимость рискового актива в момент времени  $T = t + nh$  может рассматриваться как случайная величина, распределенная по биномиальному закону относительно риск-нейтральной вероятности:

$$\tilde{P}(S_T = Su^j d^{n-j}) = C_n^j \tilde{p}^j \tilde{q}^{n-j}, j = 0, 1, \dots, n.$$

Безарбитражная цена производного инструмента в момент времени  $t$  равна математическому ожиданию дисконтированного значения его платежной функции относительно этого распределения **Ошибка! Источник ссылки не найден.:**

$$C(t) = \tilde{E}\left(\frac{f_T(S_T)}{(1+r)^n}\right).$$

В частности, для опциона колл с ценой исполнения  $K$  имеет место равенство

$$C(t) = \sum_{j=1}^n \frac{(S(t)u^j d^{n-j} - K)^+}{(1+r)^n} C_n^j \tilde{p}^j \tilde{q}^{n-j}.$$

Пусть число  $m$  — наименьшее целое решение неравенства

$$S(t)u^m d^{n-m} > K.$$

Положим  $\tilde{p}^* = \frac{\tilde{p}u}{1+r}$ ;  $\tilde{q}^* = \frac{\tilde{q}d}{1+r}$ . Тогда

$$C(t) = S(t)Bin(n, m, \tilde{p}^*) - \frac{K}{(1+r)^n} Bin(n, m, \tilde{p}).$$

Формулу называют формулой **Кокса-Росса-Рубинштейна**.

Пусть  $\sigma$  — волатильность, а  $\mu$  — ожидаемая доходность рисковому активу за единицу времени. Тогда дисперсия доходности рисковому активу за единицу времени составляет  $\sigma^2 h$  и, поскольку  $\tilde{p} = \frac{e^{\mu h} - d}{u - d}$ , получаем

$$\tilde{p}u^2 + \tilde{q}d^2 - [\tilde{p}u + \tilde{q}d]^2 = \sigma^2 h.$$

Так как  $\tilde{p}u^2 + \tilde{q}d^2 = (\tilde{p}u + \tilde{q}d)(u + d) - (\tilde{p} + \tilde{q})ud$  и  $e^{\mu h} = \tilde{p}u + \tilde{q}d$ , то

$$e^{\mu h}(u + d) - ud - e^{2\mu h} = \sigma^2 h.$$

Предположим дополнительно, что  $ud = 1$  и  $u$  имеет вид  $u = e^{x(h)}$ . Тогда с точностью до слагаемых порядка выше первого получаем

$$(1 + \mu h) \left( 1 + x + \frac{x^2}{2} + 1 - x + \frac{x^2}{2} \right) - 1 - (1 + 2\mu h) = \sigma^2 h.$$

Отсюда  $x = \sigma\sqrt{h}$ , так что  $u = e^{\sigma\sqrt{h}}$  и  $d = e^{-\sigma\sqrt{h}}$ .

С учетом этого для расчетов в рамках модели **Кокса-Росса-Рубинштейна** полагают  $u = e^{\sigma\sqrt{T/n}}$ ,  $d = e^{-\sigma\sqrt{T/n}}$ ,  $r = e^{\mu\frac{T}{n}} - 1$ .

### **35. Модель Блэка-Шоулза. Уравнение Блэка-Шоулза. Оценка производных финансовых инструментов с использованием модели Блэка-Шоулза.**

Модель Блэка-Шоулза основывается на предположении о том, что стоимость рисковому актива  $S(t)$  совершает так называемое *геометрическое броуновское движение*. В соответствии с этим предположением случайная величина  $\ln\left(\frac{S(t_0+t)}{S(t_0)}\right)$  (доходность рисковому актива в виде ставки непрерывных процентов за период  $t$ , или *логарифмическая доходность*) нормально распределена с математическим ожиданием  $\mu t$  и дисперсией  $\sigma^2 t$ , где  $\mu$  и  $\sigma$  постоянны.

Считая, что  $t_0 = 0$  и полагая  $S_0 = S(0)$ , логарифмическую доходность рисковому актива за время  $t$  можно представить в виде

$$\ln \frac{S(t)}{S_0} = \left( \mu - \frac{\sigma^2}{2} \right) t + \sigma W_t,$$

где  $W_t$  – нормально распределенная случайная величина с нулевым математическим ожиданием и дисперсией  $t$ . Эту формулу можно также переписать в виде

$$\ln \frac{S(t)}{S_0} = \left( \mu - \frac{\sigma^2}{2} \right) t + \sigma \sqrt{t} \cdot Z,$$

где  $Z$  – стандартная нормальная случайная величина.

Если  $S = S(t)$  – цена базового актива в момент  $t$ , то в момент  $T$  цена  $S(T)$  является случайной величиной с функцией распределения

$$P(S_T < x) = N(a),$$

где

$$N(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-t^2/2} dt$$

– функция стандартного нормального распределения, а число  $a$  определяется из уравнения

$$S(t) e^{\left( \mu - \frac{\sigma^2}{2} \right) (T-t) + \sigma \sqrt{T-t} \cdot a} = x.$$

Для определения цены, не допускающей арбитража используется риск-нейтральная вероятность. Относительно риск-нейтральной вероятности цена  $S_T$  может быть представлена в виде

$$S_T = S e^{\left( r - \frac{\sigma^2}{2} \right) (T-t) + \sigma \sqrt{T-t} \cdot Z},$$

где  $S = S(t)$ , а  $Z$  – стандартная нормальная случайная величина.

Безарбитражная цена производного инструмента в момент времени  $t$  равна математическому ожиданию дисконтированного значения его платежной функции относительно этого распределения:

$$C = \tilde{E} \left( f_T(S_T) e^{-r(T-t)} \right).$$

Для опциона колл с ценой исполнения  $K$  формула приобретает следующий вид:



$$C = \int_{-\infty}^{+\infty} \left( S e^{\left(r - \sigma^2 / 2\right)(T-t) + \sigma \sqrt{T-t} \xi} - K \right)^+ e^{-r(T-t)} \varphi(\xi) d\xi,$$

где  $\varphi(\xi)$  — функция плотности стандартного нормального распределения.

Вычисление интеграла дает следующую формулу (**формулу Блэка – Шоулза**):

$$C = SN(d_1) - Ke^{-r(T-t)}N(d_2),$$

где

$$d_{1,2} = \frac{\ln \frac{S}{K} + \left(r \pm \sigma^2 / 2\right)(T-t)}{\sigma \sqrt{T-t}}.$$

Для опциона пут безарбитражная цена задается формулой

$$P = -SN(-d_1) + Ke^{-r(T-t)}N(-d_2).$$

В самом деле, достаточно воспользоваться паритетом цен опционов колл и пут. Так как  $C - P = S - Ke^{-r(T-t)}$ , то

$$\begin{aligned} P &= SN(d_1) - Ke^{-r(T-t)}N(d_2) - S + Ke^{-r(T-t)} = \\ &= -S \cdot (1 - N(d_1)) + Ke^{-r(T-t)} \cdot (1 - N(d_2)) = \\ &= -SN(-d_1) + Ke^{-r(T-t)}N(-d_2). \end{aligned}$$

*Теорема.* Стоимость опциона колл удовлетворяет следующему дифференциальному уравнению:

$$\frac{\partial C}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0.$$

Это уравнение называется **уравнением Блэка-Шоулза**.

### **36. Договоры страхования с вычетом, лимитом и франшизой. Основные вероятностные и числовые характеристики**

*Договор страхования с вычетом.* Убытки страхователя можно представить в виде  $\xi = I\eta$ , где случайная величина  $\xi \geq 0$  – иск по договору (без модификации величины выплаты),  $I$  – индикатор наступления страхового случая по договору, случайная величина  $\eta > 0$  – страховая выплата по договору после наступления страхового случая (реальные убытки страхователя).

Договор страхования с вычетом предполагает, что если реальные потери страхователя  $\eta$  больше  $d$  (вычет), тогда страховая компания выплачивает сумму  $\eta - d$ ; а если потери страхователя  $\eta$  меньше или равны  $d$ , то страховая компания ничего не платит (в этом случае выплата считается равной 0). Случайная величина (СВ) выплат по этому договору (обозначается  $(\eta - d)_+$  или  $\eta_{(d)}$ ) имеет вид

$$(\eta - d)_+ = \max(\eta - d, 0) = \begin{cases} 0, & \eta \leq d, \\ \eta - d, & \eta > d. \end{cases}$$

Функция распределения  $\eta_{(d)}$  имеет вид

$$F_{\eta_{(d)}}(y) = \mathbf{P}(\eta_{(d)} \leq y) = \begin{cases} 0, & y < 0, \\ F_{\eta}(y + d), & y \geq 0. \end{cases}$$

Действительно, если  $y \geq 0$ , то

$$F_{\eta_{(d)}}(y) = \mathbf{P}(\eta_{(d)} \leq y) = \mathbf{P}(\eta - d \leq y) = \mathbf{P}(\eta \leq y + d) = F_{\eta}(y + d).$$

Если  $\eta$  обладает плотностью  $f_{\eta}(x)$ , то  $\eta_{(d)}$  – СВ смешанного типа.

Дискретная часть  $\eta_{(d)}$  имеет функцию вероятностей

$$p_{\eta_{(d)}}(y) = \mathbf{P}(\eta_{(d)} = y) = \begin{cases} F_{\eta}(d), & y = 0, \\ 0, & \text{в остальных случаях.} \end{cases}$$

Непрерывная часть  $\eta_{(d)}$  имеет плотность

$$f_{\eta_{(d)}}(y) = \begin{cases} f_{\eta}(y + d), & y > 0, \\ 0, & \text{в остальных случаях.} \end{cases}$$

$k$ -й начальный момент (если существует) для  $\eta_{(d)}$  вычисляется по правилу

$$\mathbf{E}(\eta_{(d)}^k) = \int_d^{\infty} (x - d)^k f_{\eta}(x) dx = k \int_d^{\infty} (x - d)^{k-1} (1 - F_{\eta}(x)) dx,$$

если  $\eta$  – непрерывная случайная величина с плотностью  $f_{\eta}(x)$ , и

$$\mathbf{E}(\eta_{(d)}^k) = \sum_{x_i > d} (x_i - d)^k p_{\eta}(x_i),$$

если  $\eta$  – дискретная СВ с функцией вероятностей  $p_{\eta}(x) = \mathbf{P}(\eta = x)$ .

Если не учитывать нулевую выплату (т.е. рассматриваются выплаты, производимые после предъявления иска по договору с вычетом), то мы приходим к рассмотрению случайной величины  $\eta_{(d)}^+ = [\eta - d | \eta > d]$ .

Функция распределения для  $\eta_{(d)}^+$  имеет вид

$$F_{\eta_{(d)}^+}(y) = \frac{F_{\eta}(y + d) - F_{\eta}(d)}{1 - F_{\eta}(d)}, \quad y > 0.$$

Действительно,

$$\begin{aligned} F_{\eta_{(d)}^+}(y) &= \mathbf{P}(\eta_{(d)}^+ \leq y) = \mathbf{P}(\eta_{(d)} \leq y | \eta_{(d)} > 0) = \mathbf{P}(\eta - d \leq y | \eta > d) = \\ &= \frac{\mathbf{P}(d < \eta \leq y + d)}{\mathbf{P}(\eta > d)} = \frac{F_{\eta}(y + d) - F_{\eta}(d)}{1 - F_{\eta}(d)}, \quad y > 0. \end{aligned}$$

Отсюда видно, что СВ  $\eta_{(d)}^+$  наследует тип СВ  $\eta$ .

В непрерывном случае плотности  $\eta$  и  $\eta_{(d)}^+$  связаны соотношением

$$f_{\eta_{(d)}^+}(y) = \frac{f_{\eta}(y + d)}{1 - F_{\eta}(d)}, \quad y > 0.$$

$k$ -й начальный момент (если существует) СВ  $\eta_{(d)}^+$  вычисляется по правилу

$$\mathbf{E}[\eta_{(d)}^+]^k = \frac{\mathbf{E}(\eta_{(d)}^k)}{1 - F_{\eta}(d)}.$$

$\mathbf{E}[\eta_{(d)}^+]^k$  при  $k = 1$  называется функцией эксцедента убытков (обозначается через  $e_{\xi}(d)$ ).

*Договор страхования с лимитом.* Согласно этому договору, страховщик покрывает все потери страхователя, не превышающие значение  $l$  (лимит). Если потери больше  $l$ , то страховщик выплачивает сумму  $l$ . Обозначим выплату по этому договору через  $\eta \wedge l$ . Тогда

$$\eta \wedge l = \min(\eta, l) = \begin{cases} \eta, & \eta \leq l, \\ l, & \eta > l. \end{cases}$$

Заметим, что  $\eta = \eta_{(l)} + (\eta \wedge l) = (\eta - l)_+ + (\eta \wedge l)$ . Отсюда

$$\mathbf{E}[(\eta - l)_+] = \mathbf{E}(\eta) - \mathbf{E}(\eta \wedge l).$$

Функция распределения  $\eta \wedge l$  имеет вид

$$F_{\eta \wedge l}(y) = \mathbf{P}(\eta \wedge l \leq y) = \begin{cases} \mathbf{P}(\eta \leq y) = F_{\eta}(y), & y < l, \\ 1, & y \geq l. \end{cases}$$

Если  $\eta$  – случайная величина непрерывного типа, то  $\eta \wedge l$  – случайная величина смешанного типа. Ее дискретная часть сосредоточена в точке  $l$ :

$$\mathbf{P}(\eta \wedge l = l) = \mathbf{P}(\eta \geq l) = 1 - F_{\eta}(l).$$

Непрерывная часть  $\eta \wedge l$  имеет плотность

$$f_{\eta \wedge l}(y) = \begin{cases} f_{\eta}(y), & y < l, \\ 0, & y \geq l. \end{cases}$$

**Моменты  $\eta \wedge l$ .** Если  $\eta$  является дискретной случайной величиной с законом распределения  $p_{\eta}(x_i) = \mathbf{P}(\eta = x_i)$ ,  $i = 1, 2, \dots$ , то

$$\mathbf{E}((\eta \wedge l)^k) = \sum_{x_i \leq l} x_i^k p_{\eta}(x_i) + l^k (1 - F_{\eta}(l)).$$

Если  $\eta$  является непрерывной СВ с плотностью  $f_{\eta}(x)$ , то

$$\mathbf{E}[(\eta \wedge l)^k] = \int_0^l x^k f_{\eta}(x) dx + l^k (1 - F_{\eta}(l)) = k \int_0^l x^{k-1} (1 - F_{\eta}(x)) dx.$$

*Договор страхования с франшизой.* Согласно условиям договора с франшизой страховая компания возмещает только убытки, превышающие значение  $d_f$  (франшиза):

$$\eta_{d_f} = \begin{cases} 0, & \eta \leq d_f, \\ \eta, & \eta > d_f. \end{cases}$$

Случайная величина  $\eta_{d_f}$  не обладает плотностью (даже если  $\eta$  обладает плотностью), поскольку вероятность того, что не будет предъявлен иск положительна:

$$\mathbf{P}(\eta_{d_f} = 0) = \mathbf{P}(0 < \eta \leq d_f) = F_{\eta}(d_f) > 0.$$

Реальные выплаты описываются СВ  $\eta_{d_f}^+ = [\eta | \eta > d_f]$ .

Функция распределения  $\eta_{d_f}$  имеет вид

$$F_{\eta_{d_f}}(y) = \mathbf{P}(\eta_{d_f} \leq y) = \begin{cases} 0, & y < 0, \\ F_{\eta}(d_f), & 0 \leq y \leq d_f, \\ F_{\eta}(y), & y > d_f. \end{cases}$$

Действительно, если  $y > d_f$ , то  $F_{\eta_{d_f}}(y) = \mathbf{P}(\eta_{d_f} \leq y) = \mathbf{P}(\eta \leq y) = F_{\eta}(y)$ .

Если  $\eta$  обладает плотностью  $f_{\eta}(x)$ , то  $\eta_{d_f}$  – СВ смешанного типа.

Дискретная часть  $\eta_{d_f}$  имеет функцию вероятностей

$$p_{\eta_{d_f}}(y) = \begin{cases} F_{\eta}(d_f), & y = 0, \\ 0, & \text{в остальных случаях,} \end{cases}$$

непрерывная часть  $\eta_{d_f}$  имеет плотность

$$f_{\eta_{d_f}}(y) = \begin{cases} f_{\eta}(y), & y > d_f, \\ 0, & \text{в остальных случаях.} \end{cases}$$

Далее функция распределения реальных выплат имеет вид

$$F_{\eta_{d_f}^+}(y) = \begin{cases} 0, & 0 \leq y \leq d_f, \\ \frac{F_{\eta}(y) - F_{\eta}(d_f)}{1 - F_{\eta}(d_f)}, & y > d_f \end{cases}$$

поскольку при  $y > d_f$  имеем

$$\begin{aligned} F_{\eta_{d_f}^+}(y) &= \mathbf{P}(\eta_{d_f}^+ \leq y) = \mathbf{P}(\eta \leq y | \eta > d_f) = \frac{\mathbf{P}(d_f < \eta \leq y)}{\mathbf{P}(\eta > d_f)} \\ &= \frac{F_{\eta}(y) - F_{\eta}(d_f)}{1 - F_{\eta}(d_f)}. \end{aligned}$$

Отсюда видно, что СВ  $\eta_{d_f}^+$  наследует тип случайной величины  $\eta$ .

В непрерывном случае плотности  $\eta$  и  $\eta_{d_f}^+$  связаны соотношением

$$f_{\eta_{d_f}^+}(y) = \frac{f_{\eta}(y)}{1 - F_{\eta}(d_f)}, \quad y > d_f.$$

Если  $\eta$  непрерывного типа, то

$$\begin{aligned} \mathbf{E}(\eta_{d_f}^k) &= \int_{d_f}^{\infty} x^k f_{\eta}(x) dx, \\ \mathbf{E}(\eta_{d_f}^{+k}) &= \frac{1}{1 - F_{\eta}(d_f)} \int_{d_f}^{\infty} x^k f_{\eta}(x) dx = \frac{\mathbf{E}(\eta_{d_f}^k)}{1 - F_{\eta}(d_f)}. \end{aligned}$$

Используя представление  $xf_{\eta}(x) = (x - d_f)f_{\eta}(x) + d_ff_{\eta}(x)$ , получим

$$\begin{aligned}
\mathbf{E}(\eta_{d_f}) &= \mathbf{E}(\eta - d_f)_+ + d_f (1 - F_\eta(d_f)) = \\
&= \mathbf{E}(\eta) - \mathbf{E}(\eta \wedge d_f) + d_f (1 - F_\eta(d_f)), \\
\mathbf{E}(\eta_{d_f}^+) &= \mathbf{E}(\eta - d_f)_+ + d_f = \frac{\mathbf{E}(\eta) - \mathbf{E}(\eta \wedge d_f)}{1 - F_\eta(d_f)} + d_f.
\end{aligned}$$

### 37. Индивидуальные модели риска. Вероятность разорения. Принципы назначения страховых премий.

Предположения модели:

1. Рассматривается портфель за фиксированный относительно короткий период времени;
2. Портфель состоит из фиксированного числа договоров, которое не изменяется на протяжении рассматриваемого периода времени;
3. Требование о выплате изучается на уровне договора страхования;
4. Плата за страховку полностью вносится в начале рассматриваемого периода.

Положим  $S_n = \xi_1 + \dots + \xi_n = I_1 \eta_1 + \dots + I_n \eta_n$ , где  $S_n$  – величина суммарного иска,  $\xi_k \geq 0$  – иск по  $k$ -му договору,  $I_k$  – индикатор наступления страхового случая по  $k$ -му договору,  $\eta_k > 0$  – страховая выплата по  $k$ -му договору после наступления страхового случая.

**Расчет  $\mathbf{E}(S_n)$  и  $\mathbf{Var}(S_n)$ .** Пусть  $\xi_k$ ,  $k = 1, \dots, n$  – независимые случайные величины и  $\mathbf{P}(I_k = 1) = q_k$  – вероятность предъявления иска по  $k$ -му договору,  $\mathbf{P}(I_k = 0) = 1 - q_k = p_k$ . Тогда

$$\mathbf{E}(\xi_k) = \mathbf{E}(I_k) \mathbf{E}(\eta_k) = q_k m_{\eta_k},$$

$$\mathbf{Var}(\xi_k) = p_k q_k m_{\eta_k}^2 + q_k \sigma_{\eta_k}^2.$$

Здесь  $m_{\eta_k} = \mathbf{E}(\eta_k)$ ,  $\sigma_{\eta_k}^2 = \mathbf{Var}(\eta_k)$ . Отсюда

$$\mathbf{E}(S_n) = \mathbf{E}\left(\sum_{k=1}^n \xi_k\right) = \sum_{k=1}^n \mathbf{E}(\xi_k) = \sum_{k=1}^n q_k m_{\eta_k},$$

$$\mathbf{Var}(S_n) = \mathbf{Var}\left(\sum_{k=1}^n \xi_k\right) = \sum_{k=1}^n \mathbf{Var}(\xi_k) = \sum_{k=1}^n (p_k q_k m_{\eta_k}^2 + q_k \sigma_{\eta_k}^2).$$

*Частный случай.* Пусть  $\xi_k$  одинаково распределены как  $\xi$ ,  $\eta_k$  – как  $\eta$  и  $I_k$  – как  $I$ . Положим  $p_k = p$ ,  $q_k = q$ ,  $k = 1, \dots, n$ ,  $m_{\eta_k} = m_\eta$ ,  $\sigma_{\eta_k}^2 = \sigma_\eta^2$ . Тогда

$$\mathbf{E}(S_n) = nqm_\eta, \quad \mathbf{Var}(S_n) = npqm_\eta^2 + nq\sigma_\eta^2.$$

*Вероятность разорения.* Пусть  $u$  – резервы страховой компании. Вероятность разорения определяется как

$$R(u) = \mathbf{P}(\xi_1 + \dots + \xi_n > u) = \mathbf{P}(S_n > u) = 1 - \mathbf{P}(S_n \leq u) = 1 - F_{S_n}(u),$$

где  $F_{S_n}(u)$  – функция распределения  $S_n$  или вероятность неразорения (так что  $R(u) = 1 - F_{S_n}(u)$  есть дополнительная функция распределения).

*Точный расчет вероятности разорения.* Пусть  $\xi_k$ ,  $k = 1, \dots, n$  – независимые случайные величины. Напомним некоторые факты. Пусть  $\eta_1, \eta_2$  – независимые и непрерывные случайные величины. Пусть  $f_{\eta_1}, f_{\eta_2}$  соответствующие плотности. Тогда  $\eta_1 + \eta_2$  имеет плотность (она вычисляется с помощью свертки)

$$\begin{aligned} f_{\eta_1 + \eta_2}(x) &= (f_{\eta_1} * f_{\eta_2})(x) = \int_{-\infty}^x f_{\eta_1}(y)f_{\eta_2}(x-y)dy = \\ &= \int_{-\infty}^x f_{\eta_1}(x-y)f_{\eta_2}(y)dy = (f_{\eta_2} * f_{\eta_1})(x). \end{aligned}$$

Аналогично вычисляется плотность  $n$  независимых и непрерывных случайных величин:  $f_{\eta_1 + \eta_2 + \dots + \eta_n}(x) = (f_{\eta_1} * f_{\eta_2} * \dots * f_{\eta_n})(x)$ .

Операция свертки коммутативна и ассоциативна (поэтому последняя формула приведена без расстановки скобок).

*Явная формула для  $R(u)$  и  $F_{S_n}(u)$  при  $n = 2$ .*

$$R(u) = P(\xi_1 + \xi_2 > u) = q_1p_2\bar{F}_{\eta_1}(u) + q_2p_1\bar{F}_{\eta_2}(u) + q_1q_2\bar{F}_{\eta_1+\eta_2}(u),$$

$$F_{\xi_1+\xi_2}(u) = 1 - R(u) = 1 - q_1p_2\bar{F}_{\eta_1}(u) - q_2p_1\bar{F}_{\eta_2}(u) - q_1q_2\bar{F}_{\eta_1+\eta_2}(u).$$

Подставим  $\bar{F}_\eta = 1 - F_\eta \Rightarrow$

$$R(u) = q_1p_2 + q_2p_1 + q_1q_2 - q_1p_2F_{\eta_1}(u) - q_2p_1F_{\eta_2}(u) - q_1q_2F_{\eta_1+\eta_2}(u)$$

$$F_{\xi_1+\xi_2}(u) = 1 - q_1p_2 - q_2p_1 - q_1q_2 + q_1p_2F_{\eta_1}(u) +$$

$$+q_2p_1F_{\eta_2}(u) + q_1q_2F_{\eta_1+\eta_2}(u).$$

*Доказательство.*

$$\begin{aligned} R(u) = P(\xi_1 + \xi_2 > u) &= \underbrace{P(\xi_1 + \xi_2 > u | I_1 = 0, I_2 = 0)}_{=0} \underbrace{P(I_1 = 0, I_2 = 0)}_{P(I_1=0)P(I_2=0)=p_1p_2} + \\ &+ \underbrace{P(\xi_1 + \xi_2 > u | I_1 = 1, I_2 = 0)}_{=P(\eta_1>u)=\bar{F}_{\eta_1}(u)} \underbrace{P(I_1 = 1, I_2 = 0)}_{P(I_1=1)P(I_2=0)=q_1p_2} + \\ &\underbrace{P(\xi_1 + \xi_2 > u | I_1 = 0, I_2 = 1)}_{=P(\eta_2>u)=\bar{F}_{\eta_2}(u)} \underbrace{P(I_1 = 0, I_2 = 1)}_{P(I_1=0)P(I_2=1)=q_2p_1} + \\ &\underbrace{P(\xi_1 + \xi_2 > u | I_1 = 1, I_2 = 1)}_{=P(\eta_1+\eta_2>u)=\bar{F}_{\eta_1+\eta_2}(u)} \underbrace{P(I_1 = 1, I_2 = 1)}_{P(I_1=1)P(I_2=1)=q_1q_2} \end{aligned}$$

*Приближенный расчет вероятности разорения.* Пусть  $\xi_k$ ,  $k = 1, \dots, n$  – независимые и одинаково распределенные случайные величины. Тогда применяя центральную предельную теорему, находим

$$1 - R(u) = \mathbf{P}(S_n \leq u) = \mathbf{P}\left(\frac{S_n - \mathbf{E}(S_n)}{\sqrt{\mathbf{Var}(S_n)}} \leq \frac{u - \mathbf{E}(S_n)}{\sqrt{\mathbf{Var}(S_n)}}\right) \approx \Phi\left(\frac{u - \mathbf{E}(S_n)}{\sqrt{\mathbf{Var}(S_n)}}\right).$$

*Принципы назначения страховых премий*

Как раньше пусть

$$S_n = \xi_1 + \dots + \xi_n$$

– суммарный иск и  $P_1, \dots, P_n$  – премии, где  $P_k$  – премия  $k$ -го договора.

$$u = P_1 + \dots + P_n.$$

### 1. Принцип нетто-премии

Премия  $k$ -го договора определяется по принципу

$$P_k = \mathbf{E}(\xi_k).$$

Этот принцип приводит к высокой вероятности разорения равной 0,5 (предполагается применимость нормального приближения). Действительно в этом случае

$$\begin{aligned} \mathbf{E}(S_n) &= \sum_{k=1}^n \mathbf{E}(\xi_k) = \sum_{k=1}^n P_k = u \\ R(u) = \mathbf{P}(S_n > u) &= \mathbf{P}\left(\frac{S_n - \mathbf{E}(S_n)}{\sqrt{\mathbf{Var}(S_n)}} > 0\right) \approx \Phi(0) = \frac{1}{2} \end{aligned}$$



## 2. Принципы с надбавкой

Полагают

$$P_k = \mathbf{E}(\xi_k) + l_k.$$

Отсюда

$$u = \sum_{k=1}^n P_k = \sum_{k=1}^n \mathbf{E}(\xi_k) + \sum_{k=1}^n l_k = \mathbf{E}(S_n) + l,$$

где  $l = \sum_{k=1}^n l_k$ . Далее (предполагается применимость нормального приближения)

$$R(u) = \mathbf{P}(S_n > u) = P\left(\frac{S_n - \mathbf{E}(S_n)}{\sqrt{\mathbf{Var}(S_n)}} > \frac{l}{\sqrt{\mathbf{Var}(S_n)}}\right) \approx 1 - \Phi\left(\frac{l}{\sqrt{\mathbf{Var}(S_n)}}\right).$$

Если требуемый уровень неразорения равен  $\alpha$ , то

$$l = x_\alpha \sqrt{\mathbf{Var}(S_n)}.$$

Как "справедливо" разделить  $l$ ?

### а) Принцип математического ожидания

Полагают ( $l_k$  пропорционально  $\mathbf{E}(\xi_k)$ )

$$P_k = (1 + \lambda)\mathbf{E}(\xi_k).$$

Отсюда

$$l_k = \lambda \mathbf{E}(\xi_k) \quad \left| \sum_{k=1}^n \right. \Rightarrow l = \lambda \mathbf{E}(S_n)$$

и

$$\lambda = \frac{l}{\mathbf{E}(S_n)} = \frac{x_\alpha \sqrt{\mathbf{Var}(S_n)}}{\mathbf{E}(S_n)}.$$

В результате премия равна

$$P_k = \mathbf{E}(\xi_k) \left[ 1 + \frac{x_\alpha \sqrt{\mathbf{Var}(S_n)}}{\mathbf{E}(S_n)} \right].$$

Этот принцип несправедлив к договорам с маленькой дисперсией.

### б) Принцип дисперсии

Полагают ( $l_k$  пропорционально  $\mathbf{Var}(\xi_k)$ )

$$l_k = \lambda \mathbf{Var}(\xi_k) \quad \left| \sum_{k=1}^n \right. \Rightarrow (\text{по независимости } \xi_k)$$

$$l = \lambda \mathbf{Var}(S_n) \Rightarrow$$

$$\lambda = \frac{l}{\mathbf{Var}(S_n)} = \frac{x_\alpha \sqrt{\mathbf{Var}(S_n)}}{\mathbf{Var}(S_n)} = \frac{x_\alpha}{\sqrt{\mathbf{Var}(S_n)}} \Rightarrow$$

$$P_k = \mathbf{E}(\xi_k) + \underbrace{\lambda \mathbf{Var}(\xi_k)}_{l_k} = \mathbf{E}(\xi_k) + \frac{x_\alpha}{\sqrt{\mathbf{Var}(S_n)}} \mathbf{Var}(\xi_k).$$

с) *Принцип стандартного отклонения*

Полагают ( $l_k$  пропорционально  $\sqrt{\mathbf{Var}(\xi_k)}$ )

$$l_k = \lambda \sqrt{\mathbf{Var}(\xi_k)} \quad \left| \sum_{k=1}^n \right. \Rightarrow l = \lambda \sum_{k=1}^n \sqrt{\mathbf{Var}(\xi_k)} \Rightarrow$$

$$\lambda = \frac{l}{\sum_{k=1}^n \sqrt{\mathbf{Var}(\xi_k)}} = \frac{x_\alpha \sqrt{\mathbf{Var}(S_n)}}{\sum_{k=1}^n \sqrt{\mathbf{Var}(\xi_k)}} \Rightarrow$$

$$P_k = \mathbf{E}(\xi_k) + \underbrace{\lambda \sqrt{\mathbf{Var}(\xi_k)}}_{l_k} = \mathbf{E}(\xi_k) + \frac{x_\alpha \sqrt{\mathbf{Var}(S_n)}}{\sum_{k=1}^n \sqrt{\mathbf{Var}(\xi_k)}} \sqrt{\mathbf{Var}(\xi_k)}.$$

### 38. Коллективные модели риска. Составное пуассоновское распределение.

Рассматривается портфель договоров страхования за фиксированный относительно короткий период времени. Предполагается, что плата за страховку полностью вносится в начале рассматриваемого периода.

Пусть  $\nu$  – число страховых случаев по портфелю за заданный период времени,  $\eta_1 > 0$  – величина выплаты по первому страховому случаю,  $\eta_2 > 0$  – величина выплаты по второму случаю и т. д. Тогда суммарная величина страховых выплат по портфелю за заданный период времени равна

$$S_\nu = \eta_1 + \eta_2 + \dots + \eta_\nu, \quad S_0 = 0.$$

В этой сумме слагаемые и их количество суть случайные величины.

В дальнейшем для простоты изложения будем предполагать, что

- случайные величины  $\eta_1, \eta_2, \dots$  одинаково распределены (как  $\eta$ );
- случайные величины  $\nu, \eta_1, \eta_2, \dots$  в совокупности независимы.

Распределение  $S_\nu$

Положим

$$F(x) = F_\eta(x) = \mathbf{P}(\eta \leq x) \quad (= F_{\eta_k}(x), \quad k \geq 1).$$

Покажем, что

$$\begin{aligned} F_{S_\nu}(x) &= \mathbf{P}(S_\nu \leq x) = \sum_{n=0}^{\infty} \mathbf{P}(\eta_1 + \eta_2 + \dots + \eta_n \leq x) \mathbf{P}(\nu = n) = \\ &= \sum_{n=0}^{\infty} (F * F * F * \dots * F)(x) \mathbf{P}(\nu = n) = \sum_{n=0}^{\infty} F^{*n}(x) \mathbf{P}(\nu = n), \end{aligned} \quad (38.1)$$

где  $F^{*n}(x)$  –  $n$ -кратная свертка  $F(x)$ :

$$F^{*0}(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

$$F^{*1}(x) = F(x).$$

Доказательство формулы (38.1). Имеем

$$\{S_\nu \leq x\} = \bigcap_{n=0}^{\infty} \{S_\nu \leq x, \nu = n\}.$$

Отсюда

$$\begin{aligned} F_{S_\nu}(x) &= \mathbf{P}(S_\nu \leq x) = \sum_{n=0}^{\infty} \mathbf{P}(S_\nu \leq x | \nu = n) \mathbf{P}(\nu = n) \\ &= \sum_{n=0}^{\infty} \mathbf{P}(S_n \leq x) \mathbf{P}(\nu = n). \end{aligned}$$

Напомним, что для неотрицательной и непрерывной случайной величины  $\eta$  с функцией распределения  $F(x) = F_\eta(x)$  и плотностью  $f(x) = f_\eta(x)$   $n$ -кратные свертки  $F^{*n}(x)$  и  $f^{*n}(x)$  вычисляются по формулам

$$F^{*n}(x) = P(\eta_1 + \eta_2 + \dots + \eta_n \leq x) = \int_0^x F^{*(n-1)}(x-y) f(y) dy,$$

$$f^{*n}(x) = \int_0^x f^{*(n-1)}(x-y)f(y)dy.$$

Для неотрицательной и дискретной случайной величины  $\eta$  с законом распределения  $f_j = \mathbf{P}(\eta = j) = F_\eta(j) - F_\eta(j-1)$  имеем

$$\begin{aligned} F^{*n}(x) &= \mathbf{P}\left(\underbrace{\eta_1 + \eta_2 + \dots + \eta_n}_{=S_n} \leq x\right) = \\ &= \sum_{j=0}^x \mathbf{P}\left(\underbrace{\eta_1 + \eta_2 + \dots + \eta_{n-1}}_{=S_{n-1}} \leq x-j\right) \mathbf{P}(\eta_n = j) = \\ &= \sum_{j=0}^x F^{*(n-1)}(x-j) \mathbf{P}(\eta_n = j) = \sum_{j=0}^x F^{*(n-1)}(x-j) f_j. \end{aligned}$$

Закон распределения  $S_n$  находится по правилу

$$\mathbf{P}\left(\underbrace{\eta_1 + \eta_2 + \dots + \eta_n}_{=S_n} = x\right) = f_x^{*n} = \sum_{j=0}^x f_{x-j}^{*(n-1)} \mathbf{P}(\eta_n = j) = \sum_{j=0}^x f_{x-j}^{*(n-1)} f_j.$$

Здесь  $f^{*1} = f$ .

Расчет  $E(S_v)$  и  $Var(S_v) = V(S_v)$ . Справедливы следующие формулы:

$$\mathbf{E}(S_v) = \mathbf{E}(v)\mathbf{E}(\eta) = m_v m_\eta, \quad (38.2)$$

$$\mathbf{V}(S_v) = \mathbf{E}(v)\mathbf{V}(\eta) + \mathbf{E}(\eta)^2 \mathbf{V}(v) = m_v \sigma_\eta^2 + m_\eta^2 \sigma_v^2, \quad (38.3)$$

Для доказательства этих формул напомним некоторые факты из ТВ:

для случайных величин  $\xi_1, \xi_2$  имеем

$$\mathbf{E}(\xi_1) = \mathbf{E}(\mathbf{E}(\xi_1|\xi_2)), \quad (38.4)$$

$$\mathbf{V}(\xi_1) = \mathbf{E}(\mathbf{V}(\xi_1|\xi_2)) + \mathbf{V}(\mathbf{E}(\xi_1|\xi_2)). \quad (38.5)$$

Доказательство формулы (3.2). Имеем

$$\mathbf{E}(S_v|v = n) = \mathbf{E}(S_n) = \mathbf{E}\left(\sum_{k=1}^n \eta_k\right) = \sum_{k=1}^n \mathbf{E}(\eta_k) = n\mathbf{E}(\eta).$$

Отсюда

$$\mathbf{E}(S_v|v) = v\mathbf{E}(\eta).$$

Применяя формулу (3.4) при  $\xi_1 = S_v$ ,  $\xi_2 = v$ , находим

$$\mathbf{E}(S_v) = \mathbf{E}(\mathbf{E}(S_v|v)) = \mathbf{E}(v\mathbf{E}(\eta)) = \mathbf{E}(v)\mathbf{E}(\eta).$$

*Доказательство формулы (3.3). Имеем*

$$\mathbf{V}(S_v|v = n) = \mathbf{V}(S_n) = \mathbf{V}\left(\sum_{k=1}^n \eta_k\right) = \sum_{k=1}^n \mathbf{V}(\eta_k) = n\mathbf{V}(\eta).$$

Отсюда

$$\mathbf{V}(S_v|v) = v\mathbf{V}(\eta).$$

Применяя формулу (3.5) при  $\xi_1 = S_v$ ,  $\xi_2 = v$ , находим

$$\begin{aligned}\mathbf{V}(S_v) &= \mathbf{E}(\mathbf{V}(S_v|v)) + \mathbf{V}(\mathbf{E}(S_v|v)) = \mathbf{E}(v\mathbf{V}(\eta)) + \mathbf{V}(v\mathbf{E}(\eta)) = \\ &= \mathbf{E}(v)\mathbf{V}(\eta) + \mathbf{E}(\eta)^2\mathbf{V}(v).\end{aligned}$$

*Производящая функция моментов  $S_v$  суммарных страховых выплат.*

Покажем, что производящая функция моментов  $\psi_{S_v}(t) = E(e^{tS_v})$  СВ  $S_v$  имеет вид

$$\psi_{S_v}(t) = \psi_v(\ln \psi_\eta(t)).$$

Имеем

$$\begin{aligned}\mathbf{E}(e^{tS_v}|v = n) &= \mathbf{E}(e^{tS_n}) = \mathbf{E}(e^{t(\eta_1 + \dots + \eta_n)}) = \mathbf{E}(e^{t\eta_1} \dots e^{t\eta_n}) = \\ &= \mathbf{E}(e^{t\eta_1}) \dots \mathbf{E}(e^{t\eta_n}) = \psi_{\eta_1}(t) \dots \psi_{\eta_n}(t) = \psi_\eta(t)^n.\end{aligned}$$

Отсюда

$$\mathbf{E}(e^{tS_v}|v) = \psi_\eta(t)^v$$

и

$$\begin{aligned}\psi_{S_v}(t) &= \mathbf{E}(e^{tS_v}) = \mathbf{E}(\psi_\eta(t)^v) = \mathbf{E}(e^{\ln \psi_\eta(t)^v}) = \mathbf{E}(e^{v \ln \psi_\eta(t)}) \\ &= \psi_v(\ln \psi_\eta(t)).\end{aligned}$$

*Составное пуассоновское распределение.* Пусть  $v$  имеет пуассоновское распределение:  $v \sim \text{Poiss}(\lambda)$ . В этом случае говорят, что  $S_v$  имеет составное пуассоновское распределение с параметрами  $\lambda$  и  $F(x)$ . Коротко будем писать  $S_v \sim \text{CP}(\lambda, F(x))$ .

*Напоминание*

Пусть  $v \sim \text{Poiss}(\lambda)$ :

$$P(v = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \dots, \quad \lambda > 0.$$

Тогда

$$E(v) = V(v) = \lambda,$$

$$\psi_v(t) = e^{\lambda(e^t - 1)}.$$

Для составного пуассоновского распределения с параметрами  $\lambda$  и  $F(x)$  формулы (3.2) и (3.3) примут вид

$$E(S_v) = E(v)E(\eta) = \lambda E(\eta),$$

$$V(S_v) = E(v)V(\eta) + E(\eta)^2 V(v) = \lambda V(\eta) + \lambda E(\eta)^2 = \lambda E(\eta^2).$$

Кроме того

$$\psi_{S_v}(t) = e^{\lambda(\psi_\eta(t) - 1)}.$$

*Аддитивность независимых случайных величин с составным пуассоновским распределением*

Рассмотрим следующие две задачи:

1. Предположим, что мы скомбинировали  $m$  страховых портфелей, таких, что общий размер страховых выплат по каждому из этих портфелей имеет составное пуассоновское распределение. Можно ли что-либо сказать о новом (объединенном) портфеле? Здесь предполагается, что суммарные выплаты по разным портфелям независимы.

Ответ. Да. Объединенный портфель снова описывается составным пуассоновским распределением. А именно суммарная величина выплат объединенного портфеля имеет составное пуассоновское распределение (см. теорему ниже).

2. Предположим, что мы рассматриваем некоторый страховой портфель в течение  $m$  лет, и что годовые страховые выплаты за  $m$  лет независимы, причем каждая из них имеет составное пуассоновское распределение. Мы не требуем, чтобы распределения, соответствующие указанным выплатам, были одинаковы. Тогда общий размер страховых выплат за  $m$  летний период будет иметь составное пуассоновское распределение (см. теорему ниже).

*Теорема 3.1.* Пусть  $S_{v_1}, S_{v_2}, \dots, S_{v_m}$  взаимно независимые случайные величины, такие что  $S_{v_i} \sim \text{CP}(\lambda_i, F_i)$  для каждого  $i$ . Пусть  $S = S_{v_1} + \dots + S_{v_m}$ . Тогда  $S \sim \text{CP}(\lambda, F)$ , где

$$\lambda = \sum_{i=1}^m \lambda_i,$$

$$F(x) = \sum_{i=1}^m \frac{\lambda_i}{\lambda} F_i(x).$$

**39. Эмпирическое распределение для полных и сгруппированных данных. Огива и гистограмма. Оценка Нельсона-Алена кумулятивной функции доли риска.**

*Эмпирическое распределение для полных индивидуальных данных (т.е. точно зарегистрированных данных)*

Рассмотрим выборку  $x_1, x_2, \dots, x_n$  объема  $n$ . Напомним, что эмпирическое распределение получается путем назначения вероятности  $\frac{1}{n}$  каждому элементу выборки. Более точно

Эмпирическая функция распределения (ЭФР) определяется как

$$F_n(x) = \frac{\text{число наблюдений} \leq x}{n},$$

где  $n$  – это общее число наблюдений.

Пусть  $y_1 < y_2 \dots < y_k$  – различные элементы выборки  $x_1, x_2, \dots, x_n$ , где  $k \leq n$ ,  $s_k$  – частота  $y_k$  в выборке (количество появлений наблюдения  $y_j$  в выборке). Эмпирическая функция вероятностей определяется как

$$p_n(y_k) = \frac{s_k}{n}.$$

Имеем

$$\sum_{j=1}^k s_j = n.$$

Рассмотрим случайную величину, принимающую значения  $y_j$  с вероятностями  $s_j/n$ . Тогда  $F_n(x)$  есть функция распределения этой случайной величины.

Интерес представляет число наблюдений в наборе данных больше либо равных заданному значению. И наблюдения, и число наблюдений в этом случае принято называть множеством риска. Пусть

$$r_j = \sum_{i=j}^k s_i$$

– это число наблюдений, больше либо равных  $y_j$ . ЭФР можно переписать в виде

$$F_n(x) = \begin{cases} 0, & x < y_1, \\ 1 - \frac{r_j}{n}, & y_{j-1} \leq x < y_j, \quad j = 2, \dots, k, \\ 1, & x \geq y_k. \end{cases}$$

Заметим, что если наблюдаемая случайная величина непрерывного типа, то ее плотность и функция доли риска (определение этой функции дано ниже) не могут быть вычислены с помощью производной ЭФР. Ближайшая цель – получить эмпирическую оценку кумулятивной функции доли риска.

Напомним, что кумулятивная функция доли риска  $H_\xi(x)$  определяется как

$$H_\xi(x) = -\ln S_\xi(x),$$

где  $S_\xi(x) = 1 - F_\xi(x)$ .

Если  $S_\xi(x)$  дифференцируема, то

$$H_\xi'(x) = -\frac{S_\xi'(x)}{S_\xi(x)} = \frac{f_\xi(x)}{S_\xi(x)} = h_\xi(x) \text{ (функция доли риска)}$$

и

$$H_\xi(x) = \int_{-\infty}^x h_\xi(y) dy.$$



Функция распределения может быть получена так:

$$F_{\xi}(x) = 1 - S_{\xi}(x) = 1 - e^{-H_{\xi}(x)}.$$

Таким образом оценка кумулятивной функции доли риска предоставляет альтернативный способ оценки функции распределения.

*Эмпирическая оценка Нельсона – Алена функции  $H_{\xi}(x)$*

Эмпирическая оценка Нельсона – Алена кумулятивной функции доли риска вычисляется по формуле:

$$H_{NA}(x) = \begin{cases} 0, & x < y_1, \\ \sum_{i=1}^{j-1} \frac{S_i}{r_i}, & y_{j-1} \leq x < y_j, \quad j = 2, \dots, k, \\ \sum_{i=1}^k \frac{S_i}{r_i}, & x \geq y_k. \end{cases}$$

Т.к.  $H_{\xi}(x)$  - ступенчатая функция, то ее производные (которые могли бы дать оценку функции доли риска) не могут быть вычислены непосредственно.

*Эмпирическое распределение для сгруппированных данных*

Для сгруппированных данных построить эмпирическое распределение, как было определено в случае индивидуальных данных, невозможно. Тем не менее можно построить приближение эмпирического распределения.

Пусть

$$(c_0, c_1], (c_1, c_2], \dots, (c_{k-2}, c_{k-1}], (c_{k-1}, c_k]$$

– интервалы сгруппированных данных, где  $c_0 < c_1 < \dots < c_k$  – границы групп. Обычно  $c_0 = 0, c_k = \infty$  (в этом случае вместо полуоткрытого интервала  $(c_{k-1}, c_k]$  рассматривается открытый интервал  $(c_{k-1}, c_k)$ ).

Обозначим через  $n_j$  – число наблюдений, которые попадают в интервал  $(c_{j-1}, c_j]$ . Пусть

$$n = \sum_{j=1}^k n_j$$

– общее число наблюдений.

Для таких данных мы можем определить эмпирическое распределение в каждой граничной точке по правилу

$$F_n(c_j) = \frac{1}{n} \cdot \sum_{i=1}^j n_i, \quad j = 1, 2, \dots, k,$$

где  $F_n(c_0) = 0$ .

Для получения оценки ЭФР  $F_n(x)$  необходимо для каждого интервала  $j = 1, 2, \dots, k$  интерполировать  $F_n(x)$  внутри каждого интервала  $(c_{j-1}, c_j]$ . Если применить линейную интерполяцию внутри каждого интервала  $(c_{j-1}, c_j]$ , то значения  $F_n(c_{j-1})$ ,  $F_n(c_j)$  соединяются отрезком. Тогда

$$F_n(x) = \frac{c_j - x}{c_j - c_{j-1}} \cdot F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}} \cdot F_n(c_j), \quad c_{j-1} \leq x \leq c_j.$$

Если  $c_k = \infty$ , то  $F_n(x)$  неопределена при  $c_{k-1} < x < \infty$ .

График ЭФР  $F_n(x)$  называется *огивой*.

$F_n(x)$  может быть переписана в виде

$$F_n(x) = \frac{n_j}{n(c_j - c_{j-1})} x + \frac{c_j F_n(c_{j-1}) - c_{j-1} F_n(c_j)}{c_j - c_{j-1}}, \quad c_{j-1} \leq x \leq c_j.$$

Функция  $F_n(x)$  дифференцируема во всех точках, за исключением граничных точек. Эмпирическая функция плотности (ЭФП) имеет вид (считаем ее непрерывной справа)

$$f_n(x) = \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} = \frac{n_j}{n \cdot (c_j - c_{j-1})}, \quad c_{j-1} \leq x < c_j$$

График ЭФП  $f_n(x)$  называется *гистограммой*.

Если  $c_k = \infty$ , то  $f_n(x)$  неопределена при  $c_{k-1} < x < \infty$ .

Пусть  $\xi$  – случайная величина, ассоциированная с данными (например, выплаты страховой компании по портфелю договоров страхования). Тогда оценка ожидаемой выплаты равна (при  $c_k \neq \infty$ )

$$\hat{\mathbf{E}}(\xi) = \frac{1}{n} \sum_{j=1}^k \left( \frac{c_j + c_{j-1}}{2} \right) n_j,$$

ПОСКОЛЬКУ

$$\hat{\mathbf{E}}(\xi) = \int_{c_0}^{c_k} x f_n(x) dx = \sum_{j=1}^k \int_{c_{j-1}}^{c_j} \frac{n_j x}{n(c_j - c_{j-1})} dx = \frac{1}{n} \sum_{j=1}^k \left( \frac{c_j + c_{j-1}}{2} \right) n_j.$$

Рассмотрим договор страхования с лимитом  $u$ . Если лимит полиса  $u = c_m$  (т.о. лимит – это граница класса), то оценка ожидаемой выплаты равна

$$\hat{\mathbf{E}}(\xi \wedge u) = \int_0^u x f_n(x) dx + u(1 - F_n(u)) = \frac{1}{n} \left( \sum_{j=1}^m \frac{c_j + c_{j-1}}{2} n_j + c_m \sum_{j=m+1}^k n_j \right)$$

ПОСКОЛЬКУ

$$\begin{aligned} & \int_0^u x f_n(x) dx + u(1 - F_n(u)) = \\ &= \sum_{j=1}^m \int_{c_{j-1}}^{c_j} x \frac{n_j}{n(c_j - c_{j-1})} dx + c_m \left( 1 - \frac{1}{n} \cdot \sum_{i=1}^m n_i \right) = \\ &= \frac{1}{n} \left( \sum_{j=1}^m \frac{c_j + c_{j-1}}{2} n_j + c_m \sum_{j=m+1}^k n_j \right), \end{aligned}$$

либо

$$\begin{aligned} & \int_0^u x f_n(x) dx + u(1 - F_n(u)) = \\ &= \sum_{j=1}^m \int_{c_{j-1}}^{c_j} x \frac{n_j}{n(c_j - c_{j-1})} dx + c_m \sum_{j=m+1}^k \int_{c_{j-1}}^{c_j} \frac{n_j}{n(c_j - c_{j-1})} dx = \\ &= \frac{1}{n} \left( \sum_{j=1}^m \frac{c_j + c_{j-1}}{2} n_j + c_m \sum_{j=m+1}^k n_j \right). \end{aligned}$$

Если лимит полиса  $u \in (c_{m-1}, c_m)$ , то оценка ожидаемого значения выплат следующая:

$$\begin{aligned}\hat{\mathbf{E}}(\xi \wedge u) &= \int_0^u x f_n(x) dx + u(1 - F_n(u)) = \\ &= \sum_{j=1}^{m-1} \int_{c_{j-1}}^{c_j} x \frac{n_j}{n(c_j - c_{j-1})} dx + \\ &+ u \left( 1 - \frac{n_m}{n(c_m - c_{m-1})} u - \frac{c_m F_n(c_{m-1}) - c_{m-1} F_n(c_m)}{c_m - c_{m-1}} \right).\end{aligned}$$

Если  $c_k \neq \infty$ , то получим следующую оценку для  $\mathbf{E}(\xi - d)_+$

$$\hat{\mathbf{E}}(\xi - d)_+ = \hat{\mathbf{E}}(\xi) - \hat{\mathbf{E}}(\xi \wedge d).$$

Либо

$$\hat{\mathbf{E}}(\xi - d)_+ = \int_d^{c_k} (x - d) f_n(x) dx.$$

Здесь также различаются два случая в зависимости от того вычит  $d$  совпадает с граничной точкой или нет.

Аналогично оцениваются другие числовые характеристики СВ  $\xi$ ,  $\xi \wedge u$  и  $(\xi - d)_+$  (например, моменты, дисперсия, коэффициенты вариации и асимметрии и т.д.).

#### **40. Эмпирическое распределение для цензурированных и усеченных данных. Оценка Каплана-Мейера.**

*Определение.* Наблюдение называется *усеченным снизу* (или *усеченным слева*) относительно точки  $d$ , если когда оно меньше либо равно  $d$ , оно не регистрируется, но когда оно больше  $d$ , оно регистрируется с наблюдаемым значением. Наблюдение называется *усеченным сверху* (или *усеченным справа*) относительно точки  $u$ , если когда оно больше либо равно  $u$ , оно не регистрируется, но когда оно меньше  $u$ , оно регистрируется с наблюдаемым значением.

*Определение.* Наблюдение называется *цензурированным снизу* (или *цензурированным слева*) относительно точки  $d$ , если когда оно меньше либо равно  $d$ , оно регистрируется как  $d$ , но когда оно больше  $d$ , оно регистрируется с наблюдаемым значением. Наблюдение называется *цензурированным сверху* (или *цензурированным справа*) относительно точки  $u$ , если когда оно больше либо равно  $u$ , оно регистрируется как  $u$ , но когда оно меньше  $u$ , оно регистрируется с наблюдаемым значением.

Так как усечение слева и цензурирование справа – наиболее распространенные явления в актуарной практике, то именно эти случаи будут освещены здесь. Далее «усеченный» будет означать «усеченный снизу», а «цензурированный» – «цензурированный сверху».

Левое усечение применяется, когда используется обычная франшиза. Когда величина потерь держателя страхового полиса ниже  $d$ , он знает, что страховая компания не будет платить ничего, и не сообщает о ней страховщику, но когда она выше  $d$ , он сообщает об этом страховщику.

Рассмотрим пример цензурирования справа - лимитированный полис. Если величина потерь страхователя ниже  $u$ , страховщик платит полную сумму и платит только  $u$  в противном случае.

Когда строиться таблица продолжительности жизни, не имеет смысла рассматривать людей от рождения до смерти. Принято рассматривать группы людей различных возрастов в течение нескольких лет. Когда человек включается в рассмотрение, он (или она) жив (жива) в этот момент. Возраст наступления смерти человека должен быть по крайней мере больше возраста его включения в рассмотрение, т.е. является усеченным слева. Если человек жив к завершению наблюдения, то происходит цензурирование справа, т.е. возраст наступления смерти человека неизвестен, но известно, что он больше возраста, которого достиг человек к завершению наблюдения. Цензурирование справа также затрагивает тех, кто исключается из наблюдения до его завершения по причине сдачи полиса.

Для построения эмпирического распределения для усеченных или цензурированных данных необходимо учесть *три фактора*.

*Первый* – это точка, относительно которой происходит усечение данных для данного наблюдения. Обозначим это значение  $d_i$  для  $i$ -го наблюдения. Если не было усечения, то  $d_i = 0$ .

*Второй фактор* – это само наблюдение. Используемая переменная будет зависеть так или иначе от того, было ли наблюдение цензурировано. Если не было цензурирования, обозначим его значение  $x_i$ . Если же наблюдение было цензурировано, обозначим его значение  $u_i$ . Когда такая задача представляется более формально, различие делается между случаем, когда точка цензурирования известна заранее, и случаем, когда она неизвестна. Например, при рассмотрении полиса страхования гражданской ответственности с лимитом, точка цензурирования обычно известна до поступления каких-либо требований. С другой стороны, при исследовании смертности застрахованных жизней, те, кто сдают свой полис, делают это в возрасте, который заранее (в момент продажи полиса) не был известен. Дальше мы не делаем никакого различия между этими двумя случаями.

*Третий фактор* – для построения оценки необработанные данные должны быть представлены в удобной форме. Наиболее интересные значения останутся нецензурированными наблюдениями. Пусть  $y_1 < \dots < y_k$  – это  $k$  различных значений для  $x_i$ , которые появляются в выборке,  $k$  меньше либо равно числу нецензурированных данных. Пусть  $s_i$  – это число нецензурированных наблюдений  $y_j$  в выборке. И наконец важная величина – множество риска для  $j$ -го наблюдения  $y_j$  определяется по правилу

$$r_j = (\text{количество } x_i \geq y_j) + (\text{количество } u_i \geq y_j) - (\text{количество } d_i \geq y_j). \quad (40.1)$$

Т.к. общее количество  $d_i$  равно общему количеству  $u_i$  и  $x_i$ , то получим:

$$r_j = (\text{количество } d_i < y_j) - (\text{количество } x_i < y_j) -$$

$$-(\text{количество } u_i < y_j). \quad (40.2)$$

Последняя формула немного проще для осмысления.

В терминах данных смертности или дожития  $r_j$  включает всех, кто уже включен в рассмотрение до указанного возраста или момента. Ключевым моментом является то, что множество риска – это число наблюдаемых людей, доживших до возраста  $y_j$  или число наблюдаемых живых людей в момент  $y_j$ .

Если в роли данных выступают данные о потерях, то множество риска – это число полисов с исследуемым количеством потерь (и текущее количество и максимальное количество потерь соответствуют сроку полиса) больше либо равных  $y_j$  и меньше тех, у кого франшиза больше либо равна  $y_j$ .

#### *Оценка Каплана-Мейера функции дожития*

Пусть  $T$  – продолжительность предстоящей жизни индивида или размер ущерба.

$S(0) = 1$ . Т.к., например, никто не умирает до момента  $y_1$ , то функция дожития принимает значение равное 1 до этого момента.

Далее можно провести следующие рассуждения, основанные на условных вероятностях событий. Как раз перед моментом  $y_1$  найдется  $r_1$  людей, которые могут умереть, и из которых  $s_1$  точно умрет. Т.о. вероятность дожития до момента  $y_1$  равна  $\frac{r_1 - s_1}{r_1}$ , и это значение равно  $S(y_1)$ :

$$S_n(y_1) = \mathbf{P}(T > y_1) = \frac{r_1 - s_1}{r_1} = 1 - \frac{s_1}{r_1}.$$

Аналогично, вероятность для живых в момент  $y_1$  дожить до момента  $y_2$  есть

$$\mathbf{P}(T > y_2 | T > y_1) = \frac{r_2 - s_2}{r_2} = 1 - \frac{s_2}{r_2}.$$

Отсюда

$$S_n(y_2) = \mathbf{P}(T > y_2) = \mathbf{P}(T > y_1) \mathbf{P}(T > y_2 | T > y_1) = \left(1 - \frac{s_1}{r_1}\right) \left(1 - \frac{s_2}{r_2}\right).$$

Далее

$$\begin{aligned} S_n(y_3) &= \mathbf{P}(T > y_3) = \mathbf{P}(T > y_1)\mathbf{P}(T > y_2|T > y_1)\mathbf{P}(T > y_3|T > y_2) = \\ &= \left(1 - \frac{s_1}{r_1}\right)\left(1 - \frac{s_2}{r_2}\right)\left(1 - \frac{s_3}{r_3}\right). \end{aligned}$$

и т.д.

В результате мы получим оценку Каплана-Мейера функции дожития

$$S_{KM}(t) = \begin{cases} 1, & 0 \leq t < y_1, \\ \prod_{i=1}^{j-1} \left(1 - \frac{s_i}{r_i}\right), & y_{j-1} \leq t < y_j, j = 2, \dots, k, \\ \prod_{i=1}^k \left(1 - \frac{s_i}{r_i}\right), & t \geq y_k. \end{cases}$$

Если  $s_k = r_k$ , то  $S(t) = 0$  для  $t \geq y_k$ , поскольку каждый из выборки умирает после указанного значения, и эмпирически остаться живым после этого момента невозможно. Однако, благодаря цензурированию, возможно, что на момент возраста последней смерти останутся тем не менее живые люди, но они уже были цензурированы до этого момента смерти. Мы знаем, что остаться живым после возраста наступления последней смерти возможно, но нет эмпирических данных, которые могли бы дополнить функцию дожития.

*Первый способ* решить эту проблему (этот способ используется в вышеприведенной формуле) – это оставить значение функции дожития равным ее последнему значению.

*Второй способ* – это объявить функцию дожития равной нулю после последнего наблюдаемого возраста, который представляет собой либо возраст наступления смерти, либо возраст, относительно которого происходит цензурирование. Это менее разумный способ (хотя он дает возможность вычислять моменты). *Промежуточный вариант* – это использовать убывающую экспоненциальную функцию от текущего значения до нуля. Пусть  $w = \max(x_1, \dots, x_n, u_1, \dots, u_n)$ . Тогда при  $t \geq w$  имеем:



$$S_{KM}(t) = e^{\left(\frac{t}{w}\right) \ln s^*}, \text{ где } s^* = \prod_{i=1}^k \left(1 - \frac{s_i}{r_i}\right).$$

**41. Оценка параметров распределений для полных данных. Метод моментов, метод эмпирических квантилей и метод максимального правдоподобия. Метод моментов для цензурированных данных.**

*Метод моментов и эмпирических квантилей*

Будем предполагать, что все  $n$  наблюдений  $x_1, \dots, x_n$  относятся к одному параметрическому семейству распределения. В частности, пусть функция распределения задается следующим образом:

$$F(x; \boldsymbol{\theta}), \boldsymbol{\theta}^T = (\theta_1, \theta_2, \dots, \theta_l),$$

где  $\boldsymbol{\theta}$  – вектор, состоящий из  $l$  параметров, которые нужно оценить.

*Метод моментов*

Обозначим теоретические моменты для соответствующей СВ  $\xi$  (генеральной совокупности) через

$$\mu_k(\boldsymbol{\theta}) = \mathbf{E}(\xi^k) \text{ (} k \text{ – й момент),}$$

а эмпирические или выборочные моменты через

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

(выборочный  $k$  – й момент или эмпирическая оценка  $k$  – го момента).

*Определение.* Оценка по методу моментов для параметра  $\boldsymbol{\theta}$  – это решение системы из  $l$  уравнений:

$$\mu_k(\boldsymbol{\theta}) = \hat{\mu}_k, \quad k = \overline{1, l}. \quad (41.1)$$

Традиционное определение метода моментов использует положительные целые значения для моментов. Но отрицательные или дробные моменты тоже могут быть использованы. В частности, когда оцениваются параметры обратного экспоненциального распределения,

отрицательные моменты могут служить хорошей оценкой (обычные моменты для этого распределения не существуют).

Нет никакой гарантии, что система уравнений будет иметь решение или, если решение существует, то оно будет единственным.

Если выборка получена после цензурирования по значению  $u$ , то (41.1) понимается так

$$\mu_k(\theta) = \mathbf{E}(\xi \wedge u)^k,$$

а  $\hat{\mu}_k$  – выборочный  $k$ -й момент для (точечной) выборки после цензурирования.

*Метод эмпирических квантилей*

Для непрерывных распределений квантиль уровня  $g$  находится как решение уравнения

$$F_{\xi}(x_g; \theta) = g.$$

Оценка методом эмпирических квантилей для параметра  $\theta$  – это решение системы из  $l$  уравнений вида:

$$F(\hat{\pi}_{g_k} | \theta) = g_k, k = \overline{1, l}.$$

Мотивация для этой оценки состоит в том, что она дает модель с  $l$  квантилями, которые "подгоняют" данные (как если бы они были представлены эмпирическим распределением).

Как и с методом моментов, нет никакой гарантии, что система уравнений будет иметь решение или, даже если решение есть, то оно будет единственным.

Единственная проблема с этим определением – это то, что квантили для дискретной случайной величины не всегда четко определены.

*Определение.* Сглаженная эмпирическая оценка квантиля находится следующим образом:

$$\hat{\pi}_g = (1 - h) \cdot x_{(j)} + h \cdot x_{(j+1)}, j = [(n + 1)g], h = (n + 1)g - j.$$

В этих обозначениях  $[x]$  обозначает целую часть числа  $x$  и  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  – это порядковые статистики из выборки.

За исключением случая, когда есть два и более данных с одинаковыми значениями, никакие два квантиля не совпадают. Одно из свойств этого определения – это то, что  $\hat{\pi}_g$  не может быть получено для:

$$g < \frac{1}{1+n},$$

и

$$g > \frac{n}{1+n},$$

т.к. предполагается, что нереально получить значение для большого или маленького квантиля для маленькой выборки.

#### *Метод максимального правдоподобия (ММП)*

Оценки, полученные по методу моментов (ОММ) или эмпирических квантилей склонны давать не слишком хороший результат. Основная причина этого – они используют только некоторые свойства данных, а не полный набор наблюдений.

Особенно важно использовать как можно больше информации, если генеральная совокупность имеет правый тяжелый хвост. Например, когда оцениваются параметры для нормального распределения, выборочное среднее и вариация вполне приемлемы. Но когда оцениваются параметры распределения Парето, важно знать все предельные наблюдения для того, чтобы правильно оценить параметр  $\alpha$ .

Другой недостаток этих методов – это то, что они требуют, чтобы все наблюдения относились одной и той же случайной величине. Если это будет не так, то тогда неясно, что брать в качестве моментов или квантилей генеральной совокупности. Например, если половина наблюдений имеет франшизу 50, а другая половина – франшизу 100, то неясно, к чему приравнивать выборочное среднее.

Наконец, эти методы позволяют аналитику принимать случайные решения на основе используемых квантилей или моментов.

Существует много оценок, полученных на основе индивидуальных данных. Все они реализуются с помощью целевой функции, а затем определяются параметры, которые оптимизируют эту функцию. Например, можно было бы оценить параметры с помощью минимизации функции, представляющей разность между функцией распределения параметрической модели и оценкой Нельсона – Алена для функции распределения.

Среди огромного разнообразия оценок будем использовать здесь только оценку максимального правдоподобия (ОМП).

Пусть набор данных состоит из  $n$  событий  $A_1, \dots, A_n$ , где  $A_j$  – это результат  $j$ -го наблюдения. Например,  $A_j$  может состоять из единственной точки или интервала. Последнее происходит в случае сгруппированных или цензурированных данных. Например, если было цензурирование по значению  $u$  и цензурированное наблюдение произошло, то событием будет интервал от  $u$  до бесконечности.

Далее предположим, что событие  $A_j$  – это результат наблюдения СВ  $\xi_j$ . Случайные величины  $\xi_1, \dots, \xi_n$  не обязательно должны быть одинаково распределены, но их распределения должны зависеть от одного и того же параметрического вектора  $\theta$ .

*Определение.* Функция правдоподобия – это функция вида:

$$L(\theta) = \prod_{j=1}^n P(\xi_j \in A_j; \theta)$$

ОМП параметра  $\theta^T = (\theta_1, \theta_2, \dots, \theta_l)$  – это вектор, максимизирующий функцию правдоподобия (ФП).

Заметим, что нет никакой гарантии, что функция правдоподобия имеет максимум в подходящих (приемлемых) значениях параметра. Возможно, что в зависимости от того, когда различные параметры стремятся к нулю или бесконечности, функция правдоподобия будет возрастать. Нужно быть осторожным при максимизации ФП, т.к. могут быть локальные максимумы в дополнение к глобальным.

Часто не удастся аналитически максимизировать ФП (вычислив частные производные и приравняв их к нулю). Но чаще используются численные методы, дающие приближительный ответ.

Т.к. наблюдения предполагаются независимыми, произведение вероятностей в определении функции правдоподобия представляет совместную вероятность  $P(\xi_1 \in A_1, \dots, \xi_n \in A_n; \theta)$ , т.е. ФП – это вероятность получения результатов выборки при конкретном значении параметра.

Часто вместо функции правдоподобия  $L(\theta)$  используют логарифмическую функцию правдоподобия  $l(\theta) = \ln L(\theta)$  (ее легче максимизировать). Так как функция  $\ln x$  монотонно растет при  $x > 0$ , то максимумы этих функций совпадают.

#### *Метод максимального правдоподобия для полных данных*

Когда нет ни цензурирования, ни усечения данных и записаны значения всех наблюдений, легко выписать ФП:

$$L(\theta) = \prod_{j=1}^n f_{\xi_j}(x_j; \theta),$$

$$l(\theta) = \sum_{j=1}^n \ln f_{\xi_j}(x_j | \theta).$$

Из такой записи видно, что не обязательно, чтобы все наблюдения имели одинаковое распределение.

## **42. Оценка максимального правдоподобия для полных сгруппированных, усеченных и цензурированных данных.**

### *ОМП для полных сгруппированных данных*

Напомним некоторые обозначения для полных сгруппированных данных.

Граничные точки группировки – множество чисел  $c_0 < c_1 < \dots < c_k$ , где  $c_0$  – наименьшее возможное наблюдение (часто  $c_0 = 0$ ) и  $c_k$  – наибольшее возможное наблюдение (часто  $c_k = \infty$ ).

Пусть  $n_j$  – число наблюдений, принадлежащих интервалу  $(c_{j-1}, c_j] = A_j$ .

Для таких данных ФП и ЛФП имеют соответственно вид:

$$L(\boldsymbol{\theta}) = \prod_{j=1}^k [F(c_j; \boldsymbol{\theta}) - F(c_{j-1}; \boldsymbol{\theta})]^{n_j},$$

$$l(\boldsymbol{\theta}) = \sum_{j=1}^k n_j \ln[F(c_j; \boldsymbol{\theta}) - F(c_{j-1}; \boldsymbol{\theta})].$$

*ОМП для усеченных и цензурированных данных*

*Цензурирование справа полных данных.* Цензурирование справа по значению  $u$  просто создает интервал, начинающийся от точки цензурирования и до бесконечности:

$$\mathbf{P}(\xi_j \in A_j) = \mathbf{P}(\xi_j > u) = 1 - F(u; \boldsymbol{\theta}) = S(u; \boldsymbol{\theta}).$$

Отсюда ( $k$  значений без цензурирования,  $n_u$  – количество цензурированных данных по значению  $u$ )

$$L(\boldsymbol{\theta}) = (1 - F(u; \boldsymbol{\theta}))^{n_u} \prod_{j=1}^k f_{\xi_j}(x_j; \boldsymbol{\theta}),$$

$$l(\boldsymbol{\theta}) = n_u \ln(1 - F(u; \boldsymbol{\theta})) + \sum_{j=1}^k \ln f_{\xi_j}(x_j; \boldsymbol{\theta}).$$

*Усечение слева полных данных*

Рассмотрим договор страхования с вычетом. Если для этого договора страховщиком зарегистрирована выплата  $y > 0$ , то это означает, что потери страхователя составляют  $x = y + d > d$ . Соответствующий множитель для  $y$  в ФМП есть (здесь  $\xi$  – случайная величина реальных потерь страхователя, т.е. выплат по договорам без вычета)

$$\frac{f_{\xi}(y + d; \boldsymbol{\theta})}{1 - F_{\xi}(d; \boldsymbol{\theta})} = \frac{f_{\xi}(y + d; \boldsymbol{\theta})}{S_{\xi}(d; \boldsymbol{\theta})}.$$

В самом деле выплаты страховщика описываются СВ

$$(\xi - d)_+^+ = \xi - d | \xi > d.$$

Отсюда при  $y > 0$  (для  $y \leq 0$  СВ  $(\xi - d)_+^+$  не определена) имеем

$$\begin{aligned} F_{(\xi-d)_+^+}(y; \theta) &= \mathbf{P}(\xi - d \leq y | \xi > d) = \\ &= \frac{\mathbf{P}(\xi \leq y + d)}{\mathbf{P}(\xi > d)} = \frac{F_\xi(y + d; \theta)}{1 - F_\xi(d; \theta)} = \frac{F_\xi(y + d; \theta)}{S_\xi(d; \theta)} \end{aligned}$$

Если случайная величина  $\xi$  непрерывного типа, то и  $(\xi - d)_+^+$  непрерывного типа с плотностью

$$f_{(\xi-d)_+^+}(y) = \frac{d}{dy} F_{(\xi-d)_+^+}(y; \theta) = \frac{f_\xi(y + d; \theta)}{1 - F_\xi(d; \theta)}.$$

Если требуется оценить параметры для известного распределения усеченных выплат, то ОМП сводится к ОМП полных данных (полученных после усечения).

*Данные регистрируются с учетом вычета  $d$  и лимита  $u$*

Пусть сначала применяется лимит, а затем вычет. Тогда максимальная возможная выплата равна  $u - d$ .

Если зарегистрированная выплата  $y < u - d$ , то к убыткам страхователя применяется только вычет, и убытки страхователя составляют  $x = y + d$ , и это приводит к множителю

$$\frac{f_\xi(y + d; \theta)}{1 - F_\xi(d; \theta)} = \frac{f_\xi(y + d; \theta)}{S_\xi(d; \theta)}$$

в ФМП.

Если  $y = u - d$ , то к убыткам страхователя применяется вычет и лимит, и убытки страхователя точно неизвестны, известно только, что  $x \geq u$  и это приводит к множителю

$$\frac{1 - F_\xi(u; \theta)}{1 - F_\xi(d; \theta)} = \frac{S_\xi(u; \theta)}{S_\xi(d; \theta)}$$

в ФМП.

#### **43. Модель ценообразования финансовых активов (САРМ), ее предпосылки и роль в анализе доходности акций.**

Одной из известных теорий рынка капитала является модель оценивания финансовых активов, которая относится к равновесным моделям, то есть при заданных предположениях о рынке и поведении инвесторов модель определяет теоретическую (или равновесную) стоимость актива.

**Capital Asset Pricing Model (САРМ)**, или Модель Оценки Финансовых Активов— это равновесная модель ценообразования, согласно которой ожидаемая доходность ценной бумаги является линейной функцией чувствительности доходности бумаги к изменению доходности рыночного портфеля.

Модель САРМ оценивает чувствительность доходности конкретного финансового актива к систематическому (рыночному) риску, мерой этой чувствительности является *бета-коэффициент*.

Двумя другими параметрами модели являются доходность безрискового актива и доходность рыночного портфеля. Эта модель применима для оценки требуемой доходности по акциям

Модель САРМ была впервые предложена Уильямом Шарпом в 1964 году и была основана на портфельной теории Г.Марковица.

Модель оценивания финансовых активов базируется на теории портфеля и предполагает выполненными ряд условий.

*Предпосылки модели САРМ:*

1. Инвесторы оценивают инвестиционные портфели, основываясь на ожидаемых доходностях и их стандартных отклонениях за период владения.

1. Инвесторы никогда не бывают пресыщенными, то есть при выборе между двумя портфелями они предпочтут тот, который при прочих равных условиях дает наибольшую ожидаемую доходность.



2. Инвесторы не желают рисковать и при выборе между двумя портфелями предпочтут тот, который имеет наименьшее стандартное отклонение.

3. Активы считаются бесконечно делимы, и при желании инвестор может купить часть акции.

4. Существует безрисковая процентная ставка, одинакова для всех инвесторов.

5. Налоги и транзакционные издержки несутся (отсутствуют).

6. Для всех инвесторов временной период вложений одинаков.

7. Информация свободно и незамедлительно доступна для всех инвесторов.

8. Инвесторы имеют однородные (гомогенные) ожидания, то есть одинаково оценивают ожидаемые доходности, стандартные отклонения и ковариации доходностей ценных бумаг.

9. Нормальное распределение случайных величин доходностей ценных бумаг.

В модели CAPM зависимость между риском и ожидаемой доходностью описывается с помощью линии рынка капитала CML (Capital Market Line).

Модель CAPM дает представление о том, какое должно быть соотношение между риском вложения в актив и доходностью этого вложения. Эта формула нашла широкое применение в теории современного инвестиционного анализа в самых различных его областях: оценки прибыльности проектов, портфельных инвестиций.

Модель оценки финансовых активов при анализе рынка акций дает хорошие результаты и является инструментом прогнозирования издержек по привлечению капитала для реализации инвестиционных проектов. В условиях хорошо развитого финансового рынка новая информация находит

быстрое отражение в курсовой стоимости ценных бумаг, а модель CAPM описывает взаимосвязь между риском и ожидаемой доходностью актива.

Крупнейшие мировые рыночные институты используют потенциал, заложенный в данную модель, и регулярно рассчитывают  $\beta$ -коэффициенты всех крупных компаний, котирующихся на фондовых биржах.

В области финансов, модель ценообразования финансовых активов (CAPM) используется для определения соответствующей требуемой нормы доходности актива в условиях, что этот актив добавляется к уже существующему хорошо диверсифицированному портфелю с учетом несистематического риска этого актива.

Модель CAPM представляет собой идеальную модель рынка капиталов, которая основывается на предположениях портфельной теории и исходит из равной информированности инвесторов относительно доходности и рискованности ценных бумаг.

В условиях модели CAPM справедлива теорема о разделении, в соответствии с которой оптимальный портфель рискованных активов одинаков для всех инвесторов и соответствует по структуре рыночному портфелю - совокупности всех рискованных активов, представленных на рынке. Индивидуальные портфели различаются лишь пропорциями безрисковых вложений и инвестиций в рыночный портфель.

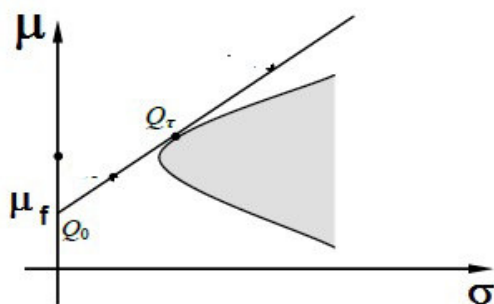
#### **44. Касательный портфель в модели ценообразования финансовых активов (CAPM).**

Модель CAPM описывает взаимосвязь между доходностью актива или портфеля и его риском. Предполагается, что функции полезности инвесторов зависят от ожидаемой доходности и риска портфелей, причем инвесторы одинаково оценивают ожидаемые доходности и риски отдельных активов. Существует безрисковый актив, который отвечает возможностью инвесторам брать займы или давать в долг финансовые ресурсы по безрисковой процентной ставке. Операционные издержки отсутствуют, а рынок находится в равновесии.

Портфель называется *рыночным*, если доля средств, вложенных в каждый актив портфеля, равна доле средств, вложенных в этот актив.

Рыночный портфель в предложениях модели CAPM является касательным портфелем. Всевозможные комбинации касательного портфеля с безрисковым активом образуют множество эффективных портфелей. На практике нет возможности определить реальный рыночный портфель, состоящий из всех активов, имеющих хождение на рынке. Поэтому инвесторы часто используют его подобие, а при операциях с обыкновенными акциями инвесторы заменяют рыночный портфель представительными индексами, например S&P 500.

Критериальное множество (в «финансовых» координатах риск  $\sigma$ , доходность  $\mu$ ) стандартной модели является частью плоскости, ограниченной парой лучей с вершиной соответствующей оценке безрискового актива. Внутри этого множества находится критериальное множество модели Блека, соответствующей рисковому части стандартной модели (т.е. образованной ее рисковыми активами), которая представляет собой часть критериальной плоскости, ограниченную ветвями гиперболы. Критериальное множество рискованной части и минимальную границу модели, образованную двумя лучами, выходящими из точки соответствующей безрисковому активу  $\mu_f$ , можно представить:



Верхний луч  $Q_0Q_\tau$ , представляющий собой эффективное множество стандартной модели, касается верхней ветви гиперболы — эффективной границы рискованных активов в точке, соответствующей оценке  $Q_\tau$  касательного портфеля  $\tau$ .

Из совпадения оценок параметров рынка всеми инвесторами (предпосылка модели CAPM) следует, что критериальные множества, их границы (как минимальная, так и эффективная) и касательные портфели так же будут совпадать для всех инвесторов. Поскольку в силу условия рациональности, каждый инвестор выбирает оптимальные по Марковицу портфели, то эти портфели будут эффективными. Оценки этих портфелей для всех инвесторов будут лежать на одной и той же эффективной границе – верхнем луче  $QOQt$ . На этом же луче лежат оценки двух фиксированных портфелей: безрискового  $f$  и касательного  $t$

Оптимальный портфель любого инвестора является линейной комбинацией всего двух фиксированных портфелей – безрискового и касательного (рискового) портфелей. На плоскости  $\mu, \sigma$  множество эффективных портфелей изображается полупрямой, проходящей через точки, отвечающие безрисковому и касательному портфелю. В теории модели оценки ценообразования финансовых активов эта прямая называется рыночной линией капитала (*capital market line, CML*). Эта прямая линия отражает взаимосвязь между доходностью и риском эффективных портфелей и проходит через оценки безрискового и касательного портфеля, ее уравнение можно записать в виде:

$$\mu = \mu_f + \frac{\mu_t - \mu_f}{\sigma_t} \cdot \sigma,$$

где  $\mu_t$  - ожидаемая доходность касательного портфеля;

$\mu_f$  - ожидаемая доходность безрискового портфеля;

$\mu$  - ожидаемая доходность произвольного эффективного портфеля;

$\sigma_t$  – риск касательного портфеля;

$\sigma$  – риск произвольного эффективного портфеля;

$\sigma_f = 0$ .

Рыночная линия капитала соотносит риск и доходность эффективных портфелей. Такая же связь (не обязательно эффективного) составляет основной результат теории CAPM.

Рассмотрим структуру касательного портфеля. Пусть на рынке обращается  $n$  активов  $A_1, A_2, \dots, A_n$ , и на рынке имеется  $q$  участников - инвесторов  $B^1, B^2, \dots, B^q$ , покупающих и продающих эти активы. Пусть каждый участник-инвестор  $B^k$  обладает начальным капиталом  $V^k$ , где  $k=1, 2, \dots, q$ . Каждый из этих участников сформирует оптимальный, с учетом своего отношения к риску, портфель

$$\underline{x}^k = (x_0^k, x^k) \quad k=1, \dots, q,$$

где  $x_0^k$  – доля начального капитала, инвестируемая в безрисковый актив;

$x^k$  – рисковая часть портфеля инвестора  $B^k$ .

Начальный инвестируемый капитал  $V^k$  инвестора  $B^k$  разобьется на две части:

безрисковую  $V_0^k = x_0^k \cdot V^k$

и рисковую  $W^k = (1 - x_0^k) \cdot V^k$ ;  $V^k = V_0^k + W^k$ .

Тогда общий рисковый капитал, вложенный во все рисковые активы, равен:  $W = W^1 + W^2 + \dots + W^q$ .

Так как все инвесторы вкладывают рисковую часть в один и тот же касательный портфель  $\tau = (\tau_1, \tau_2, \dots, \tau_n)$ , то портфель, соответствующий рисковей части  $x^k$

$$u^k = \frac{x^k}{1 - x_0^k}$$

будет совпадать с касательным портфелем  $u^k = \tau$ ,  $k=1, 2, \dots, q$ .

Отсюда следует, что  $k$ -й инвестор вложит в  $j$ -й рисковый актив  $\tau_j$ -ю долю своего рискового капитала  $W^k$ , т.е.

$$W_j^k = \tau_j \cdot W^k.$$

Общая сумма средств, инвестированная в  $j$ -й рисковый актив, равна

$$W_j = \tau_j \cdot W^1 + \tau_j \cdot W^2 + \dots + \tau_j \cdot W^q = \tau_j W$$

$$\tau_j = \frac{W_j}{W}.$$

Таким образом, вес  $j$ -го актива данного портфеля равен рыночной доли этого актива, т.е. отношению суммы средств, вложенных в этот актив, к общей сумме средств, вложенных во все рискованные активы. Это приводит к важнейшему выводу: в условиях рыночного равновесия цены на рискованные активы установятся таким образом, что будет выполняться соотношение  $\tau_j = \frac{w_j}{W}$ . Поэтому касательный портфель называют также *рыночным портфелем*  $\pi_M$  с вектором весов  $x_M$ , доходностью  $r_M$ , вариацией  $V_M$ , стандартным отклонением  $\sigma_M$ . Уравнение эффективной линии рынка можно переписать в виде:

$$r = r_0 + \frac{r_M - r_0}{\sigma_M} \cdot \sigma.$$

Угловым коэффициентом этой прямой  $\frac{r_M - r_0}{\sigma_M}$  называется *ценой за риск* и показывает, насколько возрастает ожидаемая доходность эффективного портфеля при увеличении среднеквадратического отклонения на одну единицу.

Величина  $\frac{r_M - r_0}{\sigma_M} \cdot \sigma$  называется *премией за риск*, которую требует инвестор при вложении капитала в рискованные активы. Превышение рыночной доходности над безрисковой ( $r_M - r_0$ ) называется *рыночной премией*.

Согласно терминам премий уравнение эффективной линии рынка позволяет утверждать, что премия эффективного портфеля всегда пропорциональна рыночной премии.

#### **45. Мера риска актива в модели ценообразования финансовых активов (CAPM).**

Уравнение эффективной линии рынка связывает линейным образом доходность и риск эффективных портфелей. Связь между доходностью и риском любых портфелей линейна, а мерой риска в этой связи является не традиционно понимаемая характеристика – стандартное отклонение, а величина, называемая бетой портфеля (актива).

Бета-коэффициент, как мера оценки систематического риска актива, позволяет в модели САРМ увязать требуемую доходность с остающимся после диверсификации риском.

Для измерения рыночного риска актива применяют величину «бета». Бета иллюстрирует зависимость между доходностью актива и доходностью рынка. Доходность рынка – это доходность рыночного портфеля. На практике невозможно сформировать собственно рыночный портфель, включающий все активы, поэтому используют фондовый индекс.

Бета доходности  $R_\pi$  портфеля  $\pi$  относительно доходности  $R_M$  рыночного портфеля  $\pi_M$  называется бетой портфеля:

$$\beta_\pi = \frac{\text{cov}(R_\pi, R_M)}{\sigma_M^2},$$

где  $\text{cov}(R_\pi, R_M)$  – ковариация доходности портфеля  $\pi$  с доходностью рыночного портфеля;

$\sigma_M^2$  – дисперсия доходности рыночного портфеля.

Бета доходности актива  $A$  относительно доходности  $R_M$  рыночного портфеля  $\pi_M$  называется бетой портфеля (актива):

$$\beta_A = \frac{\text{cov}(R_A, R_M)}{\sigma_M^2},$$

где  $\text{cov}(R_A, R_M)$  – ковариация доходности актива  $A$  с доходностью рыночного портфеля.

С учетом того, что  $\text{cov}(R_\pi, R_M) = \text{cor}(R_\pi, R_M) \cdot \sigma_\pi \cdot \sigma_M$ , получаем:

$$\beta_\pi = \frac{\text{cor}(R_\pi, R_M) \cdot \sigma_\pi \cdot \sigma_M}{\sigma_M^2} = \rho_{\pi M} \frac{\sigma_\pi}{\sigma_M},$$

где  $\rho_{\pi M}$  – коэффициент корреляции портфеля с рынком.

Ковариация является симметрической билинейной функцией своих аргументов, и бета портфеля  $\pi$  с вектором весов  $x = (x_1, x_2, \dots, x_n)$  является линейной комбинацией «бет», составляющих портфель активов:

$$\beta_\pi = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n.$$

Бета рыночного портфеля равна единице, так как ковариация доходности рыночного портфеля с самим собой есть его дисперсия:

$$\beta_M = \frac{\sigma_M^2}{\sigma_M^2} = 1.$$

Бета актива без риска равна нулю.

Величина  $\beta$  актива показывает, насколько риск актива отличается от риска рыночного портфеля. Активы с бетой больше единицы обладают большим риском, чем рыночный портфель, а активы с бетой меньше единицы, соответственно, менее рискованны.

Активы делят на *агрессивные* и *защитные*. Бета агрессивных активов больше единицы ( $\beta > 1$ ), а защитных – меньше единицы ( $\beta < 1$ ). Бета может быть как положительной, так и отрицательной величиной. Положительная бета свидетельствует о том, что доходности актива и рынка при изменении конъюнктуры изменяются в одном направлении, а отрицательная бета, что доходности актива и рынка меняются в противоположных направлениях. В реальной жизни большинство активов характеризуются положительной бетой.

Значение $\beta$	Интерпретация $\beta$
$\beta < 0$	доходность акции демонстрирует разнонаправленное движение с доходностью рыночного портфеля
$\beta = 0$	отсутствие связи между доходностью акции и доходностью рыночного портфеля
$0 < \beta < 1$	доходности акции и рыночного портфеля демонстрируют однонаправленное движение, доходность акции менее чувствительна к систематическому риску
$\beta = 1$	риски акции и рыночного портфеля равны, их доходности демонстрируют однонаправленное движение
$\beta > 1$	доходности акции и рыночного портфеля демонстрируют однонаправленное движение, доходность акции более чувствительна к систематическому риску



Зная бету актива, можно оценить, насколько должна измениться его ожидаемая доходность при изменяемой доходности рынка.

Например, пусть бета актива равна +2,5. Тогда при увеличении ожидаемой доходности рыночного портфеля на 1% произойдёт рост доходности актива на 2,5%.

Если бета актива равна 0,8, то при увеличении ожидаемой доходности рынка на 1% ожидаемая доходность актива должна возрасти на 0,8%.

Если бета равна -2,5, то при повышении доходности рыночного портфеля на 1% доходность актива снизится на 2,5% и наоборот.

Активы с отрицательной бетой являются особо ценными инструментами для диверсификации портфеля поскольку в этом случае можно построить безрисковый портфель с «нулевой бетой».

Бета актива рассчитывается на основе доходности актива и рынка за предыдущие периоды времени. Информацией о значениях беты располагают аналитические компании.

Основное уравнение CAPM записывается в виде:

$$r_A = r_f + \beta_A \cdot (r_M - r_f),$$

где  $r_A$ - требуемая (ожидаемая) доходность актива  $A$ ;

$r_M$ - рыночная доходность (доходность фондового индекса);

$r_f$ - безрисковая ставка доходности.

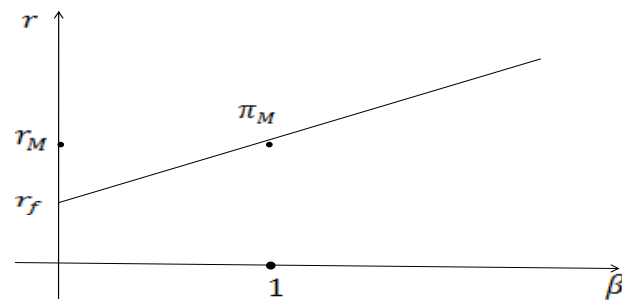
и

$$r_\pi = r_f + \beta_\pi \cdot (r_M - r_f),$$

где  $r_\pi$ - требуемая (ожидаемая) рыночная доходность.

Премия за риск актива  $A$  равна:  $(r_A - r_f)$ ; премия за риск рыночного портфеля (рыночная премия за риск) равна:  $(r_M - r_f)$ .

Основное уравнение CAPM описывает линейную связь между ожидаемой доходностью портфеля или актива и его бетой, которая представляет характеристику риска. Прямая линия, соответствующая этому уравнению, называется *характеристической линией рынка*.



Коэффициент бета портфеля или актива играет роль «коэффициента усиления», преобразующего рыночную премию за риск актива  $A$  в премию по портфелю  $\pi$ . Бета является мерой чувствительности доходности портфеля или актива к изменениям рыночной доходности.

Таким образом,  $\Delta r_{\pi} = \beta_{\pi} \cdot \Delta r_M$ ,

где  $\Delta r_{\pi}, \Delta r_M$  - изменения рыночной и портфельной доходностей.

Прирост ожидаемой доходности (или премии) активов, у которых бета больше единицы, будут больше, чем прирост рыночной премии, а тех, у которых бета меньше единицы, будут меньше, чем прирост рыночной премии.

**46. Градиентные методы поиска локального экстремума функции многих переменных: метод наискорейшего подъема (или спуска), метод проекции градиента, метод условного градиента.**

Задача поиска безусловного экстремума (без ограничений) имеет вид

$$f(x) \rightarrow \text{extr}, x \in X = R^n. \quad (46.1)$$

Классические необходимые условия экстремума (локального или глобального, максимума или минимума) для этой задачи дает теорема Ферма.

*Теорема.* (теорема Ферма) Пусть функция  $f(x)$  дифференцируема в точке решения задачи (1)  $x^*$ , тогда

$$\frac{\partial f(x^*)}{\partial x_j} = 0, j = 1, \dots, n. \quad (46.2)$$

Непосредственное решение системы уравнений (46.2) может оказаться достаточно сложным, поэтому на практике могут использоваться, например, итеративные градиентные методы. Градиентные методы – это методы первого порядка, т.е. методы поиска экстремума, использующие первые производные. Фактически в этих методах производится линеаризация, связанная с заменой максимизируемой функции  $f(x)$  линейным членом разложения ее в ряд Тейлора. Для задачи на максимум, выбирая произвольную начальную точку  $x^{(0)}$ , строят итеративный процесс

$$x^{(k+1)} = x^{(k)} + \alpha_k f'(x^{(k)}), k=0, 1, 2, \dots \quad (46.3)$$

Число  $\alpha_k$  называют длиной шага или просто шагом. Если все  $\alpha_k$  равны между собой, то имеем процесс с постоянным шагом. Процесс (46.3), лежащий в основе градиентных методов, представляет собой движение в сторону возрастания функции  $f(x)$ , так как если градиент в текущей точке  $f'(x^{(k)}) \neq 0$ , то всегда можно выбрать  $\alpha_k$ , так, что  $f(x^{(k+1)}) > f(x^{(k)})$ . Для задачи на минимум, естественно, движение осуществляется в направлении антиградиента.

Существуют разные способы выбора  $\alpha_k$ . Вообще говоря, наилучшим является выбор такого  $\alpha_k$ , при котором обеспечивается максимальный рост функции  $f(x)$ . Такое  $\alpha_k$  находится из условия

$$f\left(x^{(k)} + \alpha_k f'(x^{(k)})\right) = \max_{\alpha \geq 0} f\left(x^{(k)} + \alpha f'(x^{(k)})\right). \quad (46.4)$$

Градиентный метод поиска экстремума (3) с выбором шага по способу (46.4) называется **методом скорейшего подъема** (или спуска для задачи на минимум). Такой метод требует наименьшего числа итераций, но зато на каждом шаге приходится решать дополнительную задачу поиска экстремума (правда, в одномерном случае). На практике часто довольствуются нахождением любого  $\alpha_k$ , обеспечивающего рост функции. Для этого берут произвольное  $\alpha_k$  и проверяют условие роста, если оно не

выполняется, то дробят  $\alpha_k$  до тех пор, пока это условие не будет выполнено (такое достаточно малое  $\alpha_k$  при  $f'(x^{(k)}) \neq 0$  существует всегда).

Процесс (3), очевидно, останавливается, когда выполнено условие (46.2). Так как это условие не обеспечивает наличия экстремума, необходимо провести дополнительное исследование функции  $f(x)$  в окрестности найденной точки (например, с помощью достаточных условий). Однако даже если найдена точка максимума, определить локальный это максимум или глобальный не всегда возможно. Поэтому градиентные методы в общем случае дают лишь локальные экстремумы (при этом можно попытаться найти глобальный экстремум, применяя итеративный процесс многократно с разными начальными точками). Градиентные методы обеспечивают нахождение глобального экстремума только для вогнутых (выпуклых) функций; если функция  $f(x)$  вогнута, то найденная стационарная точка будет глобальным максимумом, а если выпукла – глобальным минимумом.

Как правило, итеративный процесс (46.3) бесконечный, поэтому необходимо ввести правило остановки. Для этого задается требуемая точность решения  $\varepsilon$ , процесс останавливают, когда, например, либо расстояние между двумя последовательными точками становится меньше  $\varepsilon$ , либо градиент по модулю меньше  $\varepsilon$ .

Если рассматривается задача максимизации  $f(x)$  при ограничениях, т.е. когда в (46.1)  $X$  не совпадает с  $R^n$ , то непосредственное применение градиентного процесса может привести к нарушению ограничений, даже если начальная точка  $x^{(0)} \in X$ . Однако эту трудность можно преодолеть, например, если получаемую очередную точку проектировать на множество  $X$ . Если обозначить операцию проектирования точки  $x$  на множество  $X$  через  $\text{Pr}_X(x)$ , то соответствующий итеративный процесс имеет вид

$$x^{(k+1)} = \text{Pr}_X \left( x^{(k)} + \alpha_k f'(x^{(k)}) \right). \quad (46.5)$$

Полученный метод носит название *метода проекции градиента*. Шаг  $\alpha_k$  в методе (46.5) может выбираться различными способами (например, как в методе скорейшего подъема). Стационарная точка этого процесса является решением задачи  $\max_{x \in X} f(x)$  в случае вогнутой функции  $f(x)$ , а в общем случае требуется дополнительное исследование.

Недостатком метода проекции градиента является необходимость проведения операции проектирования, которая в общем случае эквивалентна некоторой задаче поиска экстремума. Однако, когда  $X$  является шаром, параллелепипедом, гиперплоскостью, полупространством или ортантом, задача проектирования решается просто и в явном виде.

Еще одной разновидностью градиентных методов является *метод условного градиента*, который также предназначен для решения экстремальных задач с ограничениями и состоит в решении вспомогательной задачи максимизации на множестве  $X$  линейной функции  $\langle f'(x^{(k)}), x - x^{(k)} \rangle$  (здесь  $\langle \cdot \rangle$  означает скалярное произведение), представляющей собой главную часть приращения функции  $f(x)$  в точке  $x^{(k)}$ . Эта вспомогательная задача может быть непростой, но если  $X$  задается линейными ограничениями, то она представляет собой задачу линейного программирования (ЛП), которая решается за конечное число шагов стандартными методами (например, симплекс-методом). Если решение вспомогательной задачи  $\tilde{x}^{(k)}$  найдено, то следующее приближение для исходной задачи строится по формуле

$$x^{(k+1)} = x^{(k)} + \alpha_k (\tilde{x}^{(k)} - x^{(k)}) \quad (46.6)$$

Если множество  $X$  выпуклое, то  $x^{(k+1)} \in X$ . Шаг  $\alpha_k$  выбирается из условия максимального роста функции  $f(x)$  или любым другим способом, обеспечивающим рост  $f(x)$ . На практике обычно решают вспомогательную задачу не точно, а приближенно. В процессе (46.6) направление движения не совпадает с градиентом функции  $f(x)$  в точке  $x^{(k)}$ , но определяется им, так

как его компоненты берутся в качестве коэффициентов линейной целевой функции вспомогательной задачи.

#### **47. Постановка задачи целочисленного программирования. Метод Гомори. Метод ветвей и границ.**

Задачу оптимизации, в которой требуется, чтобы все или некоторые переменные принимали только целочисленные (или дискретные) значения, называют задачей целочисленного (или дискретного) программирования. Иначе говоря, задачей дискретного или *целочисленного программирования* называется задача  $f(x) \rightarrow \max, x \in X, X$  – дискретное множество. Как правило,  $X \subset \mathbb{Z}^n$  – множество точек с целочисленными координатами или  $X \subset B^n, B = \{0, 1\}$  – множество точек, координаты которых принимают значение 0 или 1 (булевы переменные). Среди методов решения целочисленных задач можно выделить два направления: методы отсечения и комбинаторные методы.

В **методах отсечения** для решения целочисленных задач используются методы решения непрерывных задач. Рассмотрим это на примере задачи ЛП:  $cx \rightarrow \max, Ax \leq b, x \geq 0$ , где  $c \in R^n, b \in R^m, A$  – матрица размера  $m \times n, x \in \mathbb{Z}^n$ . Сначала решается задача ЛП без условия целочисленности, которую назовем задачей 1'. Если найденное решение  $x^1$  целочисленно, то оно является решением исходной задачи. Если нет, то к ограничениям задачи 1' добавляется дополнительное ограничение, которое отсекает  $x^1$  и сохраняет в допустимом множестве исходной задачи все его целочисленные точки. Такое ограничение называется правильным отсечением. На следующем этапе находится решение задачи с дополнительно введенным ограничением, которую назовем задачей 2'. Если решение  $x^2$  не является целочисленным, то для него вводится правильное отсечение. И так далее до тех пор, пока решение очередной задачи ЛП не окажется целочисленным. Конкретные способы построения правильных

отсечений приводят к конкретным вычислительным алгоритмам. Наиболее известны алгоритмы Гомори.

Из симплексной таблицы, содержащей оптимальный нецелочисленный план, возьмем уравнение

$$\sum_{j=1}^n a_{ij}x_j = b_i, \quad (47.1)$$

причем  $b_i$  и некоторые коэффициенты  $a_{ij}$ , являются дробными. Поскольку  $x_j$  – неотрицательные целые числа, то к ограничениям исходной задачи можно добавить новое ограничение  $\sum_{j=1}^n [a_{ij}]x_j \leq [b_i]$ . Знак “ $\leq$ ”, так как  $\sum_{j=1}^n [a_{ij}]x_j$  – целое и его всегда можно ограничить целой частью  $b_i$ . Откуда следует, что неравенство сохраняет все целые точки многогранного множества  $X$ . Можно преобразовать последнее неравенство к виду

$$\sum_{j=1}^n [a_{ij}]x_j + x_{n+1} = [b_i], \quad (47.2)$$

добавив свободную переменную  $x_{n+1}$  с условием ее неотрицательности и целочисленности. Условие (47.2) будет выполняться при целых  $x_j$ . Вычитая из (47.2) уравнение (47.1), получаем ограничение, содержащее дробные части:

$$\sum_{j=1}^n (-\{a_{ij}\})x_j + x_{n+1} = -\{b_i\}. \quad (47.3)$$

Уравнение (47.3) называется **отсечением Гомори**. Из (47.3) следует неравенство

$$\sum_{j=1}^n \{a_{ij}\}x_j \geq \{b_i\}. \quad (47.4)$$

Нецелочисленный план не удовлетворяет (47.4), т.к.  $\forall i \{b_i\} > 0$ , а левая часть (47.4) равна нулю ( $\sum_{j=1}^n \{a_{ij}\}x_j = 0$ ), т.к. свободные переменные равны нулю. Любой целочисленный план удовлетворяет (47.4) как строгому равенству, т.к. в этом случае  $\forall i \{b_i\} = 0$ .

Заметим, что (9) уже не содержит базисную переменную из (47.1), а новая базисная переменная  $x_{n+1}$  будет выражена через прежние свободные.

Представление о комбинаторных методах дает наиболее известный и широко используемый на практике **метод ветвей и границ**. Основная идея метода состоит в разбиении допустимого множества на подмножества

(ветвление) и вычислении оценок (границ) значений целевой функции на этих подмножествах, позволяющие отбрасывать подмножества, заведомо не содержащие решений. Графически указанный процесс можно представить структурной схемой в виде дерева, у которого каждая вершина соответствует некоторому подмножеству планов (и, обычно, связанной с ним вспомогательной задаче ЛП), а выходящие из нее ветви отвечают разбиению этого подмножества на новые подмножества планов.

Для решения задачи целочисленного программирования методом ветвей и границ можно использовать алгоритм Лэнда-Дойга. Прежде всего, решается исходная задача ЛП без условия целочисленности. Если в полученном решении все переменные целочисленные, то задача решена. Если некоторая компонента  $x_k$  нецелочисленная, т.е.  $x_k = [x_k] + \{x_k\}$ , где  $[x_k]$  – целая часть  $x_k$ ,  $\{x_k\}$  – дробная часть  $x_k$ , то исходная задача разбивается на две подзадачи ЛП. Одна подзадача решается дополнительным ограничением  $x_k = [x_k]$ , а другая – с дополнительным ограничением  $x_k = [x_k] + 1$ . Вычисляется оценка (значение целевой функции). Если одна из этих двух подзадач, например, с ограничением  $x_k = [x_k]$ , дает нецелочисленное решение с компонентой  $x_r = [x_r] + \{x_r\}$ , то переходим к следующему шагу, на котором выбранная подзадача разветвляется на две новые с дополнительными ограничениями  $x_k = [x_k]$ ,  $x_r = [x_r]$  и  $x_k = [x_k]$ ,  $x_r = [x_r] + 1$  соответственно. Вычисления заканчиваются, когда кандидатов на ветвление больше нет.

Описанный алгоритм применим и для задач, в которых вместо требования целочисленности содержится требование, чтобы переменная принимала только значения 0, 1.

#### **48. Постановка задачи многокритериальной оптимизации. Метод целевого программирования. Метод идеальной точки.**

*Многокритериальная оптимизация* – это раздел математического программирования, посвященный проблемам выбора принципов



оптимальности и методов нахождения их реализаций в экстремальных задачах с несколькими критериями. Пусть на допустимом множестве  $X$  задано несколько частных критериев  $f_1(x), f_2(x), \dots, f_m(x)$ . Цель задачи оптимизации состоит в получении результата, имеющего как можно более высокие оценки по каждому критерию. Такая задача называется *задачей многокритериальной (или векторной) оптимизации*.

Рассмотрим задачу с двумя критериями, считая их равноправными (постановку задачи можно обобщить на любое количество критериев). На плоскости  $x_1Ox_2$  задано допустимое множество  $X$  и в каждой точке этого множества определены две непрерывные функции  $f_1(x)=\Phi(x_1, x_2)$  и  $f_2(x)=\Psi(x_1, x_2)$ . Требуется найти точку  $(x_1^*, x_2^*) \in X$ , в которой  $\Phi(x_1^*, x_2^*)$  и  $\Psi(x_1^*, x_2^*)$  принимают максимальные значения, т.е. решить совместно две экстремальные задачи  $\Phi(x_1, x_2) \rightarrow \max$  и  $\Psi(x_1, x_2) \rightarrow \max, (x_1, x_2) \in X$ . Наибольшее значение  $f_1$  ( $f_{1\max}$ ) и наибольшее значение  $f_2$  ( $f_{2\max}$ ) достигаются в разных точках, а точка с координатами  $(f_{1\max}, f_{2\max})$  лежит вне множества  $Y$  значений этих критериев. Тем самым в исходной постановке задача, вообще говоря, неразрешима – удовлетворить обоим требованиям одновременно невозможно. Следовательно, нужно искать какое-то компромиссное решение. Процесс решения многокритериальных задач связан с экспертными оценками как самих критериев, так и взаимоотношений между ними.

Среди известных методов решения задач многокритериальной оптимизации можно отметить метод идеальной точки. Метод идеальной точки состоит в отыскании на множестве Парето точки, ближайшей к *точке утопии*, задаваемой ЛПР. Обычно ЛПР формулирует цель в виде желаемых значений показателей, и часто в качестве координат целевой точки выбирается сочетание наилучших значений всех критериев (обычно эта точка не реализуется при заданных ограничениях, поэтому ее и называют точкой утопии).

**Множество Парето-оптимальных решений** определяется равенством  $P_f(X) = \{x^* \in X \mid \nexists x \in X, \text{ что } f(x) \geq f(x^*)\}$ ,  $f(x) = (f_1(x), f_2(x), \dots, f_m(x))$ . Вектор  $f(x^*)$  при Парето-оптимальном решении  $x^*$  называют **Парето-оптимальным вектором решения**  $x^*$  или просто Парето-оптимальным вектором, а множество всех таких векторов – **множеством Парето-оптимальных векторов**.

Мы рассмотрели задачу, в которой оба критерия максимизируются. На практике часто встречаются случаи, когда требования выглядят по-иному, например,  $\Phi(x_1, x_2) \rightarrow \min$ ,  $\Psi(x_1, x_2) \rightarrow \max$ . Можно решать такие задачи непосредственно (для этой задачи множество Парето представляет собой «северо-западную» границу), но, поменяв в случае необходимости знак у критерия на противоположный, можно свести любую двухкритериальную задачу к рассмотренной задаче на максимум частных критериев, для которой множество Парето представляет собой «северо-восточную» границу.

Точка  $x^*$  называется **идеальной точкой**, если она является оптимальной всем критериям, т.е.  $\forall i = 1, \dots, m \quad \forall x \in X \quad f_i(x^*) \geq f_i(x)$ . Если идеальная точка на множестве  $X$  существует, то она является решением задачи многокритериальной оптимизации. Но такая ситуация является нетипичной. В общем случае можно найти идеальную точку в пространстве критериев  $y^* = (y_1^*, \dots, y_m^*)$ , где  $y_i^* = \max_{x \in X} f_i(x)$  и ближайшую к ней точку множества  $Y = \{y \in R^m \mid y_i = f_i(x), x \in X\}$ . Такой подход к решению задачи многокритериальной оптимизации называется *методом идеальной точки*.

Обобщением метода идеальной точки является *целевое программирование*. Этот термин был предложен А. Чарнсом и В. Купером для определенного типа задач математического программирования, состоящих в нахождении решений, обеспечивающих «как можно более близкое» приближение к множеству одновременно недостижимых целей.

Обозначим  $Y^*$  целевое множество в пространстве критериев. Данное множество может быть задано, например,  $f_i(x) = y_i^*$  или  $f_i(x) \geq y_i^*$  ( $f_i(x) \leq y_i^*$ ), или  $f_i(x) \in [y_i^1, y_i^2]$ ,  $i=1, \dots, m$ .

Кандидатами в решение являются точки из  $Y$ , ближайшие к  $Y^*$  в смысле какой-то метрики  $\rho$  в пространстве критериев. Тогда в общем виде задача формулируется следующим образом:  $\min_{x \in X} \rho(f(x), Y^*)$ , где  $\rho(f(x), Y^*) = \min_{y^* \in Y^*} \rho(f(x), y^*)$ .

Метрика может быть выбрана, например, полиэдральная (обобщенная линейная):  $\rho_0(f(x), y^*) = \max_i \alpha_i |f_i(x) - y_i^*|$ ,  $\rho_1(f(x), y^*) = \sum_i \alpha_i |f_i(x) - y_i^*|$ , или не полиэдральная:  $\rho_2^2(f(x), y^*) = \sum_i \alpha_i (f_i(x) - y_i^*)^2$ . Если метрика полиэдральная, задача целевого программирования может быть сведена к задаче ЛП.

#### 49. Постановка взаимно двойственных задач линейного программирования. Экономический смысл двойственности. Основная теорема двойственности и ее следствия.

Рассмотрим пару двойственных задач линейного программирования

$$\left\{ \begin{array}{l} AX \leq B, \\ \tilde{X} \geq 0, \\ z = C^t X + c_0 \rightarrow \max \end{array} \right. \quad \text{и} \quad \left\{ \begin{array}{l} A^t Y \geq C, \\ \tilde{Y} \geq 0, \\ T = B^t Y + c_0 \rightarrow \min \end{array} \right.$$

где  $A = \|a_{ij}\|$  —  $m \times n$  матрица,  $X = (x_1, x_2, \dots, x_n)^t$ ,  $Y = (y_1, y_2, \dots, y_m)^t$ ,  $B = (b_1, b_2, \dots, b_m)^t$ ,  $C = (c_1, c_2, \dots, c_n)^t$  — векторы-столбцы соответствующей размерности, а вектора  $\tilde{X} = (x_{i_1}, x_{i_2}, \dots, x_{i_k})^t$ ,  $\tilde{Y} = (y_{j_1}, y_{j_2}, \dots, y_{j_l})^t$  с индексами  $1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$ ,  $1 \leq j_1 \leq j_2 \leq \dots \leq j_k \leq m$  — части векторов  $X, Y$  (то есть тривиальные ограничения налагаются лишь на часть координат векторов  $X, Y$ ). Нетривиальные ограничения в обеих задачах могут быть как типа неравенств (со знаком

« $\leq$ » или « $\geq$ »), так и типа уравнений. Чтобы понять связь между этими задачами, построим расширенные матрицы

$$\tilde{A} = \left[ \begin{array}{cccccc} a_{11} & a_{12} & \dots & a_{1n} & \leq & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & = & b_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & \geq & b_m \\ \geq & \sim & \dots & \geq & & \\ c_1 & c_2 & \dots & c_n & & c_0 \end{array} \right]_{\rightarrow max}$$

и

$$\tilde{A}' = \left[ \begin{array}{cccccc} a_{11} & a_{21} & \dots & a_{m1} & \geq & c_1 \\ a_{12} & a_{22} & \dots & a_{m2} & = & c_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} & \geq & c_n \\ \geq & \sim & \dots & \geq & & \\ b_1 & b_2 & \dots & b_m & & c_0 \end{array} \right]_{\rightarrow min}$$

обеих задач. Таким образом,

Матрица нетривиальных ограничений двойственной задачи получается из соответствующей матрицы исходной задачи транспонированием.

Правые части нетривиальных ограничений двойственной задачи являются коэффициентами целевой функции исходной задачи, и, наоборот, коэффициенты целевой функции двойственной задачи совпадают с правыми частями ограничений исходной задачи.

Если исходная задача является задачей на максимум (на минимум), то двойственная задача будет задачей на минимум (на максимум), и строка тривиальных ограничений переходит в столбец нетривиальных ограничений без изменения знака неравенства с « $\leq$ » на « $\geq$ » и наоборот (соответственно, с изменением знака). Если на переменную в исходной задаче тривиальное ограничение отсутствует, то соответствующее ограничение в двойственной задаче будет типа уравнения; иными словами, « $\sim$ » переходит в « $=$ ». Столбец нетривиальных ограничений переходит в строку тривиальных ограничений с изменением знака неравенства с « $\leq$ » на « $\geq$ » и наоборот (соответственно, без изменения знака), а « $=$ » переходит в « $\sim$ ».

Основная связь между двойственными задачами изложена в следующих утверждениях.

*Теорема (основное неравенство для двойственных задач).* Для всех допустимых решений  $X, Y$  пары двойственных задач имеет место неравенство  $z(X) \leq T(Y)$ .

*Теорема (первая теорема двойственности-основная теорема двойственности).* Если исходная задача имеет оптимальное решение, то и двойственная ей имеет оптимальное решение. При этом оптимальные значения обеих целевых функций равны, то есть  $z_{\max} = T_{\min}$

Чтобы решить задачу линейного программирования, иногда проще решить двойственную задачу, а затем найти решение исходной задачи.

Сформулируем теперь *теорему равновесия* (вторую теорему двойственности), которая позволяет не только установить связь между оптимальными значениями целевых функций, но и между точками, в которых эти значения достигаются.

*Теорема (вторая теорема двойственности - теорема равновесия).* Оптимальные решения  $X^* = (x_1^*, x_2^*, \dots, x_n^*)$  и  $Y^* = (y_1^*, y_2^*, \dots, y_m^*)$  пары двойственных задач связаны между собой равенствами

$$\begin{cases} \left( \sum_{i=1}^m a_{ik} y_i^* - c_k \right) \cdot x_k^* = 0, k = 1, \dots, n, \\ \left( \sum_{k=1}^n a_{ik} x_k^* - b_i \right) \cdot y_i^* = 0, i = 1, \dots, m. \end{cases}$$

Понятие двойственности имеет естественную экономическую интерпретацию. Рассмотрим задачу об использовании ресурсов:

$$\begin{aligned} f = C^T X &\rightarrow \max; \\ \begin{cases} AX \leq B \\ X \geq 0 \end{cases} \end{aligned},$$

где  $A = (a_{ij})$  - матрица норм расхода ресурсов  $R_1, R_2, \dots, R_m$  для производства продукции вида  $P_1, P_2, \dots, P_n$ ,  $C = (c_1, c_2, \dots, c_n)^T$  - вектор прибыли от реализации единицы соответствующей продукции,  $B = (b_1, b_2, \dots, b_m)^T$  - запасы ресурсов,  $X = (x_1, x_2, \dots, x_n)^T$  - соответствующий производственный план.

Альтернативой производства продукции является продажа имеющихся ресурсов по ценам  $p = (p_1, p_2, \dots, p_m)^T$ . Ясно, что выручка от продажи ресурсов должна быть не меньше прибыли, полученной при продаже продукции, при чем покупатель заинтересован заплатить за ресурсы наименьшую стоимость. Данные требования равносильны выполнению условий:

$$g = B^T p \rightarrow \min;$$

$$\begin{cases} A^T p \geq C \\ p \geq 0 \end{cases}$$

Мы получили задачу, двойственную к задаче об использовании ресурсов.

## **50. Общая постановка многокритериальной оптимизации. Парето-эффективная граница. Методы решения многокритериальной оптимизации.**

Задача вида

$$\begin{cases} f_i(x) \rightarrow \max(\min) \\ x \in D \end{cases},$$

где  $i > 1$ ,  $D \in R^n$  – некоторое допустимое множество, а  $f_i(x)$  – гладкие функции на  $D$ , называется *задачей многокритериальной оптимизации*.

В экономических задачах область допустимых решений  $D$  обычно задается системой уравнений и неравенств, к которой могут быть добавлены

некоторые дополнительные ограничения, например, целочисленность переменных.

*Определение.* Пусть  $X$  и  $Y$  – два допустимых решения. Говорят, что  $X$  доминирует  $Y$ , если для всех  $i = 1, 2, \dots, n$  выполняется неравенство  $f_i(X) \geq f_i(Y)$  и найдется такое  $k$ , что  $f_k(X) > f_k(Y)$ .

*Определение.* Решение  $Z$ , называется недоминируемым (эффективным), если нет решения  $X$ , которое бы доминировало  $Z$ .

*Определение.* Множество эффективных (недоминируемых) решений называется *множеством Парето*. Геометрическое изображение множества Парето называется *Парето-эффективной границей* (Парето-оптимальной границей).

В задаче многокритериальной оптимизации наилучшее решение следует искать в множестве Парето.

*Алгоритм построения Парето-эффективной границы:*

1. Строим допустимое множество  $D$ , заданное системой ограничений, как пересечение полуплоскостей, соответствующих каждому неравенству, входящему в эту систему.

2. Для каждой функции  $f_i = c_{i1}x_1 + c_{i2}x_2 + c_{io}$  строим линию уровня как прямую, перпендикулярную соответствующему вектору нормали  $\bar{n}_i = (c_{i1}, c_{i2})$ . Каждая из этих линий уровня разбивает плоскость  $XOY$  на две полуплоскости. Пусть  $\Pi_i$  – полуплоскости, содержащие вектор градиента целевой функции  $f_i$ , а их пересечение  $\Pi = \bigcap_{i=1}^n \Pi_i$ ;

3. Перемещая данную область  $\Pi$  по границе допустимого множества  $D$ , находим те точки границы, которые являются единственными точками пересечения областей  $\Pi$  и  $D$ . Данные точки являются оптимальными по Парето, а множество всех таких точек – Парето-эффективной границей.

*Пример.* Найти Парето-эффективную границу задачи

$$\begin{cases} f_1 = 4x_1 + x_2 \rightarrow \max \\ f_2 = x_1 + 2x_2 \rightarrow \max \\ x_1 + x_2 \leq 7 \\ x_1 \leq 5 \\ x_2 \leq 4 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

*Решение.* Область допустимых решений D задается системой неравенств

$$\begin{cases} x_1 + x_2 \leq 7 \\ x_1 \leq 5 \\ x_2 \leq 4 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

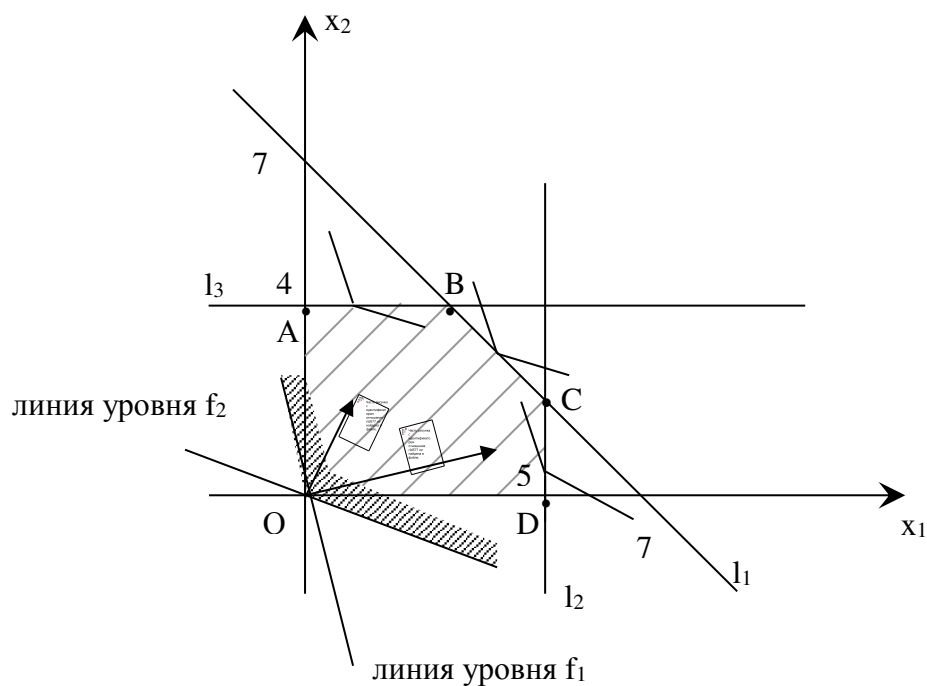
Построим данную область.

$$l_1 : x_1 + x_2 = 7$$

$$l_2 : x_1 = 5$$

$$l_3 : x_2 = 4$$

В качестве допустимого множества получаем область OABCD с угловыми точками O (0;0), A(0;4), B(3;4), C(5;2), D (5;0).





Для функций  $f_1$  и  $f_2$  построим линии уровня ( $f_1=const$ ,  $f_2=const$ ) как прямые, перпендикулярные соответствующим векторам нормали  $\vec{n}_1=(4;1)$  и  $\vec{n}_2=(1;2)$ . Каждая из этих линий уровня разбивает плоскость  $XOY$  на две полуплоскости. Рассмотрим те из них  $\Pi_i$ , которые содержат соответствующий вектор нормали (вектор градиента целевой функции). Пусть  $\Pi=\Pi_1\cap\Pi_2$ . Перемещая область  $\Pi$  по границе множества  $D$  легко определить, что Парето-эффективной границей будет отрезок  $[BC]$ , т.е. множество точек  $(1-t)(3;4)+t(5;2)=(3+2t; 4-2t)$ ,  $t \in [0; 1]$ .

*Ответ.*  $[BC]$  – Парето-эффективная граница, т.е. множество точек  $(3+2t; 4-2t)$ ,  $t \in [0; 1]$

*Замечание.* Задача минимизации одной или нескольких целевых функций в задаче многокритериальной оптимизации легко сводится к задаче максимизации умножением на  $(-1)$ .

### *Методы решения многокритериальной задачи оптимизации*

*Метод обобщенного критерия.* Метод перехода от нескольких критериев  $f_1, f_2, \dots, f_m$  к одному, задаваемому новой функцией  $z = \sum_{j=1}^m (\alpha_j \cdot f_j)$  называется сверткой или методом обобщенного критерия. Числа  $\alpha_j$  называются весовыми коэффициентами. Чем больше  $\alpha_j$ , тем больший «вклад» вносит  $j$ -й критерий в обобщенный критерий  $z$ . Иногда требуют, чтобы  $\sum_{j=1}^m \alpha_j = 1$ . Выбор весовых коэффициентов  $\alpha_j$  имеет субъективный характер.

*Метод приоритетов.* Метод приоритетов решения многокритериальных задач применяется в том случае, когда критерии  $f_i$  упорядочены по их относительной важности.

На первом шаге решения задачи отбирают множество исходов, которые имеют максимальную оценку по важнейшему критерию. Если исход единственный, то он и является оптимальным. Если же исходов несколько, то среди них выбирают те, которые имеют максимальную оценку по

второму по важности критерию. Если опять исходов несколько, то процесс повторяют для следующих критерий.

*Метод идеальной точки (Метод Салуквадзе).* Метод идеальной точки является «геометрическим» методом для многокритериальных задач.

Метод состоит из двух этапов: на первом этапе решаются  $n$  задач вида:

$$\begin{cases} f_i^* = \max(\min) f_i(x) \\ x \in D \end{cases}$$

Здесь  $f_i^*$  - максимальное (минимальное) значение  $i$ -го критерия на множестве  $D$ . Значения  $f_i^*$  являются координатами некоторой «идеальной» точки  $f^*$ . Если эта точка принадлежит множеству  $D$ , то она и является наилучшей. Если же  $f^*$  не принадлежит множеству  $D$ , то на втором этапе находится точка из множества  $D$ , ближайшая к идеальной точке как решение задачи  $R(f(x), f^*) \rightarrow \min, x \in D$ , где  $R$  – расстояние от  $f(x)$  до  $f^*$ .

Продemonстрируем его на следующем примере:

*Пример:* Для выпуска двух видов продукции используется два вида ресурсов. Известны  $A = \begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix}$  – матрица норм расхода сырья,  $Q = (2, 1)$  – стоимость ресурсов,  $P = (6, 9)$  – цена реализации продукции,  $B = \begin{pmatrix} 18 \\ 16 \end{pmatrix}$  – запасы ресурсов. Найти план производства, максимизирующий одновременно выручку и прибыль (или, что эквивалентно, максимизирующий прибыль и минимизирующий затраты на ресурсы).

*Решение.* Если  $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$  – план производства, то стоимость ресурсов  $QAX = 4x_1 + 8x_2$ , выручка  $f_1 = PX = 6x_1 + 9x_2$ , прибыль  $f_2 = PX - QAX =$

$2x_1 + x_2$ . Расписав условие  $AX \leq B$ , т.е.  $\begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 18 \\ 16 \end{pmatrix}$ , получим

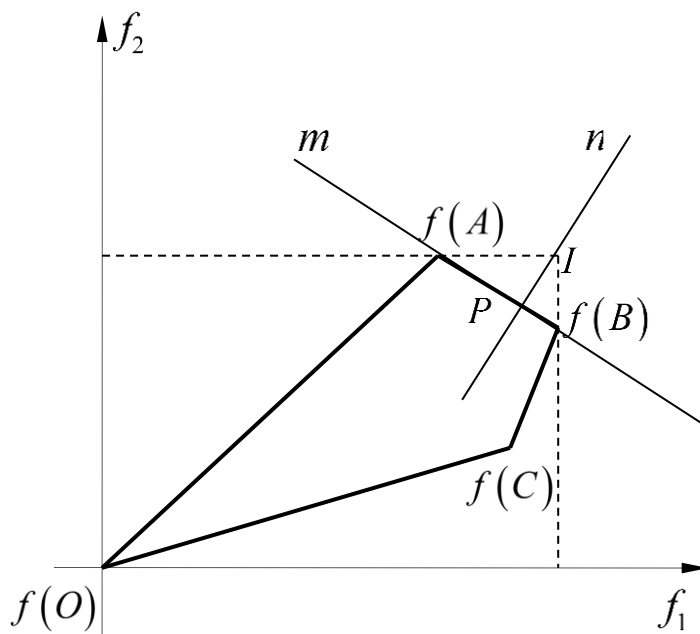
$$\text{задачу: } \begin{cases} f_1 = 6x_1 + 9x_2 \rightarrow \max, \\ f_2 = 2x_1 + x_2 \rightarrow \max, \\ x_1 + 3x_2 \leq 18, \\ x_1 + 2x_2 \leq 16, \\ x_1 \geq 0, x_2 \geq 0, \end{cases}.$$

Введем линейное преобразование  $f: R^2 \rightarrow R^2$ , определенное критериями  $f_1$  и  $f_2$ :

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} 6x_1 + 9x_2 \\ 2x_1 + x_2 \end{pmatrix} = \begin{pmatrix} 6 & 9 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

При этом  $f(O) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ ,  $f(A) = \begin{pmatrix} 6 & 9 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 54 \\ 6 \end{pmatrix}$ ,  $f(B) = \begin{pmatrix} 63 \\ 11 \end{pmatrix}$ ,  $f(C) = \begin{pmatrix} 48 \\ 16 \end{pmatrix}$ . В силу линейности  $f$  мы можем легко получить образ области  $D$  под действием преобразования  $f$  на плоскости  $(f_1, f_2)$  – это четырехугольник с вершинами в точках  $f(O)$ ,  $f(A)$ ,  $f(B)$ ,  $f(C)$ . При этом идеальной является точка  $I$  с координатами  $(f_{1,\max}, f_{2,\max})$ , не принадлежащая образу  $D$  в силу того, что оптимальные значения

соответствующих целевых функций достигаются в различных точках.



Естественно предположить, что образом компромиссной точки является точка, принадлежащая  $D$  и ближайшая к  $I$ , т.е. основание перпендикуляра, опущенного из  $I$  на отрезок, соединяющий точки  $f(A)$  и

$f(B)$ . Найдя уравнение прямой  $m$ , проходящей через эти точки, а затем

прямой  $n$ , ей перпендикулярной и проходящей через  $I$ , получим, что точка  $P = t \cap n$  имеет координаты  $P \left( \frac{123}{2}; \frac{23}{2} \right)$ . Наконец, находим компромиссную точку как прообраз  $P$ :  $x^* = f^{-1}(P) = \begin{pmatrix} 6 & 9 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 123/2 \\ 23/2 \end{pmatrix} = \begin{pmatrix} 7/2 \\ 9/2 \end{pmatrix}$ .

## Практико-ориентированные задания

*Задание 1.* Анализируется зависимость количества туристов ( $Y$ ), воспользовавшихся услугами фирмы, от затрат на рекламу ( $X$ , в тыс. у.е.). В качестве спецификации выбрана модель:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (1)$$

Результаты оценивания получены по выборочным данным, включающим 20 наблюдений:

$$\hat{Y}_t = -118,3 + 83,84 \cdot X_t, \quad \hat{\sigma} = 81,98, R^2 = 0,66, ESS = 120963,3, \quad (2)$$

(143,28)      (14,28)

где  $ESS$ — сумма квадратов остатков. Используя тест *RESET* проверьте правильность выбора спецификации (1), если оценка вспомогательной регрессии теста имеет вид:

$$Y_t = -372,71 + 200,0 \cdot X_t - 0,0007 \cdot \hat{Y}_t^2 + e_t, \quad R^2 = 0,66,$$

(1148,9)      (270,0)      (0,002)      (83,9)

$$ESS_{UR} = 119662,5,$$

где  $ESS_{UR}$ — сумма квадратов остатков вспомогательной регрессии. Критическое значение статистики теста равно  $F_{\alpha}(1; 17) = 4,45$  (при уровне значимости  $\alpha = 0,05$ ). Постройте прогноз для величины затрат на рекламу в 20 тыс. у.е.

*Решение.* Вычислим по формуле (15.13)  $F$ -статистику теста *RESET*, и сравним её с критическим значением:

$$F = \frac{(120963,3 - 119662,5)/1}{119662,5/17} = 0,186 < F_{\alpha=0,05}(1; 17) = 4,45,$$

следовательно, нулевая гипотеза теста не отклоняется, спецификация модели не нуждается в изменении. Прогноз количества туристов для величины затрат на рекламу в 20 тыс. у.е. равен  $\hat{Y}_t = -118,3 + 83,84 \cdot 20 \approx 1795$ .

*Задание 2.* Проверьте остатки модели (2) на гомоскедастичность при помощи теста Голдфелда Квандта, если суммы квадратов остатков вспомогательных моделей, оцененных по упорядоченным (в порядке возрастания) по  $X$  данным, равны:  $ESS_1 = 20400, ESS_2 = 76875$ ,

критическое значение статистики теста —  $F_{\alpha}(5; 5) = 5,05$  (при уровне значимости  $\alpha = 0,05$ ).

*Решение.* Вычислим по формуле (15.8)  $GQ$ -статистику, и сравним её с критическим значением:

$$GQ = ESS_2/ESS_1 = 76875/20400 = 3,77 < F_{\alpha=0,05}(5,5) = 5,05,$$

следовательно, нулевая гипотеза теста не отклоняется, возмущение — гомоскедастично.

*Задание 3.* Проверьте остатки модели (2) на отсутствие автокорреляции при помощи теста Дарбина-Уотсона, если  $\sum_{t=2}^{20} (e_t - e_{t-1})^2 = 266751,8$ , а нижняя и верхняя границы критического значения статистики равны:  $d_L = 1,201$ ,  $d_U = 1,411$ . Оцените коэффициент авторегрессии остатков первого порядка.

*Решение.* Вычислим по формуле (15.9)  $DW$ -статистику, и сравним её с верхней и нижней границами критического значения:

$$DW = \sum_{t=2}^n (e_t - e_{t-1})^2 / ESS = 266751,8 / 120969,3 = 2,21,$$

т.к.  $d_U = 1,411 < DW = 2,21 < 4 - d_U = 2,59$ , нулевая гипотеза теста не отклоняется, автокорреляция отсутствует (см. рис. 15.1). Используя приближенное равенство  $DW \approx 2(1 - \hat{\rho})$ , в формуле (15.9), оценим коэффициент авторегрессии остатков первого порядка:

$$\hat{\rho} \approx 1 - DW/2 = 1 - 2,21/2 = -0,105.$$

*Задание 4.* При исследовании зависимости потребления  $Y$  от дохода  $X$  по 12 наблюдениям получена следующая модель:

$$\hat{Y}_t = \underset{(12,0)}{-144,0} + \underset{(0,2)}{0,80} \cdot X_t, \quad t_{\alpha} = 3,1, \quad \hat{\sigma}^2 = 10,$$

$\alpha = 0,01$  — уровень значимости. Проверьте значимость оценки параметра при регрессоре и постройте для неё интервальную оценку. Оцените ковариацию между коэффициентами модели, если

$$(X^T X)^{-1} = \begin{pmatrix} 14,4 & -1,5 \\ -1,5 & 0,004 \end{pmatrix}.$$

*Решение.* Проверим статистическую значимость оценки параметра при регрессоре при помощи неравенства (16.20)

$$|t_{\hat{\beta}_2}| = |\hat{\beta}_2 / s_{\hat{\beta}_2}| = \frac{0,8}{0,2} = 4 > t_\alpha = 3,1,$$

следовательно, оценка параметра  $\hat{\beta}_2 = 0,8$  — статистически значима, и доход значимо влияет на потребление. Интервальная оценка параметра  $\hat{\beta}_2$  определяется границами (см. (16.18)):

$$\beta^- = \hat{\beta}_2 - t_{kp} \cdot s_{\hat{\beta}_2} = 0,8 - 3,1 \cdot 0,2 = 0,18,$$

$$\beta^+ = \hat{\beta}_2 + t_{kp} \cdot s_{\hat{\beta}_2} = 0,8 + 3,1 \cdot 0,2 = 1,42.$$

Автоковариационная матрица (16.5) оценок параметров модели равна

$$\hat{\sigma}^2(X^T X)^{-1} = 10 \begin{pmatrix} 14,4 & -1,5 \\ -1,5 & 0,004 \end{pmatrix}, \text{ и } cov\{\hat{\beta}_1, \hat{\beta}_2\} = -15.$$

*Задание 5.* Построить доверительный интервал для эндогенной переменной 16-го наблюдения  $Y_{t=16} = 2,5$  (тыс. руб.) по результатам оценивания модели зависимости накоплений от дохода по выборочным данным, включающим 15 наблюдений:

$$\hat{Y}_t = \underset{(s_{\hat{\beta}_1})}{\hat{\beta}_1} + \underset{(s_{\hat{\beta}_2})}{\hat{\beta}_2} \cdot X_t = \underset{(1,5)}{-2,0} + \underset{(0,05)}{0,14} \cdot X_t, \quad t_{\hat{\beta}_1} = -1,33, \quad t_{\hat{\beta}_2} = 2,8, \quad t_\alpha = 3,$$

$$Y_{t=16} = 2,5, \quad X_{t=16} = 30, \quad \sqrt{var(Y_{16} - \hat{Y}_{16})} = 1,2.$$

Проверить адекватность модели.

*Решение.* По условию задачи обучающая выборка включает 15 наблюдений, контролирующая — 1 наблюдение. Для построения доверительного интервала по формуле (16.21) для момента  $p = 16$ , вычислим прогноз эндогенной переменной на этот момент:

$$\hat{Y}_{16} = -2,0 + 0,14 \cdot X_{16} = -2,0 + 0,14 \cdot 30 = 2,2.$$

Так как  $s_p = 1,2$  (стандартная ошибка прогноза), границы доверительного интервала принимают значения:

$$Y^- = \hat{Y}_p - t_\alpha \cdot s_p = 2,2 - 3 \cdot 1,2 = -1,4;$$

$$Y^+ = \hat{Y}_p + t_\alpha \cdot s_p = 2,2 + 3 \cdot 1,2 = 5,8,$$

$$Y^- = -1,4 < Y_{p=16} = 2,5 < Y^+ = 5,8$$

следовательно, модель адекватна.

*Задание 6.* По ежегодным данным оценена модель зависимости объёма выпуска ( $Y$ , в % к уровню базового года) от инвестиционных расходов ( $X$ , млрд долл.) на приобретение новых заводов и оборудования:

$$\hat{Y}_t = 599,092 + 1,932 \cdot X_t + 0,695 \cdot X_{t-1} + 0,672 \cdot X_{t-2} + 1,863 \cdot X_{t-3}.$$

(51,840)      (0,177)      (0,129)      (0,121)      (0,172)

Определить и проинтерпретировать характеристики лаговой структуры данной модели.

*Решение.* Оценка краткосрочного мультипликатора  $\hat{\beta}_0 = 1,932$  показывает, что при увеличении инвестиционных расходов  $X$  на 1 млрд долл. в текущем периоде приводит к росту объёма выпуска продукции на 1,932 млрд долл. в этом же периоде; оценка долгосрочного мультипликатора

$$\hat{\beta} = \sum_{i=0}^k \hat{\beta}_i = 1,932 + 0,695 + 0,672 + 1,863 = 5,163,$$

означает, что увеличение инвестиций в экономику страны на 1 млрд долл. в текущем периоде приведет через 3 года к росту объёма выпуска на 5,163 млрд долл. По формуле (17.4) вычислим оценки относительных коэффициентов:

$$\hat{b}_0 = \hat{\beta}_0 / \hat{\beta} = 1,932 / 5,163 = 0,374; \quad \hat{b}_1 = \hat{\beta}_1 / \hat{\beta} = 0,695 / 5,163 = 0,135;$$

$$\hat{b}_2 = \hat{\beta}_2 / \hat{\beta} = 0,672 / 5,163 = 0,130; \quad \hat{b}_3 = \hat{\beta}_3 / \hat{\beta} = 1,863 / 5,163 = 0,361.$$

Относительные коэффициенты интерпретируются как доли (проценты) инвестиций, реализуемые в  $i$ -том году:  $\hat{b}_0 = 0,374$  — 37,4% инвестиций реализуются в виде объёма выпуска в текущем году,  $\hat{b}_1 = 0,135$  — 13,5% инвестиций реализуются в виде объёма выпуска в следующем году, и т.д. Средняя величина лага, вычисляемая по формуле (17.5),

$$\bar{l} = \sum_{i=1}^k i \cdot \hat{b}_i = 0 \cdot 0,374 + 1 \cdot 0,135 + 2 \cdot 0,130 + 3 \cdot 0,365 = 1,478,$$

показывает, что в среднем, увеличение инвестиций в экономику приведет к увеличению объёма выпуска через 1,48 года.



*Задание 7.* По ежегодным данным о динамике объемов ВВП США ( $Y$ , в % к уровню базового года) и валовых внутренних инвестиций в экономику США ( $X$ ) (в млрд. долл. США) за 32 года построена модель с распределёнными лагами:

$$Y_t = \alpha + \beta_0 \cdot X_t + \beta_1 \cdot X_{t-1} + \beta_2 \cdot X_{t-2} + \beta_3 \cdot X_{t-3} + \varepsilon_t$$

Оцените параметры этой модели методом Алмон с полиномом второй степени, если оценка вспомогательной регрессии

$$Y_t = \alpha + a_0 Z_{t0} + a_1 Z_{t1} + a_2 Z_{t2} + v_t$$

имеет вид:  $Y_t = 599,092 + 1,932Z_{0t} - 1,843Z_{1t} + 0,607Z_{2t} + \hat{v}_t$ .  
(51,840)      (0,177)      (0,368)      (0,121)      (31,807)

Определите стандартные ошибки оценок параметров  $\hat{\alpha}$  и  $\hat{\beta}_0$  заданной модели.

*Решение.* Оценки параметров исходной модели связаны с оценками параметров вспомогательной модели полиномом второй степени:

$$\beta_i = a_0 + a_1 i + a_2 i^2, \quad i = 0, \dots, 3,$$

$$\hat{\beta}_0 = \hat{a}_0 = 1,932, \quad \hat{\beta}_1 = \hat{a}_0 + \hat{a}_1 + \hat{a}_2 = 1,932 - 1,843 + 0,607 = 0,695,$$

$$\hat{\beta}_2 = \hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2 = 1,932 - 2 \cdot 1,843 + 4 \cdot 0,607 = 0,672,$$

$$\hat{\beta}_3 = \hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2 = 1,932 - 3 \cdot 1,843 + 9 \cdot 0,607 = 1,863.$$

Параметр  $\alpha$  исходной модели и вспомогательной в методе Алмон — совпадают, поэтому стандартные ошибки их оценок также совпадают,  $s_{\hat{\alpha}} = 51,840$ . Параметр  $\beta_0$  исходной модели равен параметру  $a_0$  вспомогательной модели, поэтому

$$Var\{\hat{\beta}\} = Var\{\hat{a}_0\}, \text{ и стандартная ошибка } s_{\hat{\beta}_0} = 0,177.$$

Таким образом, результат оценивания модели с распределёнными лагами методом Алмон можно представить в виде:

$$\hat{Y}_t = 599,092 + 1,932 \cdot X_t + 0,695 \cdot X_{t-1} + 0,672 \cdot X_{t-2} + 1,863 \cdot X_{t-3}.$$
(51,840)      (0,177)

*Задание 8.* Рассмотрим систему, включающую два уравнения:

$$\begin{cases} Y_{1t} &= a_{12}Y_{2t} + b_{11}X_{1t} + v_{1t} \\ Y_{2t} &= a_{21}Y_{1t} + b_{22}X_{2t} + v_{2t} \end{cases}, \quad (4)$$

с вектором эндогенных переменных  $Y_t = (Y_{1t}, Y_{2t})^T$  и вектором predetermined переменных  $X_t = (X_{1t}, X_{2t})^T$ , где  $Y_1$  — годовое потребление продукта на душу населения (в кг.);  $Y_2$  — оптовая цена за кг. (в долл.);  $X_1$  — доход на душу населения (в долл.);  $X_2$  — расходы по обработке продукта (в % к цене).

Система (4) получена в результате математической формализации экономических закономерностей, характеризующих моделируемый объект, и представляет собой *структурную форму спецификации*. Ее можно записать в матричном виде, предварительно перенеся все члены, кроме случайных возмущений, в левую часть:

$$\begin{cases} Y_{1t} - a_{12}Y_{2t} - b_{11}X_{1t} = v_{1t} \\ Y_{2t} - a_{21}Y_{1t} - b_{22}X_{2t} = v_{2t} \end{cases}.$$

Сформировав вектор эндогенных переменных модели, включающий два элемента ( $m = 2$ ),  $Y_t = (Y_{1t}, Y_{2t})^T$ , и вектор<sup>11</sup> predetermined переменных ( $k = 2$ ),  $X_t = (X_{1t}, X_{2t})^T$  заполним матрицы структурных параметров ( $A$  — при векторе  $Y$ ,  $B$  — при векторе  $X$ )

$$\begin{pmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} + \begin{pmatrix} -b_{11} & 0 \\ 0 & -b_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}. \quad (5)$$

Матрица  $A(m \times m)$  — квадратная (число строк равно числу уравнений системы, число столбцов равно числу эндогенных переменных<sup>12</sup>). Матрица  $B(m \times k)$  — прямоугольная (число строк равно числу уравнений, число столбцов равно размерности вектора predetermined переменных). Вектор эндогенных переменных системы (4) не выражен в явном виде через вектор predetermined переменных, поэтому для решения задачи оценки или прогнозирования необходимо составить *приведенную форму спецификации*:

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \quad (6)$$

<sup>11</sup> Вектор predetermined переменных включает дополнительный элемент — единицу, если хотя бы в одном из уравнений присутствует свободный член.

<sup>12</sup> В правильно составленной спецификации число уравнений равно числу эндогенных переменных.

Элементы матрицы  $M$  приведенных параметров связаны с элементами матриц  $A$  и  $B$  структурных параметров соотношением (18.3).

*Задание 9.* Оценить нижнюю границу стоимости четырехмесячного опциона покупателя на акцию, по которой не выплачиваются дивиденды, если цена акции 168, цена исполнения опциона — 150, безрисковая ставка — 8% годовых при непрерывном начислении процентов. Ответ обосновать. Описать и обосновать арбитражную стратегию, если опцион недооценен.

*Решение.* Для стоимости опционов продавца и покупателя справедливо соотношение

$$C_t - P_t = S_t - Ke^{-r(T-t)},$$

называемое паритетом цен. В приведенной формуле  $S_t$ ,  $C_t$  и  $P_t$  — цены в момент времени  $t$  соответственно на актив и опционы покупателя и продавца на этот актив с датой исполнения  $T$  и ценой исполнения  $K$ , а  $r$  — безрисковая процентная ставка. Так как  $P_t \geq 0$ , то

$$C_t \geq S_t - Ke^{-r(T-t)}.$$

С учетом этого получаем  $C_t \geq 168 - 150 \cdot e^{-0,08 \cdot 4/12} = 21,947$ . Предположим, что опцион продается по цене  $x < 21,947$ . Построим соответствующую арбитражную стратегию. Поскольку опцион недооценен, следует его купить. На это потребуется сумма  $x$ . Чтобы ее получить, нужно продать акцию, а разницу  $168 - x$  разместить по безрисковой ставке. В момент времени  $T$  стоимость опциона составит  $\max(0, S_T - 150)$ , на возврат базового актива потребуется потратить  $S_T$ , а сумма, размещенная по безрисковой ставке, вырастет до величины  $(168 - x)e^{0,08 \cdot 4/12}$ . Таким образом, на момент времени  $T$  баланс составит величину

$$V = \max(0, S_T - 150) + (168 - x)e^{0,08 \cdot 4/12} - S_T.$$

Как бы ни изменялась цена базового актива, эта величина будет положительной. В самом деле, если  $S_T > 150$ , то

$$V = S_T - 150 + (168 - x)e^{0,08 \cdot \frac{4}{12}} - S_T > (168 - 168 + 150 \cdot e^{-0,08 \cdot \frac{4}{12}}).$$

$$e^{0,08 \cdot \frac{4}{12}} - 150 = 0.$$

Если же  $S_T < 150$ , то

$$V = (168 - x)e^{0,08 \cdot \frac{4}{12}} - S_T > \left(168 - 168 + 150 \cdot e^{-0,08 \cdot \frac{4}{12}}\right) \cdot e^{0,08 \cdot \frac{4}{12}} - 150 = 0.$$

При реализации описанной стратегии портфель, стоивший 0 в начальный момент времени, имеет положительную стоимость через четыре месяца.

*Задание 10.* Составить спред типа «бабочка» из трех трехмесячных опционов покупателя с ценами исполнения 130, 135, 140. Цены этих опционов составляют соответственно 8, 4, 2. Представить графически зависимость дохода, который приносит спред, от цены базового актива.

*Решение.* Для составления спреда типа «бабочка» составим портфель из опционов с ценами исполнения 130 и 140 в длинной позиции и двух опционов с ценой исполнения 135 в короткой позиции. Затраты на составление портфеля составят величину  $8 + 2 - 2 \cdot 4 = 2$ . На момент исполнения опционов стоимость портфеля составит величину

$$V_T = \max(0, S_T - 130) + \max(0, S_T - 140) - 2 \cdot \max(0, S_T - 135) - 2e^{0,25r},$$

где  $S_T$  — цена базового актива, а  $r$  — безрисковая ставка

График зависимости  $V_T = V_T(S_T)$  представлен на рисунке.

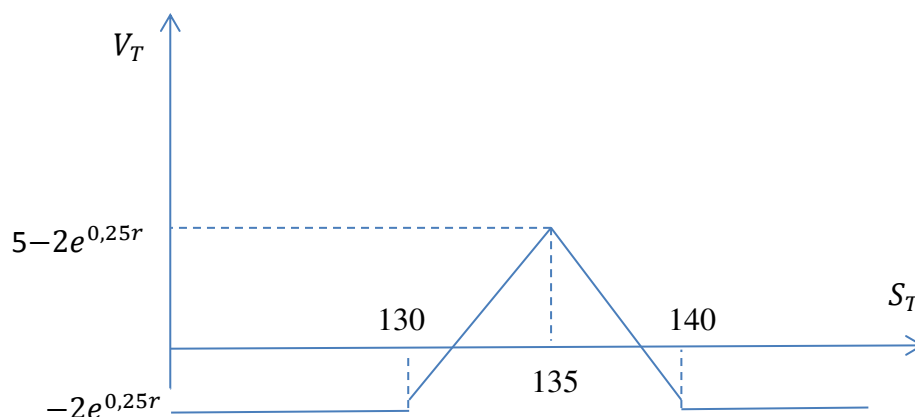


Рис. Спред «бабочка»

Если  $130 + 2e^{0,25r} < S_T < 140 - 2e^{0,25r}$ , стратегия принесет доход.

**Задание 11.** Составить бычий спрэд из двух трехмесячных опционов продавца с ценами исполнения 130, 135. Цены этих опционов составляют соответственно 4, 8. Представить графически зависимость дохода, который приносит спрэд, от цены базового актива.

**Решение.** Для составления бычьего спреда составим портфель из опционов с ценами исполнения 130 в длинной позиции и опциона с ценой исполнения 135 в короткой позиции. Доход от составления портфеля составит величину  $8 - 4 = 4$ . На момент исполнения опционов стоимость портфеля составит величину

$$V_T = \max(0; 130 - S_T) - \max(0; 135 - S_T) + 4e^{0,25r},$$

где  $S_T$  — цена базового актива, а  $r$  — безрисковая ставка

График зависимости  $V_T = V_T(S_T)$  представлен на рисунке.

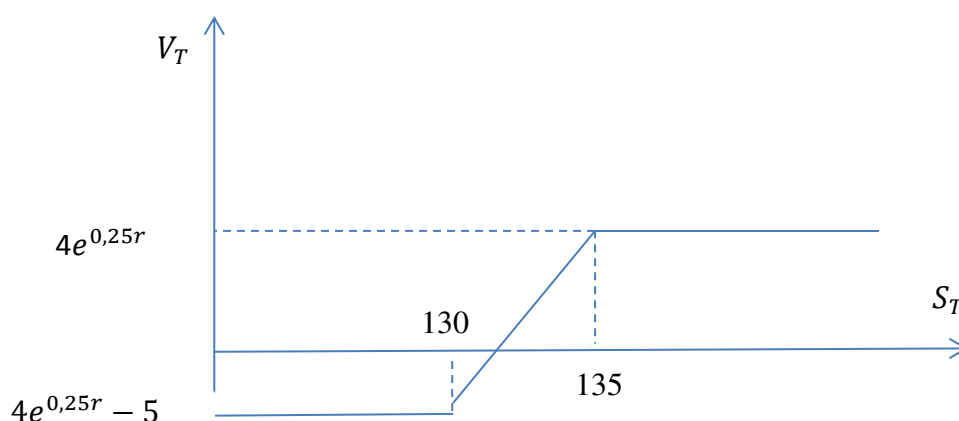


Рис. Бычий спрэд

Если  $135 - 4e^{0,25r} < S_T$ , стратегия принесет доход.

**Задание 12.** Текущая цена акции равна  $S$ . Используя трехпериодную биномиальную модель, оценить стоимость американского опциона "пут" с ценой исполнения  $K$  и сроком истечения через  $n$  месяцев, если годовая волатильность акции равна  $\sigma\%$ , а годовая безрисковая ставка при непрерывном начислении процентов составляет  $\rho\%$ .

Используя модель Блэка-Шоулза, вычислить стоимость аналогичных европейских опционов «пут» и «колл».

Данные для заданий приведены в таблице

$S$	$K$	$n$	$\sigma$	$\rho$
100	100	6	30	10.0

*Решение.* На рисунке схематически представлен процесс изменения цен рискового актива в трехпериодной модели.

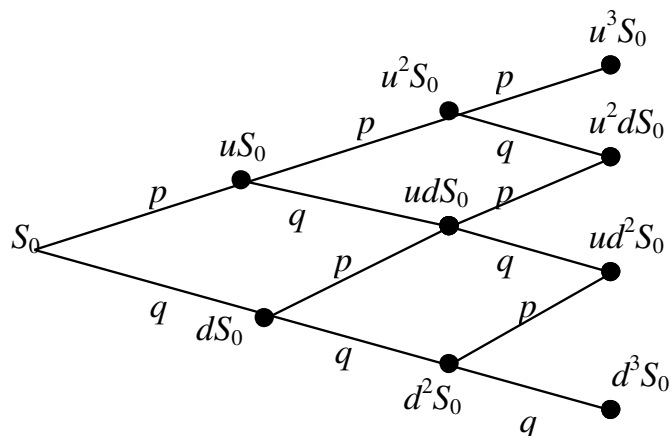


Рис. Трехпериодная модель

Один период соответствует двум месяцам. Определим параметры модели:

$$u = e^{\sigma\sqrt{T/3}} = e^{0,3 \cdot \sqrt{\frac{1}{3} \cdot \frac{6}{12}}} = 1,130; d = e^{-\sigma\sqrt{T/3}} = e^{-0,3 \cdot \sqrt{\frac{6}{12}}} = 0,885.$$

Коэффициент наращения по безрисковой ставке за один период составляет

$$1 + r = e^{\frac{\rho}{6}} = e^{\frac{0,1}{6}} = 1,017.$$

Вычислим риск-нейтральную вероятность:

$$p = \frac{1+r-d}{u-d} = \frac{1,017-0,885}{1,130-0,885} = 0,538; q = 0,462.$$

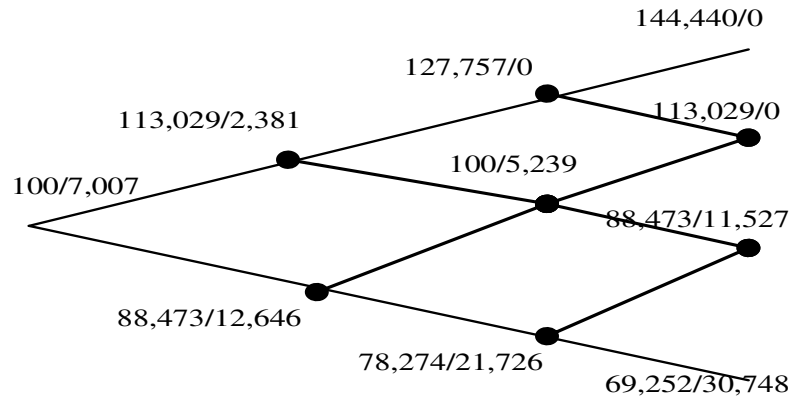


Рис. Расчет цены американского опциона

На рисунке в вершинах графа через дробную черту указаны цена базового актива и цена опциона. На момент истечения срока (в конце третьего периода) опцион пут будет исполнен в том случае, если цена рискованного актива окажется ниже цены исполнения опциона. Цена опциона в этом случае составит  $K - S_T$ . Соответственно, если цена базового актива повышалась в двух периодах, опцион не исполняется и его цена равна нулю. При понижении цены базового актива в двух периодах цена опциона оказывается равной 11,527, при понижении в трех периодах — 30,748.

Рассчитаем цены опциона на конец второго периода:

$$P_{uu} = 0;$$

$$P_{ud} = \max\left(K - S_{ud}, \frac{pP_{udu} + qP_{udd}}{1+r}\right) = 5,239$$

$$P_{dd} = \max\left(K - S_{dd}, \frac{pP_{ddu} + qP_{ddd}}{1+r}\right) = 21,726$$

Аналогичным образом рассчитываем цены на конец первого периода и цену опциона на текущий момент времени. В итоге получаем  $P^{am} = 7,007$ .

Вычислим теперь цену аналогичных европейских опционов, используя формулу Блэка-Шоулза. Для опционов «колл» и «пут» имеем соответственно

$$C = SN(d_1) - Ke^{-rT}N(d_2), P = -SN(-d_1) + Ke^{-rT}N(-d_2),$$

где

$$d_{1,2} = \frac{\ln \frac{S}{K} + (r \pm \sigma^2 / 2)T}{\sigma \sqrt{T}}.$$

Проведя расчеты, получаем:  $P = 6,029$ ;  $C = 10,907$

*Задание 13.* Текущая цена фьючерсного контракта на поставку актива равна  $F$ . Используя трехпериодную биномиальную модель, найти стоимость американского фьючерсного опциона пут с ценой исполнения  $K$  и датой истечения срока через  $n$  месяцев. Годовая волатильность базового актива составляет  $\sigma\%$ , безрисковая ставка при непрерывном начислении процентов равна  $\rho\%$ .

Используя модель Блэка-Шоулза, вычислить стоимость аналогичных европейских опционов «пут» и «колл».

Данные для заданий приведены в таблице

$F$	$K$	$n$	$\sigma$	$\rho$
100	101	6	40.0%	19.0%

*Решение.* Схема решения та же, что и в задаче 12.

Один период соответствует двум месяцам. Определим параметры модели:

$$u = e^{\sigma \sqrt{T/3}} = e^{0,4 \cdot \sqrt{\frac{1}{3} \cdot \frac{6}{12}}} = 1,177; d = e^{-\sigma \sqrt{T/3}} = e^{-0,4 \cdot \sqrt{\frac{6}{12}}} = 0,849.$$

Коэффициент дисконтирования по безрисковой ставке за один период составляет

$$(1 + r)^{-1} = e^{-\frac{\rho}{6}} = e^{-\frac{0,19}{6}} = 0,969.$$

Вычислим риск-нейтральную вероятность:

$$p = \frac{1-d}{u-d} = \frac{1-0,849}{1,177-0,849} = 0,459; q = 0,541.$$



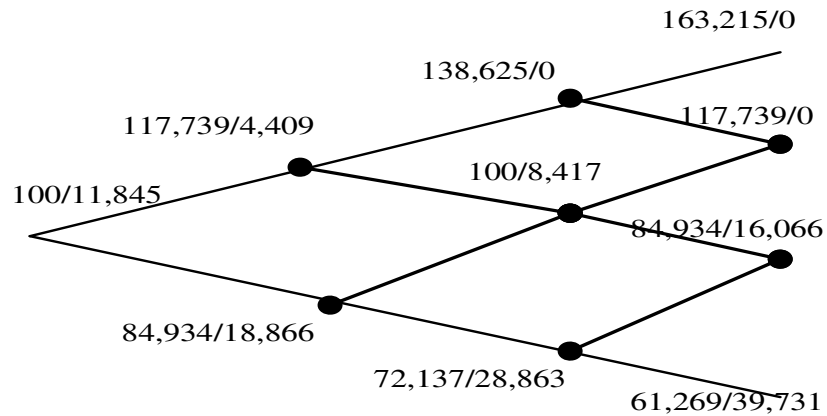


Рис. Расчет цены фьючерсного опциона

На рисунке в вершинах графа через дробную черту указаны цена базового актива и цена опциона.

Вычислим теперь цену аналогичных европейских опционов, используя формулу Блэка-Шоулза. Для опционов «колл» и «пут» имеем соответственно

$$C = SN(d_1) - Ke^{-rT}N(d_2), P = -SN(-d_1) + Ke^{-rT}N(-d_2),$$

где

$$d_{1,2} = \frac{\ln \frac{S}{K} \pm \frac{\sigma^2}{2} T}{\sigma \sqrt{T}}.$$

Проведя расчеты, получаем:  $P = 10,739$ ;  $C = 9,830$

**Задание 14.** Финансовый менеджер должен через 3 года произвести выплату в размере 200 000 у.е. Определить, сколько 2-летних бескупонных облигаций типа А (цена одной такой облигации составляет 857,3 у.е. и номинал 1000 у.е.) и сколько 4-летних облигаций типа В (купонная ставка – 10%, номинал 1000 у.е., доходность к погашению 8%) ему следует купить, чтобы защитить средства от изменений процентной ставки.

**Решение.** Дюрация долга составляет  $D = 3$  года (долг можно рассматривать как бескупонную облигацию со сроком погашения 3 года). Для облигации типа А дюрация  $D(A)$  составляет 2 года, а доходность к моменту погашения найдём из уравнения  $857,3 = \frac{1000}{(1+y)^2}$ , откуда  $y = 8\%$ .

Для облигации типа  $B$  при той же доходности найдем ее дюрацию  $D(B)$  и цену  $P(B)$ :

$$P(B) = \sum_{i=1}^4 \frac{0,1 \cdot 1000}{(1+0,08)^i} + \frac{1000}{(1+0,08)^4} = 1066,2;$$

$$D(B) = \frac{1}{1066,2} \left( \sum_{i=1}^4 i \cdot \frac{0,1 \cdot 1000}{(1+0,08)^i} + 4 \cdot \frac{1000}{(1+0,08)^4} \right) = 3,5.$$

Выравнивая дюрации долга и портфеля, состоящего из облигаций  $A$  и  $B$ , определим доли «иммунизированного» портфеля из следующей системы:

$$\begin{cases} w_A + w_B = 1, \\ 2w_A + 3,5w_B = 3. \end{cases}$$

Получаем:  $w_A = 33,33 \%$ ,  $w_B = 66,67 \%$ . Для формирования «иммунизированного» портфеля потребуется  $P = \frac{200000}{1,08^3} = 158\,766$  у.е. Из этих средств в облигации типа  $A$  нужно вложить  $P \cdot w_A = 52917$  у.е., а так как цена одной такой облигации составляет  $857,3$  у.е., то необходимо приобрести  $n_A = \frac{52917}{857,3} \approx 62$  облигации типа  $A$ . В облигации типа  $B$  потребуется вложить  $P \cdot w_B = 106306$  у.е. и с учетом цены будет приобретено  $n_B = \frac{106306}{1066,2} \approx 99$  облигаций типа  $B$ .

**Задание 15.** На рынке имеется акция, текущая цена которой равна 20. Ценовая динамика акции описывается геометрическим Броуновским движением с параметрами  $\mu$  и  $\sigma$ . Годовые опционы пут на эту акцию с ценой исполнения 20 и 40 продаются по цене соответственно 2 и 4. Составить бычий спрэд и найти условия на ожидаемую доходность  $\mu$  и волатильность  $\sigma$ , при которых стратегия даст неотрицательный результат с вероятностью более 50%.

**Решение.** Составим бычий спрэд. Для этого опцион с ценой исполнения 20, производим короткую продажу опциона с ценой исполнения 40 и вырученную сумму 2 размещаем по безрисковой ставке. На момент

исполнения опционов стоимость портфеля будет равна следующей величине:

$$V = \max(0, 20 - S) - \max(0, 40 - S) + 2e^r,$$

где  $S$  — цена акции на момент исполнения опционов,  $r$  — безрисковая ставка. Условие  $V \geq 0$  равносильно тому, что  $S \geq 40 - 2e^r$ .

Цена  $S$  является случайной величиной и имеет следующий вид:

$$S = 20 \cdot e^{\mu - \frac{\sigma^2}{2} + \sigma Z},$$

где  $Z$  — стандартная нормальная величина. Условие неотрицательности спреда равносильно тому, что

$$20 \cdot e^{\mu - \frac{\sigma^2}{2} + \sigma Z} \geq 40 - 2e^r.$$

Это означает, что

$$Z \geq \frac{1}{\sigma} \left( \ln \frac{40 - 2e^r}{20} - \mu + \frac{\sigma^2}{2} \right).$$

Вероятность такого события превысит 50%, если

$$\frac{1}{\sigma} \left( \ln \frac{40 - 2e^r}{20} - \mu + \frac{\sigma^2}{2} \right) < 0.$$

Это имеет место при условии, что ожидаемая доходность достаточно высока:

$$\mu > \ln \frac{40 - 2e^r}{20} + \frac{\sigma^2}{2}.$$

Например, при безрисковой ставке 10% и волатильности 40% ожидаемая доходность акции должна быть выше 77%.

*Задание 16.* Ценовая динамика двух акций описывается геометрическим Броуновским движением с общим Винеровским процессом. Ожидаемая доходность по акциям и волатильность акций составляют соответственно 10%, 20% и 20%, 40%. Цены акций в начальный момент совпадают. Найти вероятность того, что в момент времени  $t$  цена первой акции будет не ниже, чем цена второй.

*Решение.* Ценовую динамику акций можно представить следующими уравнениями

$$S_1(t) = Se^{\left(\mu_1 - \frac{\sigma_1^2}{2}\right)t + \sigma_1\sqrt{t}Z};$$

$$S_2(t) = Se^{\left(\mu_2 - \frac{\sigma_2^2}{2}\right)t + \sigma_2\sqrt{t}Z},$$

где  $Z$  — стандартная нормальная величина. Условие  $S_1(t) \geq S_2(t)$  равносильно тому, что

$$\left(\mu_1 - \frac{\sigma_1^2}{2}\right)t + \sigma_1\sqrt{t}Z \geq \left(\mu_2 - \frac{\sigma_2^2}{2}\right)t + \sigma_2\sqrt{t}Z,$$

т. е.

$$Z \leq \left(\frac{\sigma_2 + \sigma_1}{2} - (\mu_2 - \mu_1)\right)\sqrt{t}.$$

Подставив данные из условий задачи, получаем

$$Z \leq 0,2\sqrt{t}.$$

Следовательно,

$$\Pr[S_1(t) \geq S_2(t)] = N(0,2\sqrt{t}),$$

где  $N(x)$  — функция распределения стандартной нормальной величины. Например, вероятность того, что цена первой акции превзойдет цену второй акции через год, составляет 0,579, через три месяца — 0,534.

*Задание 17.* Ценовая динамика двух акций описывается геометрическим Броуновским движением с общим Винеровским процессом. Волатильность первой акции составляет 20%, ожидаемая доходность и волатильность второй акции составляют соответственно 20% и 40%. Цены акций в начальный момент совпадают. Найти наименьшее значение ожидаемой доходности первой акции  $\mu_1$ , при котором выполняется следующее условие: вероятность того, что в момент времени  $t$  цена первой акции будет не ниже, чем цена второй, больше или равна 75%.

*Решение.* Ценовую динамику акций можно представить следующими уравнениями

$$S_1(t) = Se^{\left(\mu_1 - \frac{\sigma_1^2}{2}\right)t + \sigma_1\sqrt{t}Z};$$

$$S_2(t) = Se^{\left(\mu_2 - \frac{\sigma_2^2}{2}\right)t + \sigma_2\sqrt{t}Z},$$

где  $Z$  — стандартная нормальная величина. Условие  $S_1(t) \geq S_2(t)$  равносильно тому, что

$$\left(\mu_1 - \frac{\sigma_1^2}{2}\right)t + \sigma_1\sqrt{t}Z \geq \left(\mu_2 - \frac{\sigma_2^2}{2}\right)t + \sigma_2\sqrt{t}Z.$$

Подставив данные из условий задачи, получаем

$$(\mu_1 - 0,02)t + 0,2\sqrt{t}Z \geq 0,12t + 0,4\sqrt{t}Z.$$

Следовательно,

$$\Pr[S_1(t) \geq S_2(t)] = N((5\mu_1 - 0,7)\sqrt{t}),$$

где  $N(x)$  — функция распределения стандартной нормальной величины.  $N(x) = 0,75$  при  $x = 0,674$ . Значит,  $\Pr[S_1(t) \geq S_2(t)] \geq 0,75$  при  $(5\mu_1 - 0,7)\sqrt{t} \geq 0,674$ . Таким образом,

$$\mu_1 \geq 0,14 + \frac{0,135}{\sqrt{t}}.$$

Следовательно,  $\mu_1 = 0,14 + \frac{0,135}{\sqrt{t}}$  — наименьшее значение ожидаемой доходности первой акции, при котором  $\Pr[S_1(t) \geq S_2(t)] \geq 0,75$ . Например, если  $t = 1$ , то  $\mu_1 = 27,5\%$ .

*Задание 18.* Величина ущерба при пожаре (в случае, если он произошел) имеет плотность  $f_\eta(x) = ce^{-0,02x}$  при  $x > 0$ ,  $f_\eta(x) = 0$  в остальных случаях. Страховая компания установила верхний предел своей ответственности 34,6574 у.е.

а) Найти средний размер выплат компании по одному страховому случаю для нового договора.

б) Чему равна вероятность того, что выплата по этому договору равна 34,6574.

в) Найти функцию распределения выплат по новому договору.

*Решение.* Речь идет о договоре с лимитом  $l = 34,6574$ .  $\eta$  экспоненциально распределена и  $c = 0,02$ .

а) Средний размер выплат:

$$E(\eta \wedge l) = \int_0^l x f_\eta(x) dx + l(1 - F_\eta(l)) = \int_0^l (1 - F_\eta(x)) dx =$$

$$= \int_0^{34,6574} e^{-0,02x} dx = 50 \cdot \left(1 - e^{-\frac{34,6574}{50}}\right) = 25.00002049.$$

б) Распределение величины страховой выплаты:

$$F_{\eta \wedge l}(x) = \begin{cases} 1, & x \geq l, \\ F_{\eta}(x) = 1 - e^{-0,02x}, & x < l. \end{cases}$$

в) Вероятность того, что страховая выплата составит  $l = 34,6574$ :

$$P(\eta \wedge l = l) = P(\eta \geq l) = 1 - F_{\eta}(l) = e^{-0,02l} = 0,49999959.$$

*Задание 19.* Величина возмещения по договору страхования описывается случайной величиной с плотностью  $f_{\eta}(x) = \frac{1}{400}$  при  $0 < x < 400$ ,  $f_{\eta}(x) = 0$  в остальных случаях. Если договор предусматривает вычет по выплатам, и ожидаемое значение выплат страховщика по одному договору равно 128, то чему равен вычет? Найти функцию эксцедента убытков. Найти также плотность реальных выплат (без учета 0) по одному договору страхования.

*Решение.* Речь идет о договоре с вычетом  $d$ .

а) Нахождение вычета  $d$ :

$$\begin{aligned} E[(\eta - d)_+] &= \int_d^{400} (x - d) f_{\eta}(x) dx = \frac{1}{400} \int_d^{400} (x - d) dx = \\ &= \frac{(x - d)^2}{800} \Big|_d^{400} = \frac{(400 - d)^2}{800}. \end{aligned}$$

Дано

$$\frac{(400 - d)^2}{800} = 128 \Rightarrow$$

$$400 - d = \pm 320 \Rightarrow$$

$$d = 720 \text{ или } 80$$

720 исключается (вычет не может быть больше максимальной выплаты 400 по первичному договору). Отсюда

$$d = 80.$$

б) Функция эксцедента убытков  $e_\eta(d)$  (здесь вычет произвольный):

$$e_\eta(d) = \mathbf{E}(\eta - d | \eta > d) = \frac{\mathbf{E}[(\eta - d)_+]}{1 - F_\eta(d)} = \frac{\frac{(400 - d)^2}{800}}{1 - \frac{d}{400}} = \frac{400 - d}{2}.$$

в) Плотность реальных выплат:

$$f_{\eta_{(d)}}(y) = \frac{f_\eta(y + d)}{1 - F_\eta(d)} = \frac{\frac{1}{400}}{1 - \frac{80}{400}} = \frac{1}{320}, \quad y \in (0, 320).$$

*Задание 20.* Известно, что убытки страхователя  $\eta$  по договору следуют экспоненциальному распределению со средним 1. Пусть новый договор предусматривает франшизу  $d_f = 1$ . Найти

а) функцию распределения убытков страховщика с учетом нулевой выплаты. Найти непрерывную и дискретную части выплат;

б) среднее значение выплат для договора с франшизой (с учетом нулевой выплаты);

в) функцию распределения, плотность и среднее значение выплат без учета нулевой выплаты.

*Решение.* а) Дано

$$\begin{aligned} f_\eta(x) &= e^{-x}, & x > 0, \\ F_\eta(x) &= 1 - e^{-x}, & x > 0. \end{aligned}$$

Отсюда

$$\begin{aligned} F_{\eta_{d_f}}(y) &= \mathbf{P}(\eta_{d_f} \leq y) = \begin{cases} 0, & y < 0, \\ F_\eta(d_f), & 0 \leq y \leq d_f, \\ F_\eta(y), & y > d_f, \end{cases} \\ &= \begin{cases} 0, & y < 0, \\ 1 - e^{-1}, & 0 \leq y \leq 1, \\ 1 - e^{-y}, & y > 1. \end{cases} \end{aligned}$$

Дискретная часть  $\eta_{d_f}$  имеет функцию вероятностей

$$p_{\eta_{d_f}}(y) = \begin{cases} 1 - e^{-1}, & y = 0, \\ 0, & \text{в остальных случаях.} \end{cases}$$

Непрерывная часть  $\eta_{d_f}$  имеет плотность

$$f_{\eta_{d_f}}(y) = \begin{cases} f_{\eta}(y), & y > d_f, \\ 0, & \text{в остальных случаях} \end{cases} = \\ = \begin{cases} e^{-y}, & y > 1, \\ 0, & \text{в остальных случаях.} \end{cases}$$

б)

$$\mathbf{E}(\eta_{d_f}) = \int_{d_f}^{\infty} y f_{\eta}(y) dy = \int_1^{\infty} y e^{-y} dy = 2e^{-1}.$$

в)

$$F_{\eta_{d_f}^+}(y) = \begin{cases} 0, & 0 \leq y \leq d_f \\ \frac{F_{\eta}(y) - F_{\eta}(d_f)}{1 - F_{\eta}(d_f)}, & y > d_f \end{cases} = \\ = \begin{cases} 0, & 0 \leq y \leq 1, \\ 1 - e^{-(y-1)}, & y > 1. \end{cases}$$

Отсюда

$$f_{\eta_{d_f}^+}(y) = e^{-(y-1)}, y > 1.$$

Далее

$$\mathbf{E}(\eta_{d_f}^+) = \frac{\mathbf{E}(\eta_{d_f})}{1 - F_{\eta}(d_f)} = \frac{2e^{-1}}{e^{-1}} = 2.$$

**Задание 21.** Портфель состоит из 1000 одинаковых договоров страхования. Вероятность предъявления иска одинакова для всех договоров и равна 0,15. Для каждого договора страховая сумма при наступлении страхового случая имеет распределение Парето с параметрами  $\alpha = 3$ ,  $\theta = 100$ . Найти ожидаемое значение и дисперсию величины суммарного иска.

*Решение.* Дано:

портфель состоит из  $n = 1000$  договоров;

$q_k = q = 0,15$  – вероятность предъявления иска;

$\eta_k$  распределено как  $\eta \sim \text{Paerto}(\underbrace{3}_{\alpha}, \underbrace{100}_{\theta})$ .

Требуется найти  $\mathbf{E}(S_{1000})$ ,  $\mathbf{Var}(S_{1000})$ .

Плотность и функция распределения для  $\eta$  имеют соответственно вид



$$f_{\eta}(x) = \frac{\alpha}{\theta} \left( \frac{\theta}{x + \theta} \right)^{\alpha+1} = \frac{3}{100} \left( \frac{100}{x + 100} \right)^4,$$

$$F_{\eta}(x) = 1 - \left( \frac{100}{x + 100} \right)^3.$$

Вероятность того, что не будет предъявлен иск, равна

$$p = 1 - q = 0,85.$$

Далее

$$\mathbf{E}(\eta) = m_{\eta} = \int_0^{+\infty} (1 - F_{\eta}(x)) \, dx = \int_0^{+\infty} \left( \frac{100}{x + 100} \right)^3 \, dx = 50,$$

$$\mathbf{E}(\eta^2) = 2 \int_0^{+\infty} x(1 - F_{\eta}(x)) \, dx = 2 \int_0^{+\infty} x \left( \frac{100}{x + 100} \right)^3 \, dx = 10000.$$

Отсюда

$$\mathbf{Var}(\eta) = 10000 - 50^2 = 7500.$$

Найдем теперь  $\mathbf{E}(S_{1000})$  и  $\mathbf{Var}(S_{1000})$ . Имеем

$$\mathbf{E}(S_{1000}) = nqm_{\eta} = 1000 \cdot 0,15 \cdot \frac{100}{2} = 7500,$$

$$\begin{aligned} \mathbf{Var}(S_{1000}) &= npqm_{\eta}^2 + nq\sigma_{\eta}^2 = \\ &= 1000 \cdot 0,85 \cdot 0,15 \left( \frac{100}{2} \right)^2 + 1000 \cdot 0,15 \cdot \frac{3 \cdot 100^2}{2^2 \cdot 1} = 1\,443\,750. \end{aligned}$$

*Задание 22.* В рамках индивидуальной модели риска вероятность наступления страхового случая равна 0,1. После наступления страхового случая сумма возмещения равна 1000 долл. Исследование показало, что компания может продать полис не более чем за 115 долл. Сколько полисов должна продать компания для того, чтобы она с вероятностью 95% могла выполнить свои обязательства (без учета расходов)? При решении задачи использовать нормальную аппроксимацию.

*Решение.* Имеем

$$S_n = \xi_1 + \dots + \xi_n = 1000(I_1 + \dots + I_n),$$

где  $\mathbf{P}(I = 1) = 0,1 = q$ ,  $\mathbf{P}(I = 0) = 0,9 = p$ . Требуется найти  $n$ :

$$\mathbf{P}(S_n \leq 115 \cdot n) = 0,95.$$

Имеем

$$\begin{aligned} \mathbf{E}(S_n) &= 100n, \\ \mathbf{Var}(S_n) &= 1000^2 \cdot n \cdot \mathbf{Var}(I) = 1000^2 \cdot n \cdot p \cdot q = 90\,000 \cdot n, \\ \mathbf{P}(S_n \leq 115 \cdot n) &= \mathbf{P}\left(\frac{S_n - \mathbf{E}(S_n)}{\sqrt{\mathbf{Var}(S_n)}} \leq \frac{115 \cdot n - 100n}{300\sqrt{n}}\right) = \\ &= \mathbf{P}\left(\frac{S_n - \mathbf{E}(S_n)}{\sqrt{\mathbf{Var}(S_n)}} \leq \frac{15 \cdot \sqrt{n}}{300}\right) = 0,95. \end{aligned}$$

Следовательно

$$\frac{15 \cdot \sqrt{n}}{300} = 1,645.$$

Отсюда

$$n \approx 1083.$$

*Задание 23.* Портфель состоит из двух одинаковых договоров. Риски по двум договорам независимы. Вероятность наступления страхового случая равна 0,1. Для каждого договора страховая сумма при наступлении страхового случая имеет экспоненциальное распределение со средним 1.

- а) Найти функцию распределения величины суммарного иска.
- б) Найти зависимость вероятности разорения от имеющегося капитала.
- в) Каким начальным капиталом должна обладать страховая компания, чтобы она с вероятностью 0,95 могла выполнить свои обязательства?
- г) Найти вероятность того, что суммарный иск равен 0.
- д) Найти ожидаемое значение и дисперсию суммарного иска.

*Решение.* Пусть  $\eta_k$  ( $\xi_k$  соответственно) – величина реального убытка (величина иска соответственно) -го договора  $k = 1, 2$ . Реальные выплаты имеют соответственно плотность и функцию распределения ( $\eta_1 \sim \text{Exp}(1)$ ,  $\eta_2 \sim \text{Exp}(1)$ )

$$f_{\eta_k}(u) = \begin{cases} 0, & u < 0, \\ e^{-u}, & u \geq 0, \end{cases}$$

$$F_{\eta_k}(u) = \begin{cases} 0, & u < 0, \\ 1 - e^{-u}, & u \geq 0. \end{cases}$$

Имеем при  $u \geq 0$

$$\begin{aligned} f_{\eta_1 + \eta_2}(u) &= \int_0^u f_{\eta_1}(u-y)f_{\eta_2}(y)dy = \int_0^u e^{-(u-y)}e^{-y}dy = \\ &= e^{-u} \int_0^u 1dy = ue^{-u}. \end{aligned}$$

Отсюда

$$\begin{aligned} f_{\eta_1 + \eta_2}(u) &= \begin{cases} 0, & u < 0, \\ ue^{-u}, & u \geq 0, \end{cases} \\ F_{\eta_1 + \eta_2}(u) &= \begin{cases} 0, & u < 0, \\ \int_0^u ye^{-y} dy, & u \geq 0 \end{cases} = \\ &= \begin{cases} 0, & u < 0, \\ 1 - e^{-u}(1+u), & u \geq 0. \end{cases} \end{aligned}$$

а) Дано  $q_1 = q_2 = q = 0,1$ ,  $p_1 = p_2 = p = 0,9$ . Далее

$$\begin{aligned} F_{S_2}(u) &= 1 - q_1p_2 - q_2p_1 - q_1q_2 + q_1p_2F_{\eta_1}(u) + \\ &\quad + q_2p_1F_{\eta_2}(u) + q_1q_2F_{\eta_1 + \eta_2}(u) = \\ &= 0,81 + 0,09F_{\eta_1}(u) + 0,09F_{\eta_2}(u) + 0,01F_{\eta_1 + \eta_2}(u) = \\ &= \left( F_{\eta_1}(u) = F_{\eta_2}(u) \right) = 0,81 + 0,18F_{\eta_1}(u) + 0,01F_{\eta_1 + \eta_2}(u) = \\ &= \begin{cases} 0, & u < 0, \\ 1 - 0,19e^{-u} - 0,01ue^{-u}, & u \geq 0. \end{cases} \end{aligned}$$

Здесь  $F_{S_2}(u) = 0$  при  $u < 0$ , поскольку  $S_2$  принимает неотрицательные значения.

б)

$$\begin{aligned} R(u) &= 1 - F_{S_2}(u) = \\ &= \begin{cases} 1, & u < 0, \\ 0,19e^{-u} + 0,01ue^{-u}, & u \geq 0. \end{cases} \end{aligned}$$

в) Требуется найти  $u_0$ :  $F_{S_2}(u_0) = 0,95$ . Имеем

$$1 - 0,19e^{-u} - 0,01ue^{-u} = 0,95.$$

Корень этого уравнения находится численно:  $u_0 = 1,406412584$ .

г)  $P(S_2 = 0) = 0,81$  (величина скачка  $F_{S_2}(u)$  в точке 0).

д)

$$E(S_2) = 2qm_{\eta_k} = 2 \cdot 0,1 \cdot 1 = 0,2,$$

$$\begin{aligned} \text{Var}(S_2) &= 2pqm_{\eta_k}^2 + 2q\sigma_{\eta_k}^2 = \\ &= 2 \cdot 0,9 \cdot 0,1 \cdot 1^2 + 2 \cdot 0,1 \cdot 1 = 0,38. \end{aligned}$$

**Задание 24.** (данные задачи собраны в таблице ниже). Портфель состоит из 10 000 договоров страхования. Выплаты осуществляются по двум страховым случаям. Вероятность наступления первого страхового случая одинакова для всех договоров. Относительно вероятности наступления второго страхового случая портфель был разбит на две группы по 4000 и 6000 договоров страхования. В каждой группе вероятность наступления второго страхового случая одинакова для всех договоров. Найти премию относительно различных принципов, чтобы (для принципов с надбавкой) страховая компания могла выполнить свои обязательства с вероятностью 95% без привлечения дополнительных средств.

	Группа 1	Группа 2	
Число договоров	4000	6000	Выплата
Вероятность первого страхового случая	0,0005	0,0005	1000 000
Вероятность второго страхового случая	0,004	0,002	250 000

**Решение.** Найдем сначала  $E(\xi_k)$ ,  $\text{Var}(\xi_k)$ ,  $E(S_n)$ ,  $\text{Var}(S_n)$ .

Распределение  $\xi_k$  при  $k = 1, \dots, 4000$  имеет вид

$\xi_k$	0	1000 000	250 000
$P$	0,9955	0,0005	0,004

Распределение  $\xi_k$  при  $k = 4001, \dots, 10000$  имеет вид

$\xi_k$	0	1000 000	250 000
$P$	0,9975	0,0005	0,002

При  $k = 1, \dots, 4000$  имеем (верхний индекс указывает, для какой группы производятся расчеты)

$$\mathbf{E}(\xi_k^{(1)}) = 1000\,000 \cdot 0,0005 + 250\,000 \cdot 0,004 = 1500,$$

$$\mathbf{Var}(\xi_k^{(1)}) = 10^{12} \cdot 0,0005 + 25^2 \cdot 10^8 \cdot 0,004 - 1500^2 = 747\,750\,000.$$

При  $k = 4001, \dots, 10000$  имеем

$$\mathbf{E}(\xi_k^{(2)}) = 1000\,000 \cdot 0,0005 + 250\,000 \cdot 0,002 = 1000,$$

$$\mathbf{Var}(\xi_k^{(2)}) = 10^{12} \cdot 0,0005 + 25^2 \cdot 10^8 \cdot 0,002 - 1000^2 = 624 \cdot 10^6.$$

Отсюда

$$\mathbf{E}(S_n) = \sum_{k=1}^n \mathbf{E}(\xi_k) = 4000 \cdot 1500 + 6000 \cdot 1000 = 12 \cdot 10^6,$$

$$\begin{aligned} \mathbf{Var}(S_n) &= \sum_{k=1}^n \mathbf{Var}(\xi_k) = 4000 \cdot 747\,750\,000 + 6000 \cdot 624 \cdot 10^6 \\ &= 6735 \cdot 10^9 \end{aligned}$$

### 1. Принцип нетто-премии

Премия для  $k$ -го договора при  $k = 1, \dots, 4000$  равна

$$P_k^{(1)} = \mathbf{E}(\xi_k^{(1)}) = 1500.$$

Премия для  $k$ -го договора при  $k = 4001, \dots, 6000$  равна

$$P_k^{(2)} = \mathbf{E}(\xi_k^{(2)}) = 1000.$$

Отсюда

$$\mathbf{E}(S_n) = \sum_{k=1}^n \mathbf{E}(\xi_k) = \sum_{k=1}^n P_k = 12 \cdot 10^6 = u,$$

$$R(u) = \mathbf{P}(S_n > u) = \mathbf{P}\left(\frac{S_n - \mathbf{E}(S_n)}{\sqrt{\mathbf{Var}(S_n)}} > 0\right) \approx \Phi(0) = \frac{1}{2}.$$

## 2. Принцип математического ожидания

$$x_\alpha = x_{0,95} = 1,64485.$$

Относительная страховая надбавка для всех договоров равна

$$\lambda = \frac{x_\alpha \sqrt{\mathbf{Var}(S_n)}}{\mathbf{E}(S_n)} = \frac{1,64485 \sqrt{6735 \cdot 10^9}}{12 \cdot 10^6} = 0,3557.$$

Отсюда находим индивидуальную защитную страховую надбавку

$$l_k = \lambda \mathbf{E}(\xi_k):$$

при  $k = 1, \dots, 4000$  имеем

$$l_k^{(1)} = 0,3557 \cdot 1500 = 533,588,$$

при  $k = 4001, \dots, 10000$  имеем

$$l_k^{(2)} = 0,3557 \cdot 1000 = 355,72535.$$

Соответственно премия для первой и второй группы договоров равна

$$P_k^{(1)} = (1 + \lambda) \mathbf{E}(\xi_k) = 1500 + 533,588 = 2033,59,$$

$$P_k^{(2)} = (1 + \lambda) \mathbf{E}(\xi_k) = 1000 + 533,588 = 1355,73.$$

Суммарная защитная надбавка равна

$$l = x_\alpha \sqrt{\mathbf{Var}(S_n)} = 1,64485 \cdot \sqrt{6735 \cdot 10^9} = 4\,268\,704,16.$$

$l$  можно найти также по правилу

$$l = 4000 l_k^{(1)} + 6000 l_k^{(2)}.$$

Суммарная премия совпадает с  $u$ :

$$u = 4000 \cdot P_k^{(1)} + 6000 \cdot P_k^{(2)} = 16\,268\,704,16.$$

## 3. Принцип дисперсии

Имеем

$$\lambda = \frac{x_\alpha}{\sqrt{\mathbf{Var}(S_n)}} = \frac{1,64485}{\sqrt{6735 \cdot 10^9}} = 6,33809 \cdot 10^{-7},$$

$$l_k^{(1)} = \lambda \mathbf{Var}(\xi_k^{(1)}) = 747\,750\,000 \cdot 6,33809 \cdot 10^{-7} = 473,93074,$$

$$k = 1, \dots, 4000,$$

$$l_k^{(2)} = \lambda \mathbf{Var}(\xi_k^{(2)}) = 624 \cdot 10^6 \cdot 6,33809 \cdot 10^{-7} = 395,49686,$$

$$k = 4001, \dots, 10000.$$

Соответственно премия для первой и второй группы договоров равна

$$P_k^{(1)} = \mathbf{E}(\xi_k^{(1)}) + \underbrace{\lambda \mathbf{Var}(\xi_k^{(1)})}_{l_k^{(1)}} = 1500 + 473,93074 = 1973,93074,$$

$$k = 1, \dots, 4000,$$

$$P_k^{(2)} = \mathbf{E}(\xi_k^{(2)}) + \underbrace{\lambda \mathbf{Var}(\xi_k^{(2)})}_{l_k^{(2)}} = 1000 + 395,49686 = 1395,49686,$$

$$k = 4001, \dots, 10000.$$

Суммарная защитная надбавка равна

$$l = \lambda \mathbf{Var}(S_n) = x_\alpha \sqrt{\mathbf{Var}(S_n)} = 1,64485 \cdot \sqrt{6735 \cdot 10^9} = 4\,268\,704,16.$$

$l$  можно найти также по правилу

$$l = 4000l_k^{(1)} + 6000l_k^{(2)}.$$

Суммарная премия совпадает с  $u$ :

$$u = 4000 \cdot P_k^{(1)} + 6000 \cdot P_k^{(2)} = 16\,268\,704,16.$$

#### 4. Принцип стандартного отклонения

Имеем

$$\begin{aligned} \lambda &= \frac{x_\alpha \sqrt{\mathbf{Var}(S_n)}}{\sum_{k=1}^n \sqrt{\mathbf{Var}(\xi_k)}} = \\ &= \frac{1,64485 \cdot \sqrt{6735 \cdot 10^9}}{4000 \cdot \sqrt{747\,750\,000} + 6000 \cdot \sqrt{624 \cdot 10^6}} = 0,016465. \end{aligned}$$

Отсюда

$$l_k^{(1)} = \lambda \sqrt{\mathbf{Var}(\xi_k^{(1)})} = 450,23444, \quad k = 1, \dots, 4000,$$

$$l_k^{(2)} = \lambda \sqrt{\mathbf{Var}(\xi_k^{(2)})} = 411,2944, \quad k = 4001, \dots, 10000,$$

$$l = 4000l_k^{(1)} + 6000l_k^{(2)} = 4\,268\,704,16.$$

Соответственно премия для первой и второй группы договоров равна

$$P_k^{(1)} = \mathbf{E}(\xi_k^{(1)}) + \lambda \underbrace{\sqrt{\mathbf{Var}(\xi_k^{(1)})}}_{l_k} = 1950,23444, \quad k = 1, \dots, 4000,$$

$$P_k^{(2)} = \mathbf{E}(\xi_k^{(2)}) + \lambda \underbrace{\sqrt{\mathbf{Var}(\xi_k^{(2)})}}_{l_k} = 1411,2944, \quad k = 4001, \dots, 10000.$$

Суммарная премия совпадает с  $u$ :

$$u = 4000 \cdot P_k^{(1)} + 6000 \cdot P_k^{(2)} = 16\,268\,704,16.$$

*Задание 25.* Число страховых случаев за один год имеет пуассоновское распределение со средним значением 500. Выплаты по искам независимы и равномерно распределены на  $[0,500]$ . Выплаты не зависят от числа исков.

а) Найти среднее значение и дисперсию суммарных выплат.

б) Предположим, что компания выделила сумму 140 000 долл. для обеспечения выплат в течение года. Найти вероятность того, что компания сможет выполнить свои обязательства (использовать нормальное приближение).

*Решение.*

а) Имеем

$$\mathbf{E}(v) = \mathbf{V}(v) = \lambda = 500,$$

$$\mathbf{E}(\eta) = 250, \quad \mathbf{V}(\eta) = \frac{500^2}{12},$$

$$\mathbf{E}(S_v) = \lambda \mathbf{E}(\eta) = 500 \cdot 250 = 125\,000,$$

$$\begin{aligned} \mathbf{V}(S_v) &= \mathbf{E}(v)\mathbf{V}(\eta) + \mathbf{E}(\eta)^2\mathbf{V}(v) = 500 \cdot \frac{500^2}{12} + 250^2 \cdot 500 \\ &= 41\,666\,666,667. \end{aligned}$$

б)

$$\begin{aligned} &\mathbf{P}(S_v \leq 140\,000) = \\ &= \mathbf{P}\left(\frac{S_v - \mathbf{E}(S_v)}{\sqrt{\mathbf{V}(S_v)}} \leq \frac{140\,000 - 125\,000}{\sqrt{41\,666\,666,667}}\right) = \mathbf{P}(Z \leq 2,32) = 0,99, \end{aligned}$$

где  $Z \sim \mathbf{N}(0,1)$ .



*Задание 26.* Суммарные выплаты по двум портфелям независимы и имеют составное пуассоновское распределение с параметрами  $\lambda_1 = 40$ ,  $\lambda_2 = 60$  соответственно. Распределение индивидуальных выплат по двум портфелям приведено в таблице

Величина выплат для первого портфеля	Вероятность для первого портфеля	Величина выплат для второго портфеля	Вероятность для второго портфеля
20	0,25	40	0,3
40	0,5	60	0,4
60	0,25	80	0,3

Найти

- а) распределение индивидуальных выплат объединенного портфеля;
- б) ожидаемое значение индивидуальных выплат объединенного портфеля;

*Решение.*

- а) Пусть  $\eta_1$ ,  $\eta_2$  случайные величины выплат по двум портфелям соответственно. Тогда

$$\begin{aligned} \mathbf{P}(\eta_{\text{объед}} = k) &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \mathbf{P}(\eta_1 = k) + \frac{\lambda_2}{\lambda_1 + \lambda_2} \mathbf{P}(\eta_2 = k) = \\ &= \frac{2}{5} \mathbf{P}(\eta_1 = k) + \frac{3}{5} \mathbf{P}(\eta_2 = k). \end{aligned}$$

Отсюда получаем распределение индивидуальных выплат объединенного портфеля:

$\eta_{\text{объед}}$	20	40	60	80
вероятность	0,1	0,38	0,34	0,18

б)

$$\mathbf{E}(\eta_{\text{объед}}) = 20 \cdot 0,1 + 40 \cdot 0,38 + 60 \cdot 0,34 + 80 \cdot 0,18 = 52.$$

*Задание 27.* В рамках коллективной модели риска известно, что число страховых случаев в течение года имеет пуассоновское распределение со средним значением 2, распределение потерь при одном страховом случае приведено в таблице

Величина потерь	Вероятность
3	0,4
6	0,2
12	0,3
18	0,1

Найти вероятность того, что суммарная величина выплат страховщика в течение года равна а) 0, б) 12.

*Решение.* Пусть  $\nu$  – число страховых случаев по портфелю за заданный период времени. Тогда

$$\mathbf{P}(\nu = k) = \frac{2^k}{k!} e^{-2}, k \geq 0.$$

а)

$$\mathbf{P}(S_\nu = 0) = \mathbf{P}(\nu = 0) = e^{-2} = 0,135335.$$

б) Варианты для  $S_\nu = 12$

$\nu = 1$  и  $\eta_1 = 12$ ;

$\nu = 2$  и  $\eta_1 = \eta_2 = 6$ ;

$\nu = 3$  и  $\eta_1 = \eta_2 = 3, \eta_3 = 6$  либо  $\eta_1 = \eta_3 = 3, \eta_2 = 6$  либо  $\eta_2 = \eta_3 = 3, \eta_1 = 6$ ;

$\nu = 4$  и  $\eta_1 = \eta_2 = \eta_3 = \eta_4 = 3$ .

Отсюда

$$\begin{aligned} \mathbf{P}(S_\nu = 12) &= \mathbf{P}(\nu = 1)\mathbf{P}(\eta = 12) + \mathbf{P}(\nu = 2)\mathbf{P}(\eta = 6)^2 + \\ &+ 3\mathbf{P}(\nu = 3)\mathbf{P}(\eta = 3)^2\mathbf{P}(\eta = 6) + \mathbf{P}(\nu = 4)\mathbf{P}(\eta = 3)^4 = \\ &= 2e^{-2} \cdot 0,3 + \frac{2^2}{2} e^{-2} \cdot 0,2^2 + 3 \cdot \frac{2^3}{6} e^{-2} \cdot 0,4^2 \cdot 0,2 + \frac{2^4}{4!} e^{-2} \cdot 0,4^4 = \end{aligned}$$

$$= e^{-2} \left( 0,6 + 0,08 + 0,128 + \frac{2}{3} \cdot 0,4^4 \right) = 0,111665142.$$

*Задание 28.* Для некоторого портфеля договоров страхования данные о выплатах приведены в таблице

Диапазон выплат	Число выплат
0-2	25
2-10	10
10-100	10
100-1000	5

а) Найти ЭФР и ЭФП.

б) Пусть  $\xi$  – случайная величина выплат по портфелю. Используя эмпирическое распределение, найти оценку ожидаемого значения выплат.

в) Найти оценку ожидаемого значения выплат, если договоры предусматривают максимальную выплату 250 или вычет 250.

*Решение.* Дано

$$c_0 = 0, c_1 = 2, c_2 = 10, c_3 = 100, c_4 = 1000,$$

$$n_1 = 25, n_2 = 10, n_3 = 10, n_4 = 5, n = 50.$$

Отсюда

$$F_{50}(0) = 0,$$

$$F_{50}(2) = \frac{25}{50} = \frac{1}{2},$$

$$F_{50}(10) = \frac{25 + 10}{50} = \frac{35}{50} = \frac{7}{10},$$

$$F_{50}(100) = \frac{25 + 10 + 10}{50} = \frac{45}{50} = \frac{9}{10},$$

$$F_{50}(1000) = \frac{25 + 10 + 10 + 5}{50} = 1.$$

а) Теперь мы можем найти ЭФР  $F_{50}(x)$ :

$$F_{50}(x) = \begin{cases} \frac{x}{4}, & 0 \leq x \leq 2, \\ \frac{x}{40} + \frac{9}{20}, & 2 \leq x \leq 10, \\ \frac{x}{450} + \frac{61}{90}, & 10 \leq x \leq 100, \\ \frac{x}{9000} + \frac{8}{9}, & 100 \leq x \leq 1000, \\ \text{не определена,} & x > 1000. \end{cases}$$

Следовательно, ЭФП имеет вид

$$f_{50}(x) = \begin{cases} \frac{1}{4}, & 0 \leq x < 2, \\ \frac{1}{40}, & 2 \leq x < 10, \\ \frac{1}{450}, & 10 \leq x < 100, \\ \frac{1}{9000}, & 100 \leq x < 1000, \\ \text{не определена,} & x \geq 1000. \end{cases}$$

б) Имеем

$$\begin{aligned} \hat{\mathbf{E}}(\xi) &= \frac{1}{n} \sum_{j=1}^m \left( \frac{c_j + c_{j-1}}{2} \right) n_j = \\ &= \frac{1}{50} (1 \cdot 25 + 6 \cdot 10 + 55 \cdot 10 + 550 \cdot 5) = \frac{3385}{50} = 67,7. \end{aligned}$$

в) Для договора с лимитом 250 имеем

$$\begin{aligned} \hat{\mathbf{E}}(\xi \wedge 250) &= \int_0^{250} x f_{50}(x) dx + 250(1 - F(250)) = \\ &= \int_0^2 \frac{x}{4} dx + \int_2^{10} \frac{x}{40} dx + \int_{10}^{100} \frac{x}{450} dx + \int_{100}^{250} \frac{x}{9000} dx + 250 \left( 1 - \frac{250}{9000} - \frac{8}{9} \right) = \\ &= 0,5 + 1,2 + 11 + 2,9167 + 20,83 = 36,4467. \end{aligned}$$

Для договора с вычетом имеем

$$\hat{\mathbf{E}}(\xi - 250)_+ = \hat{\mathbf{E}}(\xi) - \hat{\mathbf{E}}(\xi \wedge 250) = 31,2533.$$

Либо

$$\hat{\mathbf{E}}(\xi - 250)_+ = \int_{250}^{1000} (x-250)f_{50}(x) dx = \frac{1}{9000} \int_{250}^{1000} (x-250) dx = 31,25.$$

*Задание 29.* Исследуется продолжительность времени необходимого адвокатам для урегулирования судебных исков, связанных с телесными травмами. Пусть  $T$  представляет собой количество месяцев с момента назначения адвоката до момента, когда дело будет урегулировано. Девять случаев наблюдались в течение периода исследования, два из которых не были урегулированы до истечения срока исследования. Для этих двух случаев время  $T$  было записано как 4 и 6 месяцев соответственно. Для остальных случаев наблюдаемые значения  $T$  таковы:

1; 3; 3; 5; 8; 8; 9

Использовать оценку Каплана-Мейера для оценки вероятности  $\mathbf{P}(3 \leq T \leq 5)$ .

*Решение.* Необходимые расчеты для оценки Каплана-Мейера приведены в следующих двух таблицах:

$j$	$d_j$	$x_j$	$u_j$
1	0	1	
2	0	3	
3	0	3	
4	0		4
5	0	5	
6	0		6
7	0	8	
8	0	8	
9	0	9	

$j$	$y_j$	$s_j$	$r_j$
-----	-------	-------	-------

1	1	1	9
2	3	2	8
3	5	1	5
4	8	2	3
5	9	1	1

Отсюда

$$S_{KM}(t) = \begin{cases} 1, & 0 \leq t < 1, \\ 1 - \frac{1}{9} = \frac{8}{9}, & 1 \leq t < 3, \\ \frac{8}{9} \left(1 - \frac{2}{8}\right) = \frac{2}{3}, & 3 \leq t < 5, \\ \frac{2}{3} \left(1 - \frac{1}{5}\right) = \frac{8}{15}, & 5 \leq t < 8, \\ \frac{8}{15} \left(1 - \frac{2}{3}\right) = \frac{8}{45}, & 8 \leq t < 9, \\ \frac{8}{45} \left(1 - \frac{1}{1}\right) = 0, & t \geq 9. \end{cases}$$

Следовательно

$$\begin{aligned} \mathbf{P}(3 \leq T \leq 5) &= \mathbf{P}(T = 3) + \mathbf{P}(3 < T \leq 5) = \\ &= \mathbf{P}(T = 3) + F(5) - F(3) = \\ &= \mathbf{P}(T = 3) + S_{KM}(3) - S_{KM}(5) = \frac{2}{9} + \frac{2}{3} - \frac{8}{15} = \frac{16}{45} = 0,3556. \end{aligned}$$

*Задание 30.* Для выборки из 10 полисов были зарегистрированы следующие выплаты

$$2; 3; 3; 5; 5^+; 6; 7; 7^+; 9; 10^+,$$

где «+» указывает на то, что выплаты превышали лимит полиса. Использовать оценку Каплана-Мейера, чтобы оценить вероятность того, что выплата превысит 8.

*Решение.* Необходимые расчеты для оценки Каплана-Мейера приведены в следующих двух таблицах:

$j$	$d_j$	$x_j$	$u_j$
-----	-------	-------	-------

1	0	2	
2	0	3	
3	0	3	
4	0	5	
5	0		5
6	0	6	
7	0	7	
8	0		7
9	0	9	
10	0		10

$j$	$y_j$	$s_j$	$r_j$
1	2	1	10
2	3	2	9
3	5	1	7
4	6	1	5
5	7	1	4
6	9	1	2

Отсюда

$$S_{KM}(t) = \begin{cases} 1, & 0 \leq t < 2, \\ 1 - \frac{1}{10} = 0,9, & 2 \leq t < 3, \\ 0,9 \left(1 - \frac{2}{9}\right) = 0,7, & 3 \leq t < 5, \\ 0,7 \left(1 - \frac{1}{7}\right) = 0,6, & 5 \leq t < 6, \\ 0,6 \left(1 - \frac{1}{5}\right) = 0,48, & 6 \leq t < 7, \\ 0,48 \left(1 - \frac{1}{4}\right) = 0,36, & 7 \leq t < 9, \\ 0,36 \left(1 - \frac{1}{2}\right) = 0,18, & t \geq 9. \end{cases}$$

Так как  $y_5 \leq 8 < y_6$ , то  $P_{KM} = S_{KM}(8) = S_{KM}(7) = 0,36$ .

*Задание 31.* В таблице представлена информация о 20 выплатах рабочим по медицинской страховке

27	82	115	126	155	161	243	294	340	384
457	680	855	877	974	1193	1340	1884	2558	15743

Методом моментов оценить параметры

а) экспоненциального распределения;

б) гамма распределения;

в) распределения Парето.

*Решение.* Найдем сначала первый и второй выборочные моменты:

$$\hat{\mu}_1 = \frac{1}{20} \cdot (27 + \dots + 15\,743) = 1424,4,$$

$$\hat{\mu}_2 = \frac{1}{20} \cdot (27^2 + \dots + 15\,743^2) = 13\,238\,441,9.$$

а) Пусть  $\xi \sim \text{Exp}(\theta)$ . Тогда

$$l = 1,$$

$$\mu_1(\theta) = \hat{\mu}_1,$$

$$\hat{\theta} = 1424,4.$$

б) Пусть  $\xi \sim G(\alpha, \theta)$ . Тогда

$$l = 2,$$

$$\alpha = \theta_1, \quad \theta = \theta_2.$$

Имеем

$$\begin{cases} \mu_1(\alpha, \theta) = \hat{\mu}_1, \\ \mu_2(\alpha, \theta) = \hat{\mu}_2. \end{cases}$$

Отсюда

$$\begin{cases} \alpha\theta = 1424,4, \\ \alpha(\alpha + 1)\theta^2 = 13238441,9. \end{cases}$$

Возведем первое уравнение в квадрат и разделим второе уравнение на первое. Получим:

$$\frac{\alpha + 1}{\alpha} = 6,52489 \Rightarrow \hat{\alpha} = \frac{1}{5,52489} = 0,18100.$$



Полученное значение подставим в первое уравнение. Получим:

$$\hat{\theta} = \frac{1424,4}{0,18100} = 7869,61.$$

в) Пусть  $\xi \sim \text{Pareto}(\alpha, \theta)$ . Тогда

$$l = 2, \\ \alpha = \theta_1, \quad \theta = \theta_2.$$

Имеем

$$\begin{cases} \mu_1(\alpha, \theta) = \hat{\mu}_1, \\ \mu_2(\alpha, \theta) = \hat{\mu}_2. \end{cases}$$

Отсюда

$$\begin{cases} \frac{\theta}{\alpha - 1} = 1424,4, \\ \frac{2\theta^2}{(\alpha - 1)(\alpha - 2)} = 13238441,9. \end{cases}$$

Возведем первое уравнение в квадрат и разделим второе уравнение на первое. Получим:

$$\frac{2(\alpha - 1)}{\alpha - 2} = 6,52489 \Rightarrow \hat{\alpha} = 2,442.$$

Полученное значение подставим в первое уравнение. Получим:

$$\hat{\theta} = 1424,4 \cdot 1,442 = 2053,985.$$

*Задание 32.* Определить оценку по методу моментов для экспоненциальной модели на основе данных задачи 31 с цензурированными наблюдениями по значению 250. Повторить задачу для равномерного распределения и распределения Парето с одним известным параметром  $\theta = 150$ .

*Решение.* В результате цензурирования данные задачи 31 предстают следующим образом:

Выплата	Вероятность( $= \frac{n_j}{n}$ )
---------	----------------------------------

27	$\frac{1}{20} = 0.05$
82	0.05
115	0.05
126	0.05
155	0.05
161	0.05
243	0.05
Более 250	$1 - 0.05 \cdot 7 = 0.65$

Тогда

$$\hat{\mu}_1 = \frac{1}{20} \cdot (27 + 82 + 115 + 126 + 155 + 161 + 243) + \frac{13}{20} \cdot 250 = 207,95,$$

$$\begin{aligned} \hat{\mu}_2 &= \frac{1}{20} \cdot (27^2 + 82^2 + 115^2 + 126^2 + 155^2 + 161^2 + 243^2) + \frac{13}{20} \cdot 250^2 \\ &= 47902,45 \end{aligned}$$

а) Пусть  $\xi \sim \text{Exp}(\theta)$ ,  $0 < l$ . Тогда

$$\mathbf{E}[(\xi \wedge 250)] = \int_0^{250} e^{-\frac{x}{\theta}} dx = \theta \left( 1 - e^{-\frac{250}{\theta}} \right).$$

Приравнивая выборочный момент и ожидаемое значение, получим (численно) оценку параметра по методу моментов:

$$\theta \left( 1 - e^{-\frac{250}{\theta}} \right) = 207,95 \Rightarrow$$

$$\hat{\theta} = 657,26.$$

б) Пусть  $\xi \sim U(0, \theta)$ ,  $0 < l < \theta$ . Тогда

$$\mathbf{E}[(\xi \wedge 250)] = \int_0^{250} (1 - F_{\xi}(x)) dx = \frac{500\theta - 250^2}{2\theta}.$$

Приравнивая выборочный момент и ожидаемое значение, получим оценку параметра по методу моментов:

$$\frac{500\theta - 250^2}{2\theta} = 207,95 \Rightarrow$$

$$\hat{\theta} = 743,1629.$$

в) Пусть  $\xi \sim \text{Pareto}(\alpha, 150)$ . Тогда при  $\alpha \neq 1$  имеем

$$\begin{aligned} \mathbf{E}[(\xi \wedge 250)] &= \int_0^{250} (1 - F_\xi(x)) dx = \int_0^{250} \left(\frac{150}{x + 150}\right)^\alpha dx = \\ &= \frac{150}{\alpha - 1} \left[1 - \left(\frac{150}{400}\right)^{\alpha-1}\right], \quad \alpha \neq 1. \end{aligned}$$

Приравнявая выборочный момент и ожидаемое значение, получим (численно) оценку параметра по методу моментов:

$$\frac{150}{\alpha - 1} \left[1 - \left(\frac{150}{400}\right)^{\alpha-1}\right] = 207,95 \Rightarrow$$

$$\hat{\alpha} = 0,330897758.$$

*Задание 33.* Использовать метод эмпирических квантилей для оценивания параметров экспоненциального распределения (квантиль уровня 0,5) и распределения Парето (квантили 0,3 и 0,8) на основе данных задачи 31.

*Решение.*

Для экспоненциального распределения  $\hat{\pi}_{0,5}$  – это обычная медиана:

$$\hat{\pi}_{0,5} = \frac{384+457}{2} = 420,5. \text{ Отсюда}$$

$$0,5 = F(420,5|\theta) = 1 - e^{-\frac{420,5}{\theta}} \Rightarrow \hat{\theta} = -\frac{420,5}{\ln 0,5} = 606,65.$$

Для распределения Парето для квантилей уровня 0,3 и 0,8 сглаженная эмпирическая оценка находится следующим образом:

$$0,3: j = [21 \cdot 0,3] = [6,3] = 6, h = 6,3 - 6 = 0,3,$$

$$\hat{\pi}_{0,3} = 0,7 \cdot 161 + 0,3 \cdot 243 = 185,6,$$

$$0,8: j = [21 \cdot 0,8] = [16,8] = 16, h = 16,8 - 16 = 0,8,$$

$$\hat{\pi}_{0,8} = 0,2 \cdot 1193 + 0,8 \cdot 1340 = 1310,6.$$

Уравнения, которые надо решить:

$$\begin{cases} 0,3 = F(185,6) = 1 - \left(\frac{\theta}{185,6 + \theta}\right)^\alpha, \\ 0,8 = F(1310,6) = 1 - \left(\frac{\theta}{1310,6 + \theta}\right)^\alpha \Rightarrow \\ \ln 0,7 = -0,356675 = \alpha \ln \left(\frac{\theta}{185,6 + \theta}\right) \\ \ln 0,2 = -1,609438 = \alpha \ln \left(\frac{\theta}{1310,6 + \theta}\right) \Rightarrow \\ \hat{\theta} = 715,03, \quad \hat{\alpha} = 1,54559. \end{cases}$$

Данные решения получены численно.

**Задание 34.** Используя данные задачи 31, определить ОМП для параметров

- а) экспоненциального распределения;
  - б) гамма-распределения при  $\alpha = 2$ ;
  - в) гамма-распределения при двух неизвестных параметрах  $\alpha$  и  $\theta$ .
- а) Для экспоненциального распределения:

$$\begin{aligned} L(\theta) &= \prod_{j=1}^n f_{\xi_j}(x_j; \theta) = \prod_{j=1}^n \frac{1}{\theta} e^{-x_j/\theta} = \frac{1}{\theta^n} e^{-\frac{n\bar{x}}{\theta}} \\ l(\theta) &= \sum_{j=1}^n (-\ln \theta - x_j \theta^{-1}) = -n \ln \theta - n\bar{x} \theta^{-1} \\ l'(\theta) &= -n\theta^{-1} + n\bar{x} \theta^{-2} = 0 \\ n\theta &= n\bar{x} \Rightarrow \hat{\theta} = \hat{\mu}_1 = \bar{x} \end{aligned}$$

Для данных задачи 31  $\hat{\theta} = \bar{x} = 1424,4$ . Значение ФП равно -165,23.

Для этого случая ОММ и ОМП совпадают.

- б) Для гамма-распределения при  $\alpha = 2$ :

$$f(x; \theta) = \frac{x^{2-1} e^{-\frac{x}{\theta}}}{\Gamma(2)\theta^2} = x\theta^{-2} e^{-\frac{x}{\theta}} \Rightarrow \ln f(x; \theta) = \ln x - 2 \ln \theta - x\theta^{-1},$$

$$l(\theta) = \sum_{j=1}^n \ln x_j - 2n \ln \theta - n\bar{x}\theta^{-1}$$

$$l'(\theta) = -2n\theta^{-1} + n\bar{x}\theta^{-2} = 0$$

$$\hat{\theta} = \frac{\bar{x}}{2} = \hat{\theta} = \frac{1424,4}{2} = 712,2$$

и значение ФП равно -179,98. Эта оценка совпадает с оценкой по методу моментов.

в) Для гамма распределения с двумя неизвестными параметрами уравнения не так просты. Имеем

$$f(x; \alpha, \theta) = \frac{x^{\alpha-1} e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^\alpha} \Rightarrow \ln f(x; \alpha, \theta) = (\alpha - 1) \ln x - x\theta^{-1} - \ln \Gamma(\alpha) - \alpha \ln \theta,$$

$$\ln f(x_j; \alpha, \theta) = (\alpha - 1) \ln x_j - x_j\theta^{-1} - \ln \Gamma(\alpha) - \alpha \ln \theta,$$

$$l(\theta) = \sum_{j=1}^n [(\alpha - 1) \ln x_j - x_j\theta^{-1} - \ln \Gamma(\alpha) - \alpha \ln \theta] =$$

$$= (\alpha - 1) \sum_{j=1}^n \ln x_j - \theta^{-1} n\bar{x} - n \ln \Gamma(\alpha) - n\alpha \ln \theta.$$

Для нахождения частной производной по  $\alpha$  нужно продифференцировать гамма-функцию. Полученное уравнение нельзя решить аналитически. С помощью Excel или Matlab оценки определяются быстро:  $\hat{\alpha} = 0,55616$ ,  $\hat{\theta} = 2561,1$ . И значение ФП равно -162,29. Эти значения уже не совпадают с ОММ.

*Задание 35.* Данные в таблице представляют собой 227 выплат по общему договору страхования гражданской ответственности.

Диапазон выплат	Число выплат
0 – 7500	99

7500 – 17500	42
17500 – 32500	29
32500 – 67500	28
67500 – 125000	17
125000 – 300000	9
Более 300000	3

На основе этих данных определить ОМП для

а) экспоненциального распределения;

б) обратного экспоненциального распределения.

*Решение.*

а) Для экспоненциального распределения

$$F(x; \theta) = 1 - e^{-\frac{x}{\theta}}, \quad x \geq 0.$$

Логарифмическая ФП имеет вид:

$$\begin{aligned}
 l(\theta) = & 99 \ln \left[ F(7500) - \underbrace{F(0)}_{=0} \right] + 42 \ln [F(17500) - F(7500)] + \\
 & + 29 \ln [F(32500) - F(17500)] + 28 \ln [F(67500) - F(32500)] + \\
 & + 17 \ln [F(125000) - F(67500)] + 9 \ln [F(300000) - F(125000)] + \\
 & + 3 \ln \left[ \underbrace{F(\infty)}_{=1} - F(300000) \right].
 \end{aligned}$$

Численные результаты дают  $\hat{\theta} = 29721$  и значение ФП, равное - 406,03.

б) Для обратного экспоненциального распределения функция распределения имеет вид:  $F(x; \theta) = e^{-\frac{\theta}{x}}$ .

Логарифмическая ФП имеет вид:

$$\begin{aligned}
 l(\theta) = & \sum_{j=0}^6 n_j \ln \left[ e^{-\frac{\theta}{c_{j+1}}} - e^{-\frac{\theta}{c_j}} \right] = \\
 = & 99 \ln \left[ e^{-\frac{\theta}{7500}} - 0 \right] + 42 \ln \left[ e^{-\frac{\theta}{17500}} - e^{-\frac{\theta}{7500}} \right] + 29 \ln \left[ e^{-\frac{\theta}{32500}} - e^{-\frac{\theta}{17500}} \right] +
 \end{aligned}$$

$$28 \ln \left[ e^{-\frac{\theta}{67500}} - e^{-\frac{\theta}{32500}} \right] + 17 \ln \left[ e^{-\frac{\theta}{125000}} - e^{-\frac{\theta}{67500}} \right] \\ + 9 \ln \left[ e^{-\frac{\theta}{300000}} - e^{-\frac{\theta}{125000}} \right] \\ + 3 \ln \left[ e^{-\frac{\theta}{300000}} - 1 \right].$$

Численные результаты дают

$$\hat{\theta} = 6662,38903999329, l(\hat{\theta}) = -365,396143287309.$$

*Задание 36.* Предположим, что данные из задачи 31 были цензурированы по значению 250. Определить ОМП для параметров

- а) экспоненциального распределения;
- б) обратного экспоненциального распределения;

Решение. Имеем

выплата	27	82	115	126	155	161	243	250
частота	1	1	1	1	1	1	1	13

Первые 7 наблюдений не цензурируются. Для них множество  $A_j$  состоит из отдельных точек, равных наблюдению  $x_j$ .

Для всех моделей

$$L(\theta) = f(27) \cdot f(82) \cdots f(243)(1 - F(250))^{13},$$

$$l(\theta) = \ln f(27) + \ln f(82) + \cdots + \ln f(243) + 13 \ln(1 - F(250)).$$

а) Экспоненциальное распределение. Имеем

$$L(\theta) = \theta^{-7} e^{-\frac{909}{\theta}} \cdot \left( e^{-\frac{250}{\theta}} \right)^{13} = \theta^{-7} e^{-\frac{4159}{\theta}},$$

$$l(\theta) = -7 \ln \theta - 4159 \theta^{-1},$$

$$l'(\theta) = -7 \theta^{-1} + 4159 \theta^{-2} = 0 \Rightarrow \hat{\theta} = \frac{4159}{7} = 594,14.$$

б) Для обратного экспоненциального распределения

Напомним вид плотности и ФР для обратного экспоненциального распределения:

$$f(x; \theta) = \frac{\theta}{x^2} \cdot \exp\left(-\frac{\theta}{x}\right) \text{ и } F(x; \theta) = \exp\left(-\frac{\theta}{x}\right).$$

Имеем

$$l(\theta) = 13 \ln\left(1 - e^{-\frac{\theta}{250}}\right) + 7 \ln \theta - \ln(27^2 \cdot \dots \cdot 243^2) - \theta \left(\frac{1}{27} + \dots + \frac{1}{243}\right) \Rightarrow$$

$$l'(\theta) = \frac{13e^{-\frac{\theta}{250}}}{250\left(1 - e^{-\frac{\theta}{250}}\right)} + \frac{7}{\theta} - \left(\frac{1}{27} + \dots + \frac{1}{243}\right) = 0.$$

Полученное уравнение решается численно:

$$\hat{\theta} = 189,7822265625, l(\hat{\theta}) = -52,9724897925807.$$

*Задание 37.* Выплаты по договору страхования следуют экспоненциальному распределению с параметром  $\theta$ . После введения вычета 3 и лимита 14 были зарегистрированы следующие выплаты

$$1; 3; 7; 9; 11; 11.$$

Найти ММП (методом максимального правдоподобия) оценку  $\theta$ .

*Решение.* Максимальная возможная выплата равна

$$u - d = 14 - 3 = 11.$$

ФП имеет вид

$$\begin{aligned} L(\theta) &= \frac{f(4) \cdot f(6) \cdot f(10) \cdot f(12)}{[1 - F(3)]^4} \left[ \frac{1 - F(14)}{1 - F(3)} \right]^2 = \\ &= \frac{1}{\theta^4} e^{-\frac{1}{\theta}(4+6+10+12+2 \cdot 14-6 \cdot 3)} = \frac{1}{\theta^4} e^{-\frac{42}{\theta}}. \end{aligned}$$

Отсюда 
$$l(\theta) = -\frac{42}{\theta} - 4 \ln \theta$$

и 
$$l'(\theta) = \frac{42}{\theta^2} - \frac{4}{\theta} = 0 \Rightarrow \hat{\theta} = 10,5.$$

*Задание 38.* Составить «медвежий спрэд» из двух трехмесячных put-опционов с ценами исполнения 350, 270. Цены этих опционов составляют соответственно 25, 10. Представить графически зависимость дохода, который приносит спрэд, от цены базового актива.

*Решение:*



Одновременно покупается опцион «пут» с высокой ценой исполнения 350 и продаётся опцион «пут» с низкой ценой исполнения 270.

Медвежий спред актуален, если ожидается, что рынок упадёт до определённого уровня. Покупатель такого спреда хочет воспользоваться преимуществами падающего рынка, однако снижает размер премии, продавая «пут», который ограничивает убыток в случае роста цены базового актива и прибыль в случае падения цены.

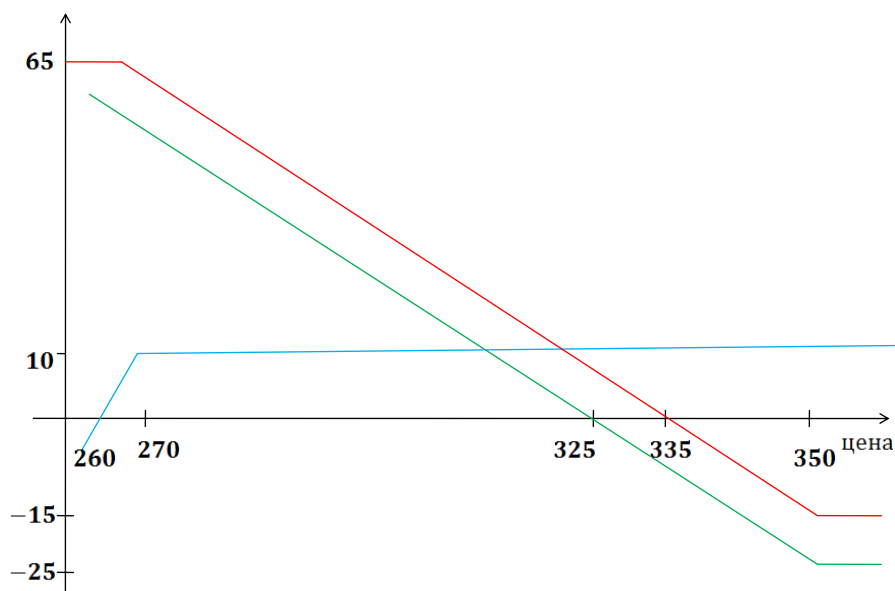
Прибыль ограничена и рассчитывается, разница между ценами исполнения опционов минус "чистая премия" (разница между премиями проданного и купленного опциона)

$$(350-270) - (25-10) = 65$$

Убыток ограничен "чистой премией" (разница между премиями проданного и купленного опциона), уплаченной за опцион.

$$10-25 = -15$$

График дохода



**Задание 39.** Составить опционную стратегию «бычий спрэд» из двух трехмесячных опционов покупателя с ценами исполнения 350, 360. Цены

этих опционов составляют соответственно 4, 2. Представить графически зависимость дохода, который приносит спред, от цены базового актива.

*Решение:*

Стратегия «бычий спред» принесет прибыль, если базовый актив повысится.

Максимальная прибыль равна разнице между ценами исполнения опционов минус максимальный риск:

$$(360-350)-(6-4)=8$$

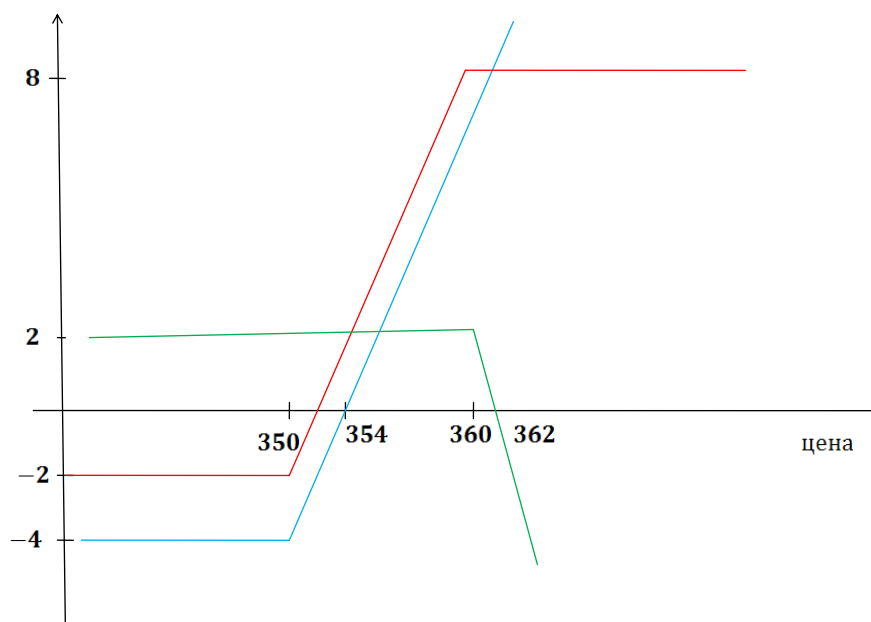
Треjder получит всю прибыль, если базовый актив поднимется выше проданного опциона.

Стратегия принесет убытки, если цена базового актива понизится, либо останется на месте.

Максимальный убыток равен премии купленного опциона минус премия проданного:

$$4-6=-2$$

График дохода:



*Задание 40.* В рамках модели Блэка-Шоулса определить цену европейского опциона колл, если курс спот акции равен 126 руб., цена

исполнения равна 138 руб., ставка без риска равна 8%, время до истечения контракта составляет 9 месяцев, мгновенное стандартное отклонение доходности акции равно 0,3

*Решение:*

Цена европейского опциона колл:

$$C = S \cdot N(d_1) - E \cdot e^{-\delta t} \cdot N(d_2)$$

$$d_1 = \frac{\ln\left(\frac{S}{E}\right) + \left(\delta + \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}} = \frac{\ln\left(\frac{126}{138}\right) + \left(0,08 + \frac{0,3^2}{2}\right) \cdot 0,75}{0,3 \cdot \sqrt{0,75}} = 0,016464$$

$$d_2 = \frac{\ln\left(\frac{S}{E}\right) + \left(\delta - \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}} = d_1 - \sigma\sqrt{t} = 0,016464 - 0,3 \cdot \sqrt{0,75} = -0,15229$$

$$N(d_1) = N(0,016464) = 0,506568$$

$$N(d_2) = N(-0,15229) = 0,439481$$

$$C = 126 \cdot 0,506568 - 138 \cdot e^{-0,3 \cdot 0,75} \cdot 0,439481 = 15,4$$

*Задание 41.* Имеется три инвестиционных инструмента с ожидаемыми доходностями  $\bar{r}_1 = 1, \bar{r}_2 = 2$ , дисперсиями  $\sigma_{11} = 4, \sigma_{22} = 9$ , и ковариацией  $\sigma_{12} = 5$  (в %). Требуется найти оптимальный состав портфеля, используя линейную свертку критериев эффективности и риска портфеля, если коэффициент риска  $\alpha = 1/6$ .

*Решение.* Постановка задачи оптимизации:

$$x_1 + 2x_2 - \frac{1}{6}(4x_1^2 + 9x_2^2 + 10x_1x_2) \rightarrow \max, x_1 + x_2 = 1.$$

Получили задачу выпуклого программирования: максимизируемая функция вогнута при линейном (вогнутом) ограничении. Поэтому необходимые условия экстремума являются и достаточными. Замена  $x_2 = 1 - x_1$  приводит к задаче без ограничений:  $2 - x_1 - \frac{1}{6}(3x_1^2 - 8x_1 + 9) \rightarrow \max$ .

Дифференцируя по  $x_1$ , и приравнявая производную к нулю, получаем уравнение  $-x_1 + \frac{1}{3} = 0$ . Откуда, с учетом  $x_1 + x_2 = 1$ , получаем оптимальный состав портфеля  $x^* = (\frac{1}{3}, \frac{2}{3})$ .

**Задание 42.** Общая сумма в 4 млн. руб. распределяется между тремя предприятиями в количествах, кратных 1 млн. руб. В результате выделения средств  $k$ -му предприятию в размере  $u$  оно дает доход  $f_k(u)$ ,  $k = 1, 2, 3, \dots$  величина которого может быть найдена из таблице «Доходы предприятий».

<b>Доходы предприятий</b>					
Доход	$u$ , млн. руб.				
	0	1	2	3	4
$f_1(u)$	0	5	9	11	12
$f_2(u)$	0	4	8	12	14
$f_3(u)$	0	7	9	10	11

Требуется распределить средства между предприятиями так, чтобы их суммарный доход был максимальным.

**Решение.** Обозначив средства, выделенные  $k$ -му предприятию ( $k = 1, 2, 3$ ), символом  $u_k$ , а сумму средств, выделенных предприятиям с номерами от 1 до  $k$ , символом  $x_k$  сформулируем исходную задачу как многошаговую:  $f(\bar{x}, \bar{u}) = \sum_{k=1}^3 f_k(u_k) \rightarrow \max$ ,  $x_k = x_{k-1} + u_k$ ,  $u_k \in [0, u - x_k] \cap N$ ,  $k = 1, 2, 3$ ,  $N$  – множество натуральных чисел,  $x_0 = 0$ ,  $x_3 = 4$ .

Для решения этой задачи применим метод динамического программирования.

**Этап I** (условная оптимизация).

Шаг 1. Найдем  $w_3(x_2) = \max_{u_3 \in [0, 4-x_2] \cap N} f_3(u_3)$ . Максимум функции  $f_3(u_3)$  достигается при максимальном допустимом значении  $u_3$ , т.е. условно-оптимальное управление на данном шаге есть  $u_3^*(x_2) = 4 - x_2$ , а условно-оптимальный результат –  $w_3(x_2) = f_3(4 - x_2)$ . Значения  $w_3(x_2)$  найденные с помощью табл. 1, представлены в таблице «Зависимость максимального дохода третьего предприятия от выделенных средств второму предприятию»:

**Зависимость максимального дохода третьего предприятия от выделенных средств второму предприятию**

$x_2$	0	1	2	3	4
-------	---	---	---	---	---

$w_3(x_2)$	11	10	9	7	0
------------	----	----	---	---	---

Шаг 2. Вычислим  $w_2(x_1) = \max_{u_2 \in [0, 4-x_1] \cap N} (f_2(u_2) + w_3(x_1 + u_2))$ . Для нахождения максимума функции  $f_2(u_2) + w_3(x_1 + u_2)$  составляем таблицу «Суммарный доход первого и второго предприятия» значений этой функции, используя данные таблице «Доходы предприятий» и таблице «Зависимость максимального дохода третьего предприятия от выделенных средств второму предприятию»:

**Суммарный доход первого и второго предприятия**

$x_1$	$u_2$				
	0	1	2	3	4
0	11	14	17	(19)	14
1	10	13	(15)	12	-
2	9	(11)	8	-	-
3	(7)	4	-	-	-
4	(0)	-	-	-	-

В таблице «Зависимость максимального дохода третьего предприятия от выделенных средств второму предприятию» скобками выделены максимальные по  $u_2$  значения функции  $f_2(u_2) + w_3(x_1 + u_2)$ , соответствующие различным значениям  $x_1$ . Используя таблицу «Зависимость максимального дохода третьего предприятия от выделенных средств второму предприятию», находим условно-оптимальное управление  $u_2^*(x_1)$  и условно-оптимальный результат  $w_2(x_1)$  на данном шаге, представив их значения в таблицах: «Зависимость максимального суммарного дохода первого и второго предприятия от выделенных средств первому предприятию» и «Зависимость условно-оптимального размера выделенных средств второму предприятию от выделенных средств первому предприятию».

**Зависимость максимального суммарного дохода первого и второго предприятия от выделенных средств первому предприятию**

$x_1$	0	1	2	3	4
$w_2(x_1)$	19	15	11	7	0

**Зависимость условно-оптимального размера выделенных средств  
второму предприятию от выделенных средств первому предприятию**

$x_1$	0	1	2	3	4
$u_2^*(x_1)$	3	2	1	0	0

Шаг 3. Так как начальное состояние  $x_0 = 0$ , то найдем только  $w_1(x_0)$  и  $u_1^*(x_0)$ . Имеем  $w_1(0) = \max_{u_1 \in [0,4] \cap N} (f_1(u_1) + w_2(0 + u_1))$ . Для определения максимума в правой части последнего равенства составим таблицу «Суммарный доход предприятий» значений функции  $f_1(u_1) + w_2(u_1)$  которые найдем с помощью таблиц «Доходы предприятий» и «Зависимость максимального суммарного дохода первого и второго предприятия от выделенных средств первому предприятию».

<b>Суммарный доход предприятий</b>					
$u_1$	0	1	2	3	4
$f_1(u_1) + w_2(u_1)$	19	(20)	(20)	19	12

Из табл. 6 видно, что  $w_1(0) = 20$ , причем  $u_1^*(0) = 1$  или  $u_1^*(0) = 2$  т.е. в данной задаче существует два оптимальных управления и две оптимальные траектории.

**Этап II** (безусловная оптимизация).

Шаг 1.

Пусть  $u_1^* = 1$ , тогда  $x_1^* = x_0 + u_1^* = 1$ . Пусть  $u_1^* = 2$ , тогда  $x_1^* = x_0 + u_1^* = 2$ .

Шаг 2.

а) Для  $x_1^* = 1$  имеем  $u_2^* = u_2^*(1) = 2$  (см. табл. 5),  $x_2^* = x_1^* + u_2^* = 3$ .

б) Для  $x_1^* = 2$  получаем  $u_2^* = u_2^*(2) = 1$  (см. табл. 5),  $x_2^* = x_1^* + u_2^* = 3$ .

Шаг 3. Так как для обеих оптимальных фазовых траекторий  $x_2^* = 3$ , то находим  $u_3^* = u_3^*(3) = 1$  и  $x_3^* = x_2^* + u_3^* = 4$ .

Окончательно получаем два оптимальных управления  $\bar{u}^* = (1,2,1)$  или  $\bar{u}^* = (2,1,1)$  и, соответственно, две оптимальных траектории  $\bar{x}^* = (0,1,3,4)$  или  $\bar{x}^* = (0,2,3,4)$ . Таким образом, существует два оптимальных варианта распределения средств предприятиям: 1) первому предприятию выделяется 1 млн.руб., второму – 2 млн.руб. и третьему – 1 млн.руб.; 2)

первому – 2, второму – 1 и третьему – 1 млн.руб. В обоих случаях суммарный доход предприятий составит  $w_1(0) = 20$  млн.руб.

*Задание 43.* Имеются два предприятия  $P$  и  $Q$ . При выделении каждому из них на год средств в количестве  $x$  единиц предприятие  $P$  обеспечивает доход  $5x$  единиц и остаток от выделенных средств  $0.3x$  единиц, а предприятие  $Q$  – доход  $4x$  единиц и остаток выделенных средств  $0.5x$  единиц. Обоим предприятиям выделено на 4 года  $a=1000$  единиц средств. Нужно распределить их в начале периода и затем остатки средств ежегодно между предприятиями, чтобы общий доход за указанный период был максимальным.

*Решение.* Разобьем весь период продолжительностью в 4 года на 4 этапа, приняв каждый год за один этап. Будем нумеровать этапы, начиная с первого года, и обозначим через  $u_k, v_k$  средства, выделяемые соответственно предприятиям  $P$  и  $Q$  на  $k$ -м этапе; сумма  $u_k + v_k = x_k$  представляет собой общее количество используемых на  $k$ -м этапе средств, оставшихся в конце предыдущего  $(k-1)$ -го этапа. Очевидно,  $x_1 = a = 1000$ . Доход, получаемый на  $k$ -м этапе, равен  $5u_k + 4v_k$ . Максимальный доход  $w_k(x_k)$ , полученный на последних этапах, начиная с  $k$ -го, т.е. с распределения  $x_k$  средств, выражается через функциональное уравнение Беллмана

$$w_k(x_k) = \max_{u_k+v_k=x_k} (5u_k + 4v_k + w_{k+1}(x_{k+1})).$$

Так как  $u_k + v_k = x_k$ , то для каждого этапа достаточно выбрать значение величины  $u_k$ , при этом  $v_k = x_k - u_k$ . Доход на  $k$ -м этапе составит  $5u_k + 4v_k = 5u_k + 4(x_k - u_k) = 4x_k + u_k$ . Количество средств, используемых на  $k$ -м этапе, выразится рекуррентной формулой  $x_k = 0.3u_{k-1} + 0.5v_{k-1} = 0.3u_{k-1} + 0.5(x_{k-1} - u_{k-1})$  или

$$x_k = 0.1(5x_{k-1} - 2u_{k-1})$$

а функциональное уравнение Беллмана примет вид:

$$w_k(x_k) = \max_{0 \leq u_k \leq x_k} (4x_k + u_k + w_{k+1}(x_{k+1})).$$

Полагая последовательно  $k = 4, 3, 2, 1$  (планирование начинается с последнего этапа), используя формулы (11) и (12) и учитывая, что  $w_5(x_5) = 0$ , получим:

$$k = 4, \quad w_4(x_4) = \max_{0 \leq u_4 \leq x_4} (4x_4 + u_4) = 5x_4, \quad \text{максимум функции } 4x_4 + u_4$$

достигается в точке  $u_4 = x_4$  – правом конце отрезка  $[0, x_4]$ ,

$$v_4 = x_4 - u_4 = 0, \quad x_4 = 0.1(5x_3 - 2u_3).$$

$$k = 3, \quad w_3(x_3) = \max_{0 \leq u_3 \leq x_3} (4x_3 + u_3 + 5x_4) =$$

$$= \max_{0 \leq u_3 \leq x_3} (4x_3 + u_3 + 5 \cdot 0.1(5x_3 - 2u_3)) = \max_{0 \leq u_3 \leq x_3} (6.5x_3 + 0 \cdot u_3) = 6.5x_3,$$

при любом значении  $u_3$ ,  $v_3 = x_3 - u_3$ ,  $x_3 = 0.1(5x_2 - 2u_2)$ .

$$k = 2, \quad w_2(x_2) = \max_{0 \leq u_2 \leq x_2} (4x_2 + u_2 + 6.5 \cdot x_3) =$$

$$= \max_{0 \leq u_2 \leq x_2} (4x_2 + u_2 + 6.5 \cdot 0.1(5x_2 - 2u_2)) =$$

$$= \max_{0 \leq u_2 \leq x_2} (7.25x_2 - 0.3u_2) = 7.25x_2,$$

максимум линейной функции  $7.25x_2 - 0.3u_2$  достигается в точке  $u_2 = 0$  – левом конце отрезка  $[0, x_2]$ ,  $v_2 = x_2 - u_2 = x_2$ ,  $x_2 = 0.1(5x_1 - 2u_1)$ .

$$k = 1, \quad w_1(x_1) = \max_{0 \leq u_1 \leq x_1} (4x_1 + u_1 + 7.25 \cdot x_2) =$$

$$= \max_{0 \leq u_1 \leq x_1} (4x_1 + u_1 + 7.25 \cdot 0.1(5x_1 - 2u_1)) =$$

$$= \max_{0 \leq u_1 \leq x_1} (7.625x_1 - 0.45u_1) = 7.625x_1,$$

максимум линейной функции  $7.625x_1 - 0.45u_1$  достигается в точке  $u_1 = 0$  – левом конце отрезка  $[0, x_1]$ ,  $v_1 = x_1 - u_1 = x_1 = a = 1000$ .

Максимальный доход за 4 года составит

$$w_1(x_1) = 7.625x_1 = 7.625 \cdot 1000 = 7625 \text{ денежных единиц.}$$

Для получения такого дохода надо в первый и второй годы все средства отдать предприятию  $Q$ :  $u_1 = 0$ ,  $v_1 = x_1 = 1000$ ,  $u_2 = 0$ ,

$v_2 = x_2 = 0.1(5x_1 - 2u_1) = 500$ ; на третий год оставшиеся средства можно распределить между предприятиями  $P$  и  $Q$  произвольным образом:

$$0 \leq u_3 \leq x_3, \quad v_3 = x_3 - u_3, \quad x_3 = 0.1(5x_2 - 2u_2) = 250;$$



и, наконец, на четвертый год все оставшиеся средства надо передать предприятию  $P$ :  $u_4 = x_4, v_4 = 0$ , остаток средств при этом будет зависеть от распределения средств в третьем году. Остаток же после четырех лет работы будет  $x_5 = 0.3x_4 = 37.5 - 0.06u_3$ . Величины  $x_4$  и  $x_5$  зависят от  $u_3$ . Максимальная прибыль за 4 года не зависит от остатка и равна 7625 единиц.

*Задание 44.* Задача управления запасами. Потребность сборочного предприятия составляет  $N$  деталей в год, причем эти детали расходуются равномерно и непрерывно. Хранение детали на складе стоит  $c_2$  руб. в сутки, а поставка одной партии деталей  $c_1$  руб. При условии отсутствия дефицита, определить наиболее экономичный объем партии  $n^0$  и интервал между поставками  $T^0$ , если  $N=120000$ ,  $c_1=10000$ ,  $c_2=0.35$ . Определить наиболее экономичный объем партии  $\tilde{n}^0$ , если отсутствие на сборке каждой детали приносит убытки в размере  $c_3=3.5$  руб.

*Решение.* Вычислим оптимальный объем партии при отсутствии дефицита по формуле Уилсона  $n^* = \sqrt{\frac{2c_1N}{c_2\theta}}$ , где  $\theta$  - рассматриваемый интервал времени:  $n^* = \sqrt{\frac{2 \cdot 10000 \cdot 120000}{0.35 \cdot 365}} = 4335$  деталей, где  $\theta=365$  дней. При этом интервал между поставками составляет  $T^* = \frac{n^* \cdot \theta}{N} = \frac{4335 \cdot 365}{120000} = 13.2$  дней. Оптимальный объем партии в условиях дефицита вычислим по формуле  $\tilde{n}^* = \frac{n^*}{\sqrt{\rho}}$ , где  $\rho = \frac{c_3}{c_2 + c_3}$  - плотность убытков:  $\rho = \frac{3.5}{0.35 + 3.5} = 0.909$ ,  $\tilde{n}^* = \frac{4335}{\sqrt{0.909}} = 4547$  деталей.

*Задание 45.* Предприятие производит два вида продукции: А и Б. Производство  $y$  единиц продукции вида Б обходится предприятию в  $y^2$  тыс. рублей, а производство  $x$  единиц продукции вида А обходится предприятию в  $x^2$  тыс. рублей. Цена единицы продукции вида А составляет 20 тыс. рублей, а продукции вида Б – 10 тыс. рублей. Методом Лагранжа найти оптимальный план  $(x, y)$  производства, максимизирующий прибыль, при условии, что затраты на производство равны 25 тыс. рублей. Издержками

считать затраты на производство. На сколько изменится прибыль при оптимальном планировании производства, если затраты увеличить на 1 тыс. рублей.

*Решение.* Задача оптимизации состоит в максимизации дохода при ограничении на затраты и естественном условии неотрицательности переменных:  $20x + 10y \rightarrow \max$ ,  $x^2 + y^2 = 25$ ,  $x, y \geq 0$ . Опустим простые ограничения ( $x, y \geq 0$ ) и найдем решение получившейся оптимизационной задачи с одним ограничением-равенством. Продифференцировав функцию Лагранжа  $L(x, y, \lambda) = 20x + 10y + \lambda(25 - x^2 - y^2)$  по всем переменным, приходим к системе уравнений:  $20 - 2x\lambda = 0$ ,  $10 - 2y\lambda = 0$ ,  $x^2 + y^2 = 25$ . Выразив из первых двух уравнений  $x$  и  $y$ :  $x = 10/\lambda$ ,  $y = 5/\lambda$ , и, подставив их в третье уравнение, найдем множитель Лагранжа  $\lambda = \pm\sqrt{5}$ . Учитывая, что переменные  $x$  и  $y$  должны быть неотрицательными, имеем  $\lambda^* = \sqrt{5}$ ,  $x^* = 2\sqrt{5}$ ,  $y^* = \sqrt{5}$ .

Окаймленная матрица Гессе имеет вид 
$$\begin{pmatrix} 0 & -2x & -2y \\ -2x & -2\lambda & 0 \\ -2y & 0 & -2\lambda \end{pmatrix}.$$

Определитель этой матрицы равен  $8\lambda(x^2 + y^2)$  и положителен в оптимальной точке  $(2\sqrt{5}, \sqrt{5})$ , следовательно,  $(x^*, y^*) = (2\sqrt{5}, \sqrt{5})$  – точка максимума.

Множители Лагранжа измеряют чувствительность оптимального значения целевой функции  $f(x^*)$  к изменениям констант  $b$  в ограничениях:  $\frac{\Delta f(x^*)}{\Delta b} \approx \lambda^*$ . По условию  $\Delta b = 1$ , значит изменение (увеличение) дохода при увеличении затрат на единицу составит  $\Delta f(x^*) \approx \sqrt{5}$ . При этом изменение прибыли  $\Delta \pi(x^*) = \Delta f(x^*) - \Delta b = \sqrt{5} - 1 \approx 2.24 - 1 = 1.24$

*Задание 46.* Функция полезности потребления имеет вид:

$$U(x_1, x_2) = (x_1 - a_1)^\alpha \cdot (x_2 - a_2)^\beta, \quad (7)$$

где  $x_1$  и  $x_2$  – спрос на первое и второе блага соответственно.

Доход потребителя –  $I$ . Вектор цен благ –  $\vec{P} = (p_1, p_2)$ .

- 1) Получите выражение для множителя Лагранжа.
- 2) Найдите функции спроса по Маршаллу.
- 3) Постройте косвенную функцию полезности.

*Решение.* Спецификация модели поведения потребителя имеет вид:

$$\begin{cases} (x_1 - a_1)^\alpha \cdot (x_2 - a_2)^\beta \rightarrow \max \\ \alpha > 0, \beta > 0, \\ 0 \leq a_1 \leq x_1, 0 \leq a_2 \leq x_2, \\ p_1 \cdot x_1 + p_2 \cdot x_2 \leq I, \\ x_1 \geq 0, x_2 \geq 0. \end{cases} \quad (8)$$

Функция Лагранжа:

$$L(x_1, x_2, \lambda) = (x_1 - a_1)^\alpha \cdot (x_2 - a_2)^\beta + \lambda(I - p_1 x_1 - p_2 x_2). \quad (9)$$

Условия существования экстремума:

$$\begin{cases} \frac{\partial L(x_1, x_2, \lambda)}{\partial x_1} = \alpha(x_1 - a_1)^{\alpha-1}(x_2 - a_2)^\beta - \lambda p_1 = 0, \\ \frac{\partial L(x_1, x_2, \lambda)}{\partial x_2} = \beta(x_1 - a_1)^\alpha(x_2 - a_2)^{\beta-1} - \lambda p_2 = 0, \\ \frac{\partial L(x_1, x_2, \lambda)}{\partial \lambda} = I - p_1 x_1 - p_2 x_2 = 0. \end{cases} \quad (10)$$

Решение системы (10):

$$\frac{\alpha(x_1 - a_1)^{\alpha-1}(x_2 - a_2)^\beta}{\beta(x_1 - a_1)^\alpha(x_2 - a_2)^{\beta-1}} = \frac{p_1}{p_2}, \quad (11.a)$$

$$\frac{\alpha}{\beta} \cdot \frac{x_2 - a_2}{x_1 - a_1} = \frac{p_1}{p_2}, \quad (11.b)$$

$$x_2 = a_2 + \frac{p_1}{p_2} \cdot \frac{\beta}{\alpha} (x_1 - a_1), \quad (11.b)$$

$$I = p_1 x_1 + p_2 \left[ a_2 + \frac{p_1}{p_2} \cdot \frac{\beta}{\alpha} (x_1 - a_1) \right], \quad (11.c)$$

$$I = p_1 x_1 + p_2 a_2 + \frac{\beta}{\alpha} \cdot p_1 (x_1 - a_1), \quad (11.d)$$

$$I - p_2 a_2 + \frac{\beta}{\alpha} \cdot p_1 a_1 = x_1 p_1 \left( 1 + \frac{\beta}{\alpha} \right), \quad (11.e)$$

$$x_1 = \frac{I - p_2 a_2 + \frac{\beta}{\alpha} \cdot p_1 a_1}{p_1 \left( 1 + \frac{\beta}{\alpha} \right)}, \quad (11.f)$$

Сделаем формулу (11. *f*) более элегантной. Для этого проведём следующие преобразования:

$$x_1 = \frac{I - p_2 a_2 - p_1 a_1 + p_1 a_1 + \frac{\beta}{\alpha} \cdot p_1 a_1}{p_1 \left(1 + \frac{\beta}{\alpha}\right)}, \quad (12. a)$$

$$x_1 = \frac{I - (p_1 a_1 + p_2 a_2) + p_1 a_1 \left(1 + \frac{\beta}{\alpha}\right)}{p_1 \left(1 + \frac{\beta}{\alpha}\right)}, \quad (12. b)$$

$$x_1^* = a_1 + \frac{I - (p_1 a_1 + p_2 a_2)}{p_1 \left(1 + \frac{\beta}{\alpha}\right)}. \quad (12. c)$$

$$x_2 = a_2 + \frac{p_1}{p_2} \cdot \frac{\beta}{\alpha} \left[ \frac{I - (p_1 a_1 + p_2 a_2)}{p_1 \left(1 + \frac{\beta}{\alpha}\right)} \right], \quad (12. d)$$

$$x_2^* = a_2 + \frac{I - (p_1 a_1 + p_2 a_2)}{p_2 \left(1 + \frac{\alpha}{\beta}\right)}, \quad (12. e)$$

Выражение для множителя Лагранжа:

$$\begin{aligned} \lambda &= \frac{\alpha \left[ \frac{I - (p_1 a_1 + p_2 a_2)}{p_1 \left(1 + \frac{\beta}{\alpha}\right)} \right]^{\alpha-1} \left[ \frac{I - (p_1 a_1 + p_2 a_2)}{p_2 \left(1 + \frac{\alpha}{\beta}\right)} \right]^{\beta}}{p_1} = \\ &= \frac{\alpha}{p_1} \cdot \frac{[I - (p_1 a_1 + p_2 a_2)]^{\alpha+\beta-1}}{\left[ p_1 \left(1 + \frac{\beta}{\alpha}\right) \right]^{\alpha-1} \left[ p_2 \left(1 + \frac{\alpha}{\beta}\right) \right]^{\beta}}. \end{aligned} \quad (13)$$

Максимальная полезность:

$$U^* = \left[ \frac{I - (p_1 a_1 + p_2 a_2)}{p_1 \left(1 + \frac{\beta}{\alpha}\right)} \right]^{\alpha} \cdot \left[ \frac{I - (p_1 a_1 + p_2 a_2)}{p_2 \left(1 + \frac{\alpha}{\beta}\right)} \right]^{\beta}. \quad (14. a)$$

$$U^* = \frac{[I - (p_1 a_1 + p_2 a_2)]^{\alpha+\beta}}{p_1^{\alpha} p_2^{\beta} \left(1 + \frac{\beta}{\alpha}\right)^{\alpha} \left(1 + \frac{\alpha}{\beta}\right)^{\beta}}. \quad (14. b)$$

Функции спроса потребителя по Маршаллу на первое и второе блага соответственно равны:

$$D_1(p_1, p_2, I) = a_1 + \frac{I - (p_1 a_1 + p_2 a_2)}{p_1 \left(1 + \frac{\beta}{\alpha}\right)}, \quad (15. a)$$

$$D_2(p_1, p_2, I) = a_2 + \frac{I - (p_1 a_1 + p_2 a_2)}{p_2 \left(1 + \frac{\alpha}{\beta}\right)}. \quad (15. b)$$

Косвенная функция полезности примет вид:

$$U(p_1, p_2, I) = \left(\frac{\alpha}{\alpha + \beta} \cdot \frac{I}{p_1}\right)^\alpha \cdot \left(\frac{\beta}{\alpha + \beta} \cdot \frac{I}{p_2}\right)^\beta \quad (16)$$

*Задание 47.* Предприниматель намерен взять в аренду отель сроком на 1 год. Имеются отели четырех типов: на 20, 30, 40 или 50 комнат. По условию аренды предприниматель должен оплатить все расходы, связанные с содержанием отеля. Эти расходы (в условных денежных единицах, уде.) состоят из трех частей:

1. Расходы, не зависящие от выбора проекта отеля:

- 1а) благоустройство территории – 10000 уде.;
- 1б) затраты на текущий ремонт и содержание – 1500 уде.;
- 1с) один ночной дежурный – 6000 уде.;
- 1д) один служащий для уборки территории – 8000 уде.

Всего – 25500 уде.

2. Расходы, пропорциональные числу комнат отеля:

- 2а) меблировка одной комнаты – 4000 уде.;
- 2б) 1 горничная на 10 комнат – 6000 уде.;
- 2с) содержание одной комнаты - 150 уде.;
- 2д) страхование на случай пожара для одной комнаты - 25 уде.

Всего на комнату – 4775 уде.

3. Расходы, пропорциональные среднему числу занятых комнат:

- 3а) стирка, уборка - 5 уде. в день.;
- 3б) электричество, газ, вода - 5 уде. день.

Всего на занятую комнату – 10 уде. в день.

Доход предпринимателя составляет 60 у.д.е. в день с каждой занятой комнаты.

Среднее число заявок на комнату в год (т.е. среднегодовой спрос) предположительно равно 20, 30, 40 или 50.

Математически формализовать данную ситуацию, построить выигрыш-функцию прибыли предпринимателя, подсчитать выигрыши и сформировать платежную матрицу игры. Какую прибыль недополучил предприниматель, если он арендовал отель на 40 комнат, а среднегодовой спрос на комнаты в этот период оказался максимальным?

*Решение.* Построим теоретико-игровую модель данной ситуации. Сознательным игроком  $A$  является предприниматель, обладающий четырьмя чистыми стратегиями:  $A_i$  - арендовать отель с  $A_i = 10(i+1)$  комнатами,  $i = 1, 2, 3, 4$ . Природой  $\Pi$  является среднегодовой спрос с состояниями  $\Pi_j$  - среднее число заявок на комнату в год равно  $\Pi_j = 10(j+1)$ ,  $j = 1, 2, 3, 4$ . Построим выигрыш-функцию прибыли предпринимателя.

Пусть предприниматель выбрал стратегию  $A_i$ , а природа находится в состоянии  $\Pi_j$ . Тогда расходы первой части составят 25500 уде., второй части -  $4775 A_i = 4775 \cdot 10(i+1) = 47750 i + 47750$  уде., третьей части (с учетом, что в году 365 дней) составят  $10 \cdot 365 \Pi_j = 3650 \cdot 10(j+1) = 36500 j + 36500$  уде. Доход  $D_{ij}$  предпринимателя будет равен

$$D_{ij} = \begin{cases} 60 \cdot 365 \Pi_j = 21900 \cdot 10(j+1) = 219000 j + 219000 \text{ уде, если } j < i, \\ 60 \cdot 365 \Pi_i = 21900 \cdot 10(i+1) = 219000 i + 219000 \text{ уде, если } j \geq i. \end{cases}$$

Следовательно, выигрыш-функция прибыли предпринимателя будет иметь вид

$$F(i, j) = F(A_i, \Pi_j) = \begin{cases} 219000 j + 219000 - 25500 - 47750 i - 47750 - 36500 j - 36500 = \\ = 182500 j - 47750 i + 109250, \text{ если } j < i, \\ 182500 i - 47750 i + 109250 = 134750 i + 109250, \text{ если } j \geq i. \end{cases} \quad (17)$$

Подсчитав выигрыши по значениям построенной выигрыш-функции, сформируем матрицу выигрышей:

$P_j \backslash A_i$	$P_1$	$P_2$	$P_3$	$P_4$
$A_1$	$a_{11} = 244000$	$a_{12} = 244000$	$a_{13} = 244000$	$a_{14} = 244000$
$A_2$	$a_{21} = 196250$	$a_{22} = 378750$	$a_{23} = 378750$	$a_{24} = 378750$
$A_3$	$a_{31} = 148500$	$a_{32} = 331000$	$a_{33} = 513500$	$a_{34} = 513500$
$A_4$	$a_{41} = 100750$	$a_{42} = 283250$	$a_{43} = 465750$	$a_{44} = 648250$

Например, элемент  $a_{23}$  подсчитывается следующим образом:  $j = 3 > 2 = i$ , поэтому подсчет проводим по нижней строчке равенства (17):  $a_{23} = 134750 \cdot 2 + 109250 = 378750$ . В игровой ситуации  $(A_3, P_4)$  доход предпринимателя составит 513500 уде. Но так как показатель благоприятности состояния  $P_4$  равен 648250 уде., то недополученная прибыль  $A$  будет равна величине риска:  $648250 - 513500 = 134750$  уед.

**Задание 48.** Рассмотрим как антагонистическую ситуацию, участниками которой являются, с одной стороны, государственная налоговая инспекция, а с другой стороны, конкретный налогоплательщик с годовым доходом 180 тыс. рублей.

У государственной налоговой инспекции два возможных способа действия. Один из них состоит в контролировании дохода налогоплательщика и взимании с него:

- налога в размере 13%, если налогоплательщик заявил свой действительный доход в размере 180 тыс. рублей;

- налога в размере 13% от 180 тыс. рублей и штрафа в размере 10% от незаявленной налогоплательщиком суммы, если налогоплательщик заявил в декларации доход меньше 180 тыс. рублей, в частности, скрыл свой доход вовсе.

Другой способ действия налоговой инспекции - вообще не контролировать доход налогоплательщика, полагаясь на его честность.

Налогоплательщик при декларировании своего дохода использует одну из следующих трех стратегий поведения: 1) заявить о действительном доходе в 180 тыс. рублей; 2) заявить доход в 90 тыс. рублей; 3) скрыть доход.

Какая из двух указанных стратегий налоговой инспекции гарантирует взимание с налогоплательщика налога, не меньшего 23 400 рублей, при любой из трех отмеченных стратегий налогоплательщика?

Какая из трех отмеченных стратегий налогоплательщика гарантирует уплату налога не больше 23 400 рублей?

*Решение.* Построим платежную матрицу игры для игрока  $A$  - государственной налоговой инспекции:

$B_j$	$B_1$	$B_2$	$B_3$	$\alpha_i$
$A_i$	*			
* $A_1$	23 400	32 400	41 400	23 400
$A_2$	23 400	11 700	0	0
$\beta_j$	23 400	32 400	41 400	<div style="text-align: right;">23400</div> <div style="text-align: left;">23400</div>

В ней, например, выигрыш  $\alpha_{12}$  имеет место при игровой ситуации  $(A_1, B_2)$ , т.е. когда игрок  $A$  придерживается чистой стратегии  $A_1$ , а игрок  $B$  – налогоплательщик - придерживается своей чистой стратегии  $B_2$ . Это означает, что  $A$  контролирует доход налогоплательщика, а налогоплательщик заявил доход меньше действительного. В этом случае налогоплательщик должен заплатить налог в размере  $180\,000 \cdot 0,13 = 23\,400$  (руб.) с действительного дохода, а также штраф  $90\,000 \cdot 0,1 = 9\,000$  (руб.), так как его заявленный доход не соответствует действительному. Таким образом,  $\alpha_{12} = 23400 + 9000 = 32400$ .



Из полученной матрицы находим нижнюю цену игры  $\alpha = 23400$  и верхнюю цену игры  $\beta = 23400$ . При этом стратегия  $A_1$  игрока  $A$  является максиминной (т. к.  $\alpha_1 = \alpha$ ), а стратегия  $B_1$  игрока  $B$  является минимаксной (т. к.  $\beta_1 = \beta$ ). Придерживаясь максиминной стратегии  $A_1$ , игрок  $A$  гарантирует себе наименьший выигрыш, т.е. государственная налоговая инспекция гарантирует себе налог, не меньший 23 400, если она будет контролировать доход налогоплательщика. Если игрок  $B$  будет придерживаться стратегии  $B_1$ , то он гарантирует себе максимальный проигрыш - для налогоплательщика сумма уплаченных с доходов налогов будет не больше 23 400 руб., если он будет честно декларировать свои доходы.

Можно отметить, что данная задача имеет решение в чистых стратегиях, т.к.  $\alpha = \beta$ , то элемент платежной матрицы  $a_{12} = 23400$  является седловой точкой игры и ценой игры, а названные максиминная и минимаксные стратегии игроков являются для них оптимальными.

*Задание 49.* Министерство желает построить один из двух объектов на территории города. Городские власти могут принять предложения министерства или отказать. Свои действия (стратегии) они применяют независимо друг от друга, и результаты определяются прибылью согласно следующим матрицам:

$$A = \begin{pmatrix} -10 & 2 \\ 1 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & -2 \\ 2 & -1 \end{pmatrix}.$$

Элементы платежных матриц вычисляются следующим образом: например, если игроки применяют свои первые стратегии, министерство решает строить 1-ый объект, а городские власти разрешают его постройку, тогда город получает выигрыш 5 млн, а министерство теряет 10 млн, и т.д.

Существует ли в этих случаях оптимальная по Нэшу ситуация для города? Определите доход/убыток города от строительства.

*Решение.* Министерство – игрок  $A$  – имеет две стратегии: строить объект 1, строить объект 2. Город – игрок  $B$  – имеет две стратегии: принять предложение министерства или отказать. Сводная матрица результатов будет иметь вид:

$$\begin{pmatrix} (-10, 5) & (2, -2) \\ \boxed{(1, 2)} & (-1, -1) \end{pmatrix}.$$

По принципу доминирования определяется оптимальная ситуация по Нэшу  $(A_2, B_1)$  в чистых стратегиях. Доход города составит 2 млн, если он даст разрешение на строительство и министерство будет строить 2-й объект. Прибыль министерства составит 1 млн.

**Перечень рекомендуемой литературы для подготовки к  
государственному экзамену по вопросам общепрофессиональных и  
профессиональных дисциплин**

**а) Основная литература:**

1. Математика в экономике. Часть 1. Линейная алгебра, аналитическая геометрия и линейное программирование.: Учебник для студ. экономич. спец. вузов / А.С.Солодовников, В.А.Бабайцев, А.В.Браилов, И.Г.Шандра. — 3-е изд., перераб. и доп. — М. : Финансы и статистика; Инфра-М, 2011.
2. Математика в экономике. Ч.2: Математический анализ: Учебник для студ. экономич. спец. вузов / А.С.Солодовников, В.А.Бабайцев, А.В.Браилов, И.Г.Шандра.— 3-е изд., перераб. и доп. — М. : Финансы и статистика; Инфра-М, 2011.
3. Математика в экономике: Учебник для студ. экономич. спец. вузов. Ч. 3. Теория вероятностей и математическая статистика / А.С. Солодовников, В.А. Бабайцев, А.В. Браилов. — М.: Финансы и статистика, 2008.
4. Алгебра и геометрия 1: Учебное пособие по дисциплине «Алгебра и геометрия» для студентов, обучающихся по направлению подготовки 01.03.02 «Прикладная математика и информатика», профиль «Анализ данных и принятие решений в экономике и финансах» (программа подготовки бакалавра). – М.: Финансовый университет, департамент анализа данных, принятия решений и финансовых технологий, 2018. – 60 с.
5. Соловьев В. И. Анализ данных в экономике. Теория вероятностей, прикладная статистика, обработка и визуализация данных в Microsoft Excel: учебник. Москва: КНОРУС, 2018. 500 с. (Бакалавриат).
6. Гисин В.Б. Дискретная математика. Учебник и практикум. — М. «Юрайт», 2016. — 383 с.
7. Бабешко Л.О. Эконометрика и эконометрическое моделирование: учебник / Л.О. Бабешко, М.Г. Бич, И.В. Орлова. — М.: Вузовский учебник: ИНФРА-М, 2017. — 400 с.

8. Золотова Т.В. Методы принятия управленческих решений (для бакалавров). Учебник: учебник / Т.В. Золотова. — М.: КноРус, 2019. — Режим доступа: <https://www.book.ru/book/929989>

9. Лабскер Л.Г., Яценко Н.А. Теория игр в экономике, финансах и бизнесе: учебник / Л.Г. Лабскер, Н.А. Яценко. — Москва: КноРус, 2017. — Режим доступа: <https://www.book.ru/book/920478>

***Дополнительная литература:***

10. Бывшев В.А. Эконометрика. — М.: Финансы и статистика, 2008.

11. Волкова Е.С., Орел Е.Н. Дополнительные главы математического анализа: учебное пособие/ под ред. В.Б. Гисина и Е.Н. Орла. — М.; Финуниверситет, 2013.

12. Гисин В.Б., Кацыло П.И., Маевский Е.В. Функциональный анализ: учебное пособие. — М.: Финансовый университет, 2014. — 220 с.

13. Гисин В.Б., Зададаев С.А., Орел О.Е. Дискретная математика. Руководство к решению задач: учебное пособие. — М.: Финансовый университет, 2012. — 136 с.

14. Бабешко, Л.О. Основы эконометрического моделирования: учебное пособие / Л.О. Бабешко. — М.: Ленанд, 2015.

15. Математические методы в экономике и финансах: учебник / коллектив авторов; под ред. В.М. Гончаренко и В.Ю. Попова. — М.: КНОРУС, 2016. — 602 с. — (Бакалавриат).

16. Агальцов, В.П. Базы данных: в 2-х кн.: учеб. Кн.1. Локальные базы данных / В.П. Агальцов .— 2-е изд., перераб. — М.: ИД «ФОРУМ»: ИНФРА-М, 2012

17. Агальцов, В.П. Базы данных: в 2-х кн.: учеб. Кн.2. Распределенные и удаленные базы данных / В.П. Агальцов .— М. : ИД «ФОРУМ»: ИНФРА-М, 2011.

18. Денежкина И.Е., Посашков С.А., Шандра И.Г. Дифференциальные уравнения. Курс лекций. М.: ФА при Правительстве РФ, 2007

19. Кондрашов Ю.Н. Эффективное использование СУБД MS SQL Server M.: RuSciIence 2018, 7,5 п.л.

20. Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 01.03.02 Прикладная математика и информатика (уровень бакалавриата). Утвержден приказом Минобрнауки России от 12.03.2015 № 228.

***Ресурсы информационно-телекоммуникационной сети «Интернет»***

1. Электронная библиотека Финансового университета (ЭБ)  
<http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf> )
2. Электронно-библиотечная система издательства «ЮРАЙТ»  
<https://www.biblio-online.ru/>
3. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
4. Электронно-библиотечная система Znanium <http://www.znaniy.com>
5. Официальный сайт Министерства финансов Российской Федерации  
[www.minfin.ru](http://www.minfin.ru).
6. Официальный сайт Федеральной службы государственной статистики Российской Федерации [www.gks.ru](http://www.gks.ru)
7. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>

**Перечень рекомендуемой литературы для подготовки к государственному экзамену по вопросам дисциплин профиля**

***Основная литература***

1. Методы оптимальных решений в экономике и финансах: учебник / И.А. Александрова [и др.]; под ред. В.М. Гончаренко, В.Ю. Попова. Москва: Кнорус, 2014. – 400 с. – [Электронный ресурс]. – 2016. – Режим доступа: <https://www.book.ru/book/920375>
2. Бауэр Дж. и др. Актуарная математика. М.: Янус, 2002.
3. Фалин Г.И. Математические основы страхования жизни и пенсионных схем. М.: Анкил, 2002.

4. Аль-Натор М.С. Основы финансовых вычислений (факты, формулы, примеры, задачи и тесты). Ч. 1-4: учебное пособие / М.С. Аль-Натор [и др.]. — Москва: Финуниверситет, 2012.

5. Количественные методы инвестиционного анализа: Учебное пособие для студ. бакалавриата / Н.И. Лахметкина [и др.]; Финуниверситет. — М. : Финуниверситет, 2012.

6. Бабайцев В.А., Гисин В.Б. Математические методы финансового анализа. Учебное пособие. — М.: Юрайт, 2018. — 215 с.

7. Рябов П.Е. Математические методы финансового анализа. Безарбитражный подход к определению цены. Курс лекций. — М.: Финансовый университет, 2016. — 108 с.

#### ***Дополнительная литература***

8. Соловьев В.И. Финансовая математика: учебное пособие. М.: КНОРУС, 2016. — 176 с.

9. Соловьев В. И. Анализ данных в экономике. Теория вероятностей, прикладная статистика, обработка и визуализация данных в Microsoft Excel: учебник. Москва: КНОРУС, 2018. 500 с. (Бакалавриат).

#### ***Ресурсы информационно-телекоммуникационной сети «Интернет»***

1. Электронная библиотека Финансового университета (ЭБ)  
<http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf> )

2. Электронно-библиотечная система издательства «ЮРАЙТ»  
<https://www.biblio-online.ru/>

3. Электронно-библиотечная система BOOK.RU <http://www.book.ru>

4. Электронно-библиотечная система Znanium  
<http://www.znanium.com>

5. Официальный сайт Министерства финансов Российской Федерации [www.minfin.ru](http://www.minfin.ru).

6. Официальный сайт Федеральной службы государственной статистики Российской Федерации [www.gks.ru](http://www.gks.ru)

7. Электронно-библиотечная система «Университетская библиотека  
ОНЛАЙН» <http://biblioclub.ru/>