ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ РФ
ФГОУ ВПО «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Е. В. Зандер, В. П. Злодеев,
Л. И. Мошкович, А. Р. Семёнова

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ В ЭКОНОМИКЕ

Учебное пособие

СФУ 2007

УДК 517.8

ББК 65в641

З-27

Рецензенты:

Зандер Е. В.

З-27 Исследование операций в экономике: учеб. пособие / Е. В. Зандер, В. П. Злодеев, Л. И. Мошкович, А. Р. Семёнова. — Сибирский федеральный ун-т. Красноярск: 2007. 202 с.

ISBN

Пособие представляет собой конспект лекций, который излагает основные разделы исследования операций в экономике, изучаемые студентами экономических специальностей вузов и необходимые для принятия управленческих решений в экономических ситуациях с использованием математических методов. Изложены базовые вопросы исследования операций в экономике: методология моделирования, использование линейных моделей в операционном анализе экономических систем, теория двойственности в анализе и принятии решений, моделирование нелинейности в экономических процессах, теория принятия решений в условиях неопределенности и риска, а также специальные задачи исследования операций (теория массового обслуживания, целочисленные задачи, модели транспортного типа и др.).

ISBN

ОГЛАВЛЕНИЕ

РАЗДЕЛ 1. ИССЛЕДОВАНИЕ ОПЕРАЦИЙ: ЛИНЕЙНЫЕ МОДЕЛИ В ЭКОНОМИКЕ

ТЕМА 1.1. ЛИНЕЙНЫЕ МОДЕЛИ В ОПЕРАЦИОННОМ АНАЛИЗЕ ЭКОНОМИЧЕСКИХ СИСТЕМ

Лекция 1.1.1. Основы методологии моделирования

Использование математических методов для анализа экономических ситуаций предполагает предварительное описание изучаемой экономической системы математическими соотношениями, другими словами, создание математической модели этой системы.

Большинство экономических систем представляются сложными системами. Они характеризуются большим числом параметров, меняющихся во времени. Сложность анализа экономических систем в том, что существенной их составляющей являются люди, принимающие решения на основе разнообразной информации с учетом различных целей. Кроме того, экономическая система непрерывно подвергается множеству случайных, трудно прогнозируемых возмущений как извне (изменение количества и номенклатуры поставок, изменение спроса и т. д.), так и изнутри (появление новых технологий, поломка оборудования, несовпадение реальных сроков с планируемыми и т. д.).

В то же время экономико-математическая модель должна быть достаточно простой и обозримой, чтобы ее можно было записать, получить необходимые для записи данные, а если модель расчетная, то поместить полученную информацию в память ЭВМ. Математическая модель всегда «беднее» реальной моделируемой системы, всегда описывает систему лишь приблизительно, выделяя одни свойства системы и пренебрегая другими. Выбор важнейших свойств для учета в модели представляет собой искусство моделирования и определяется умением исследователя операций выделить (исходя из целей анализа) главное из большого количества факторов, цифр и соображений, затрудняющих решение задачи. Исследователю операций не говорят, какие данные необходимо собрать для решения задачи, а также где и как их найти. Часто сбор необходимых данных требует изучения таких факторов, которые на первый взгляд не имеют отношения к предложенной задаче.

Модель создается тогда, когда, по убеждению исследователя, существует некая аналогия между реальной системой и его представлениями. При отсутствии достаточной информации об исследуемых объектах или процессах их изучение начинается с построения простейших моделей. Модели в исследовании операций служат для объединения опытных факторов и нахождения взаимосвязи между параметрами.

Конечная цель разработки математической модели — прогноз результатов проведения операции и выработка рекомендаций по возможным воздействиям на ход ее проведения. При этом необходим четкий анализ сферы и границ применимости полученных результатов. Обычно модели создаются на основе теоретических положений или гипотез, объясняющих полученные в результате наблюдений данные. Модели не следует считать неизменными, их нужно трактовать как инструмент анализа и понимания экономического процесса или системы. Для изучения одного и того же объекта или процесса может быть предложено несколько моделей, соответствующих целям анализа. При создании модели, обосновывающей возможное принятие решения о развитии конкретной ситуации, важное значение имеет количество выделяемого времени. Если требуется срочно выдать результаты, то создание сложной модели нецелесообразно. Во многих случаях модель нужна лишь для грубой оценки влияния отдельных факторов на результат. Подробное теоретическое исследование здесь также нецелесообразно.

Модель должна по возможности обладать «внутренней гибкостью», т. е. допускать возможность использования непредусмотренной новой информации. Сбор и обработка исходной информации являются важной частью исследования операции. К сожалению, при решении большинства практических задач не удается собрать нужный объем данных, необходимый для создания достаточно подробной модели процесса или объекта, охватывающей все аспекты задачи.

При рассмотрении экономико-математической модели оперируют следующими основными понятиями: «критерий оптимальности», «целевая функция», «система ограничений», «уравнения связи», «решение модели».

Целевую функцию нельзя отождествлять с критерием оптимальности: различные критерии могут описываться одной и той же целевой функцией.

Целевая функция математически связывает факторы модели (прежде всего переменные, «управляющие»).

Система ограничений определяет границы, сужающие область осуществляемых, приемлемых или допустимых решений, и фиксирует основные свойства моделируемого объекта или процесса. Ограничения устанавливают границы изменения параметров — характеристик объекта.

Уравнения (соотношения) связи *являются математической формализацией системы ограничений.*

Под ***критерием оптимальности*** *обычно понимают экономический показатель, содержащий в формализованном виде конкретную цель управления объектом или процессом и выражаемый математически при помощи целевой функции через факторы модели.*

Таким образом, критерий оптимальности выражает содержательное, смысловое значение целевой функции. Иногда в качестве критерия оптимальности может выступать одна из интересующих исследователя конечных характеристик объекта.

Решением экономико-математической модели *обычно называют набор переменных, удовлетворяющий уравнениям связи.* Решения, имеющие экономический смысл, называются допустимыми. Модели, имеющие более одного решения, называются вариантными. Среди допустимых решений вариантной модели находится решение, при котором целевая функция в зависимости от смысла модели имеет наибольшее или наименьшее значение. Такое решение (как и соответствующее значение целевой функции) называется оптимальным.

Применение экономико-математической модели, в особенности оптимизационной, при исследовании операций предполагает не только построение модели, но и получение решения при помощи подходящего метода. Поэтому иногда под моделированием (в более узком смысле) понимают этап нахождения решения. Выбор метода решения зависит от математической формы модели.

С точки зрения целей операционного анализа математические модели операций в экономике делятся на оптимизационные (нормативные) и дескриптивные (описательные или экономико-математические модели прямого счета). Важнейший признак дескриптивной модели — отсутствие критерия оптимальности. Решение модели сводится к вычислению выходных характеристик объекта по заданным начальным условиям. Примерами дескриптивных моделей могут быть модели расчета объемов производства по видам продукции, увязки планов производства с ресурсами (балансовые модели), а также модели оперативного учета, получения форм отчетности, анализа влияния факторов, прогнозирования показателей и ряд других. Модели этого типа позволяют решать задачи анализа, предлагающие ответ на вопрос: «Что будет, если…?».

Когда перед исследованием операции стоят задачи синтеза, требующие ответа на вопрос: «При каких значениях управляющих параметров будет достигнуто…?», необходима оптимизационная модель, характерной чертой которой является наличие одной (однокритериальная модель) или нескольких (многокритериальная) целевых функций. В общем виде однокритериальная модель может быть описана следующими соотношениями:

$$K = f\left(x_1,\ldots,x_j,\ldots,x_n;\alpha_1,\alpha_2,\ldots,\alpha_l\right) \to extr,$$

$$g_i = g_i\left(x_1,\ldots,x_j,\ldots,x_n;\alpha_1,\alpha_2,\ldots,\alpha_l\right) \begin{Bmatrix} \le \\ \ge \\ = \end{Bmatrix} 0\,,$$

где K — критерий оптимальности;

f — целевая функция — формализованное описание критерия оптимальности;

g_i — уравнения связи, $i = \overline{1,m}$;

x_j — управляемые переменные, $j = \overline{1,n}$;

α_k — неуправляемые факторы модели, $k = \overline{1,l}$.

Решение модели заключается в нахождении множества значений управляемых переменных

$$x^{*\text{т}} = (x_1^{*},\ldots,\, x_j^{*},\ldots,\, x_n^{*})\,,$$

обращающих в max или min целевую функцию и удовлетворяющих уравнениям связи.

Учет специфики конкретных операций определяет разнообразные оптимизационные экономико-математические модели. Однако для некоторых часто повторяющихся ситуаций разработаны общие методы построения моделей. Исторически и содержательно наибольший интерес представляют линейные модели, т. е. те модели, в которых целевая функция и уравнения связи являются линейными функциями.

Сфера применимости и предпосылки построения линейных моделей в экономике

Широкое распространение в экономике линейных моделей (или, как еще говорят, линейного программирования) связано, во-первых, с тем, что многие экономические операции с достаточной точностью могут быть описаны линейными моделями; во-вторых, за небольшим исключением только для этих моделей разработаны эффективные методы численного решения при приемлемой для практики размерности задачи.

Примерами использования линейных моделей могут служить модели комплексного использования сырья, транспортная, размещения производства, перспективного планирования, развития экономического комплекса, различных ситуаций оперативного управления и т. д. Широко применяются такие модели и в связи с необходимостью анализа ситуаций, возникающих в экономике из-за большого числа возможных вариантов функционирования конкретного объекта при использовании различных видов сырья, материалов, технологий и других факторов.

Использование линейных моделей опирается на возможность рассмотрения плана (вектора управляемых переменных) в расчлененной форме, составленного из элементарных процессов, которые могут протекать с различной краткостью (интенсивностью). Также предполагается, что приращение критерия оптимальности и невязок условий задачи пропорциональны изменениям соответствующих управляемых переменных. Что, в частности, означает: увеличение выпуска продукции в некоторое число раз требует увеличения потребления объектом всех других продуктов в то же самое число раз.

Для многих задач планирования и управления такие допущения выглядят вполне приемлемыми и позволяют получать хорошие результаты.

Вместе с тем необходимо четко представлять границы линейности показателей, включаемых в модель операции. Известно, что существует так называемая условно постоянная часть расходов, не зависящая от количества выпускаемой продукции в определённом диапазоне изменения объемов. При интенсификации процессов могут меняться нормы расходов на единицу выпускаемой продукции тех или иных ингредиентов. Эти и ряд других моментов приводят к нарушению линейности объектов, выделяя тем не менее достаточно обширную сферу применимости линейных моделей в операционном анализе экономических ситуаций.

Лекция 1.1.2. Основные типы линейных моделей в операционном анализе экономики. Постановка задачи и основные определения ЗЛП

В данном разделе приводятся наиболее распространенные типы линейных моделей, отражающие различные экономические ситуации. При моделировании этих ситуаций предполагается, что справедлива гипотеза о линейной зависимости описываемых показателей.

Задача оптимального планирования производства

Задача возникает при составлении планов выпуска продукции предприятием и, значит, имеет важное практическое значение.

Постановка задачи. Пусть номенклатура выпускаемой продукции состоит из n наименований. Обозначим через a_{ij} затраты i-го вида ресурсов (i = 1,2,..., m) на производство единицы продукции j-го вида (j =1,2,…, n), через $в_i$ — полные объемы имеющихся ресурсов (i =1,2,…, m), c_j — доход от реализации единицы продукта (j =1,2,…, n).

Задача состоит в том, чтобы определить объем производства каждого продукта, который позволит при наличных ресурсах получить максимальный общий доход. Сбыт всей выпущенной продукции обеспечен.

Математическая модель задачи состоит в нахождении такого n-мерного вектора выпуска продукции $X = \{x_j\}$, (где $x_j \geq 0$ при $j = \overline{1,n}$), чтобы выполнить неравенства

$$\sum_{j=1}^{n} a_{ij}\, x_j \leq в_i; \quad x_j \geq 0; \quad i = 1,2,...,m; \;\; j = 1,2,...n$$

и при этом достичь $\max \sum_{j=1}^{n} c_j x_j$, т. е. общая прибыль от производства и реализации продукции должна быть максимальной.

Задача о диете

Пусть нам известно содержание необходимых для откорма животных питательных веществ в различных применяемых кормах, а также цена единицы каждого вида корма. Требуется выбрать рацион-набор и количество кормов так, чтобы каждое питательное вещество содержалось в нем в необходимом количестве, а суммарные расходы на этот рацион были минимальны.

Математическая модель. Введем обозначения:

m — число различных питательных веществ;

n — число видов кормов;

a_{ij} — количество единиц i-го питательного вещества, содержащегося в единице j-го вида кормов;

$в_i$ — минимальная суточная потребность в i-м питательном веществе;

c_j — стоимость единицы j-го вида корма;

x_j — количество единиц j-го вида корма, используемого в рационе.

Необходимо найти x_j, $(j=1,\dots,n)$, удовлетворяющие следующим ограничениям:

$x_j \geq 0$ (количество какого-либо корма, содержащегося в рационе, не может быть отрицательным) $j = 1,\dots, n$;

$\sum\limits_{j=1}^{n} a_{ij} x_j \geq в_i$, $(i = 1,\dots,m)$; (общее количество i-го питательного вещества в данном рационе должно быть не ниже заданного), и минимизирующие суммарные затраты на составление оптимального рациона, т. е. найти $\min \sum\limits_{j=1}^{n} c_j x_j$.

Задача о раскрое

На предприятии из листового материала стандартной формы получают заготовки необходимых размеров. При этом остатки материала идут в отходы. Количество отходов зависит от принятых вариантов раскроя. Каждый вариант характеризуется количеством заготовок различного вида, выкраиваемых из листа.

Задача — составить оптимальный план раскроя, т. е. определить, сколько листов скроить по каждому из вариантов, чтобы получить необходимое количество заготовок каждого типа при минимальных суммарных отходах.

Пусть различных заготовок вида i (i= 1,…, m) требуется в количестве $в_i$, и имеется n вариантов раскроя листа, при этом количество заготовок j-го варианта раскроя равно a_{ij} , а отходы равны c_j (j=1,…, n).

Математическая модель. Обозначим количество листов, раскраиваемых по варианту, через x_j , $x_j \geq 0$, (j=1,…, n). Тогда задача сводится к нахождению x_j, удовлетворяющих ограничениям

$$\sum_{j=1}^{n} a_{ij}x_j \geq в_i, \ i = 1,...,m.$$

Суммарные отходы должны быть минимальными $\min \sum_{j=1}^{n} c_j x_j$.

Транспортная задача

Пусть имеется m пунктов с объемами производства, равными a_i (i=1,…,m), и n пунктов потребления с объемами потребления $в_j$ (j=1,…,n). Известны величины c_{ij} — затраты по перевозке единицы продукта из i-го пункта производства в j-й пункт потребления; если $\sum_{i=1}^{m} a_i \geq \sum_{j=1}^{n} в_i$ (т. е. потребление не превышает возможности производства), то задача сводится к нахождению такого плана перевозок, при котором были бы удовлетворены потребности во всех n пунктах потребления, а суммарные затраты на перевозку были бы минимальны.

Если по условию транспортной задачи потребление равно производству, т. е. выполняется $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} в_i$, *то задача называется* ***закрытого*** *типа* *(или задачей с правильным балансом), иначе —* ***открытого*** *(с неправильным балансом).*

Математическая модель. Обозначим через x_{ij} искомое количество продукта, перевозимое из i-го пункта производства в j-й пункт потребления (план перевозки). Требуется найти такой план перевозки $\{x_{ij}\}$, (i= 1,...,m; j=1,...,n), чтобы суммарные затраты на транспортировку были минимальны, т. е. $\min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$, при условиях:

1) $\sum_{i=1}^{m} x_{ij} \geq в_j$, j=1,..., n (в каждый пункт потребления завозится не больше требуемого количества продуктов);
2) $\sum_{j=1}^{n} x_{ij} \leq a_i$, i =1,..., n (из каждого пункта производства вывозится не более произведенного количества продукта);
3) $x_{ij} \geq 0$, i= 1,…,m, j=1,…, n.

Задача о назначениях

Пусть имеется n видов станков различных типов, которые требуется распределить между n видами работ. Известен ожидаемый эффект c_{ij} от использования i-го вида оборудования на j-м виде работ, измеряемый, например, количеством обрабатываемых деталей. Задача состоит в таком назначении станков на виды работ (по одному станку на каждый вид работы), чтобы суммарный эффект от использования всех станков был максимален (например, число обрабатываемых деталей).

Математическая модель. Введем переменные x_{ij} , определяемые формулой

$$x_{ij} = \begin{cases} 1, \text{ если } i\text{-й станок предназначен для } j\text{-го вида работ;} \\ 0 \text{ — в противном случае.} \end{cases}$$

Необходимо найти такие значения переменных x_{ij}, которые будут максимизировать суммарный эффект, т. е. $\max \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$ при условиях:

1) $\sum_{i=1}^{n} x_{ij} = 1$ (для каждого вида работы предназначен только один станок);

2) $\sum_{j=1}^{n} x_{ij} = 1$ (каждый станок предназначается только на одну работу).

К рассмотренным выше основным типам задач линейного программирования можно свести многие их разновидности.

Постановка задачи и основные определения ЗЛП

В общем виде задачу линейного программирования (ЗЛП) можно записать следующим образом: найти

$$\max f\,(x_1,\ldots,x_n) = \max \mathrm{c}^{\mathrm{T}} x = \max \sum_{j=1}^{n} c_j x_j = \max\left(c_1 x_1 + c_2 x_2 + \cdots + c_n x_n\right)$$

или

$$\min f\,(x_1,\ldots,x_n) = \min \mathrm{c}^{\mathrm{T}} x = \min \sum_{j=1}^{n} c_j x_j = \min\left(c_1 x_1 + c_2 x_2 + \cdots + c_n x_n\right)$$

при ограничениях $Ax = \sum_{j=1}^{n} a_{ij} x_j \begin{pmatrix} \le \\ = \\ \ge \end{pmatrix} b_i \quad i = \overline{1,m}, \quad x_j \ge 0 \quad j = \overline{1,n_1},$

где $n_1 \le n$.

Здесь

$x^{\mathrm{T}} = (x_1, x_2, \ldots, x_n)$ — n-мерный вектор переменных (план задачи),

$c^{\mathrm{T}} = (c_1, c_2, \ldots, c_n)$ — n-мерный вектор коэффициентов целевой функции,

$A = \left\{ a_{ij} \right\}_{n \times m}$ — матрица коэффициентов левой части системы

ограничений,

$b^{\mathrm{T}} = (b_1, b_2, \ldots, b_m)$ — вектор коэффициентов правой части системы ограничений.

Формы записи общей ЗЛП

Различают две основные формы записи ЗЛП: стандартная и каноническая. Будем рассматривать в качестве стандартной ЗЛП

$$\max \sum_{j=1}^{n} c_j x_j \text{, при ограничениях } \sum_{j=1}^{n} a_{ij} x_j \le b_i,\ i = \overline{1,m},\ x_j \ge 0,\ j = \overline{1,n},$$

каноническая ЗЛП записывается так:

$$\max \sum_{j=1}^{n} c_j x_j \text{, при ограничениях } \sum_{j=1}^{n} a_{ij} x_j = b_i, i = \overline{1,m},\ x_j \ge 0,\ j = \overline{1,n}.$$

От задачи на максимум можно всегда осуществить переход к задаче на минимум и наоборот. Если целевая функция сводится к нахождению

$$\min f = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n,$$

то можно перейти к задаче нахождения

$$\max f_1 = -f = -c_1 x_1 - c_2 x_2 - \ldots - c_n x_n,$$

поскольку

$$\min f = -\max(-f).$$

Переход от общей ЗЛП к стандартной осуществляется следующим образом.

1. Ограничение-неравенство ЗЛП, имеющее вид «≥», можно преобразовать к неравенству вида «≤», умножив обе части исходного неравенства на «–1».
2. Ограничение-равенство

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i$$

можно записать в виде двух неравенств:

$$\begin{cases} a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \le b_i, \\ -a_{i1}x_1 - a_{i2}x_2 - \ldots - a_{in}x_n \le -b_i. \end{cases}$$

3. Если переменная x_k не удовлетворяет условию неотрицательности, т. е.:

   а) $x_k \le 0$, то заменим ее на переменную $y_k \ge 0$ такую, что $x_k = -y_k$;

   б) x_k — не определена, то заменим ее на разность двух переменных $u_k \ge 0$ и $v_k \ge 0$, приняв $x_k = u_k - v_k$.

Переход от стандартной ЗЛП к канонической осуществляется путем преобразования ограничений — неравенства вида «≤» к строгим равенствам добавлением к левой части неравенства дополнительной переменной, т. е. $a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \le b_i$ преобразуется в неравенство

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n + x_{n+1} = b_{i,} \ (x_{n+1} \ge 0).$$

Число вводимых дополнительных неотрицательных переменных при преобразовании ограничений-неравенств в ограничения-равенства равно числу преобразуемых неравенств.

Вводимые дополнительные переменные имеют вполне определенный экономический смысл. Так, если в ограничениях исходной стандартной ЗЛП отражается расход и наличие производственных ресурсов, то числовое значение дополнительной переменной в плане задачи, записанной в канонической форме, равно объему используемого соответствующего ресурса. Обратный переход от канонической к стандартной форме записи ЗЛП проводится по правилам перехода от общей ЗЛП к стандартной.

Данные формы записи ЗЛП и правила перехода от одной формы к другой имеют практическое значение при решении ЗЛП различными методами, рассматриваемыми ниже. Далее введем основные определения, связанные с решением ЗЛП.

Всякое неотрицательное решение системы ограничений задачи линейного программирования называется ***планом****.*

Будем называть любой план задачи линейного программирования ***допустимым,*** *если он удовлетворяет условиям ограничений задачи.*

Допустимый план будем называть ***опорным,*** *если в нем отличны от нуля не более m+n–1 компонент, а остальные — равны нулю.*

Если в опорном плане число отличных от нуля компонент равно в точности n+m–1, то план является ***невырожденным,*** *иначе —* ***вырожденным.***

Допустимый план, при котором целевая функция задачи линейного программирования принимает свое минимальное значение, называется ***оптимальным планом.***

Лекция 1.1.3. Графический метод нахождения решения линейных моделей. Геометрическая интерпретация решения. Метод прямого перебора

Графический метод решения ЗЛП

Когда ЗЛП содержит всего две переменные, нетрудно получить ее геометрическую интерпретацию и решить задачу графически. Случай двух переменных не имеет особого практического значения, однако его рассмотрение делает ясными отдельные свойства задачи ЛП, а также геометрический смысл методов ее решения.

Найдем решение задачи, состоящей в определении максимального значения функции

$$f = c_1 x_1 + c_2 x_2 \rightarrow \max \tag{1.1}$$

при условиях

$$a_{i1}x_1 + a_{i2}x_2 \le b_i\ ,\ i = \overline{1,m}\,, \tag{1.2}$$

$$x_j \ge 0\ ,\ \ j = \overline{1,2}. \tag{1.3}$$

Каждое из неравенств системы ограничений (1.2) геометрически определяет полуплоскость, граница которой задается прямой $a_{i1}\, x_1 + a_{i2}\, x_2 = b_i$, $(i = \overline{1,m})$. Чтобы найти, какую именно из двух полуплоскостей определяет данное неравенство $a_{i1}x_1 + a_{i2}x_2 \le b_i \quad (i = \overline{1,m})$, достаточно подставить в него координаты любой точки, не лежащей на граничной прямой. Если при этом неравенство будет выполняться, то искомая полуплоскость та, в которой лежит взятая точка, а если нет, то противоположная ей.

В том случае, если система неравенств (1.2), (1.3) совместна, область ее решений есть множество точек, принадлежащих всем указанным полуплоскостям. *Так как множество точек пересечения данных полуплоскостей — выпуклое, то областью допустимых решений задачи (1.1)—(1.3) является выпуклое множество, которое называется* ***многоугольником решений.***

ЗЛП состоит в нахождении такой точки многоугольника решений, где целевая функция принимает максимальное решение. Эта точка существует тогда, когда многоугольник решений не пуст и на нем целевая функция ограничена сверху (для задачи на максимум). Тогда необходимо построить **линию уровня** $c_1 x_1 + c_2 x_2 = h$ (где h — произвольная константа), затем эта линия передвигается в направлении вектора $\vec{c}^T = (c_1; c_2)$.

*Вектор $\vec{c}^T = (c_1; c_2)$ называется **градиентом** целевой функции или вектором роста функции, и показывает направление ее роста (соответственно, вектор $\vec{c}^T = (-c_1; -c_2)$ показывает направление уменьшения значений целевой функции).* Перемещение осуществляется до тех пор, пока линия не пройдет через последнюю ее общую точку с многоугольником решений. Координаты указанной точки и определяют оптимальный план данной задачи.

При нахождении решения задачи ЛП графически могут встретиться случаи, когда:

а) целевая функция принимает максимальное значение в единственной точке допустимого множества (единственное решение);

б) целевая функция принимает максимальное значение в любой точке отрезка (если линия уровня совпадает с ограничением, тогда имеем случай бесконечного множества решений);

в) целевая функция может быть не ограничена на множестве допустимых решений (т. е. $f_{\max} = +\infty$ — это также случай бесконечного множества решений);

г) решения задачи (1.1)—(1.3) не существует, поскольку система ограничений этой задачи несовместна.

Нахождение минимального значения линейной функции при данной системе ограничений отличается от нахождения ее максимального значения при тех же ограничениях лишь тем, что линия уровня c1 x1 +c2 x2 = h передвигается не в направлении вектора $\vec{c}^T$ =(c1; c2), а в противоположном.

Алгоритм графического метода решения ЗЛП включает в себя, таким образом, следующие этапы.

1. Построение прямых, уравнения которых получаются путем замены в ограничениях (1.2), (1.3) знаков на равенства.
2. Нахождение полуплоскостей, определяемых каждым ограничением задачи.

3. Определение многоугольника решений.
4. Построение вектора $\vec{c}^T = (c_1; c_2)$.
5. Построение прямой $c_1 x_1 + c_2 x_2 = h$.
6. Перемещение прямой $c_1 x_1 + c_2 x_2 = h$ в направлении роста вектора $\vec{c}$, в результате этого либо определяется точка (точки), в которой целевая функция принимает максимальное значение, либо устанавливается неограниченность функции сверху на множестве допустимых решений.
7. Определение координат точки максимума функции и значения в этой точке.

Пример. *Для производства двух видов изделий A и B предприятие использует сырье двух видов. Нормы расхода сырья каждого вида на изготовление единицы продукции данного вида, прибыль от реализации одного изделия каждого вида и общее количество сырья каждого вида приведены в табл. 1.1.*

Таблица 1.1

Вид сырья	Нормы расхода сырья на одно изделие, кг		Общее количество сырья, кг
	A	*B*	
I	1	1	30
II	1	4	84
Прибыль от реализации одного изделия, руб.	3	4	—

Решение. Построим экономико-математическую модель задачи. Пусть x_1 — количество изделий типа A, x_2 — количество изделий типа B. Тогда прибыль от реализации можно записать как $3x_1 + 4x_2 \to \max$. Ограничения по использованию сырья будут выглядеть следующим образом:

$$\begin{cases} x_1 + x_2 \le 30, \\ x_1 + 4x_2 \le 84; \end{cases}$$

$x_1 \geq 0$, $x_2 \geq 0$, так как объем производства не может быть отрицательным.

Строим график (рис.1.1).

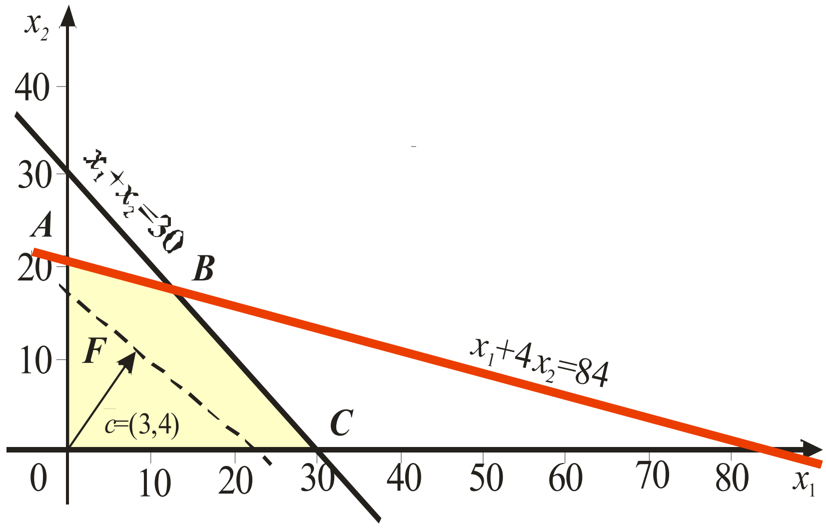


Рис. 1.1. Графическое решение к примеру

1. Прямая $x_1 + x_2 = 30$ проходит через точки (0; 30) и (30; 0). Прямая $x_1 + 4x_2 = 84$ проходит через точки (0; 21) и (84; 0).
2. Полуплоскости, определяемые неравенствами задачи, находят путем подстановки произвольной точки плоскости, например точки (0; 0) в неравенства. В данной задаче точка (0; 0) удовлетворяет обоим неравенствам, следовательно, полуплоскости лежат в той стороне от граничных прямых, где находится точка (0; 0). Ограничения $x_1 \geq 0$, $x_2 \geq 0$ определяют первый координатный угол в декартовой системе координат.
3. Многоугольником допустимых решений является четырехугольник $OABC$. Строим вектор $\vec{c}^{\,т}$=(3; 4) или совпадающий с ним по направлению $\vec{c}^{\,т}$= (30; 40).
4. Строим прямую $3x_1 + 4x_2 = 12$, которая проходит через точки (0; 3) и (4; 0).
5. Передвигаем прямую F в направлении вектора $\vec{c}$. Последней общей точкой с многоугольником решений является точка B, значит, это и есть точка максимума.

6. Определяем координаты точки *B*. Для этого необходимо решить систему из двух уравнений

$$\begin{cases} x_1 + x_2 = 30, \\ x_1 + 4x_2 = 84. \end{cases}$$

Решением системы является точка $x_1^* = 12$, $x_2^* = 18$. Следовательно, если предприятие изготовит 12 изделий типа А и 18 изделий типа В, то получит прибыль, равную $3 \cdot 12 + 4 \cdot 18 = 108$ (руб.).

Метод прямого перебора

Данный метод решения применим, когда ЗЛП приведена к канонической форме и число ограничений *m* меньше числа переменных *n*.

Если n — количество переменных, m — количество ограничений в задаче, то количество базисных переменных (не равных нулю) будет равно m, а небазисных (свободные переменные, которые приравниваются нулю) — n–m. В этом случае число возможных вариантов базисных наборов для данной задачи определяется по формуле $C_n^m = \dfrac{n!}{m!(n-m)!}$. Далее строится таблица решения, в которой для каждого базисного набора вычисляют значения базисных переменных из ограничений задачи, остальные переменные (небазисные) принимают нулевые значения. Для каждого базисного набора переменных вычисляют значения целевой функции *f*. Базисные наборы переменных, у которых при вычислениях получаются $x_j < 0$, не принимаются к рассмотрению, что вытекает из экономического смысла задачи. Затем выбирается среди вычисленных значений функции наибольшее $f_{\max}$ (либо $f_{\min}$, в зависимости от условий задачи) и по нему восстанавливается набор переменных, который и будет оптимальным планом задачи.

Пример. *Решить задачу методом прямого перебора:*

$$f = 3x_1 + 4x_2 \to \max$$

$$\begin{cases} x_1 + x_2 \le 30, \\ x_1 + 4x_2 \le 84; \end{cases}$$

$$x_1 \ge 0, x_2 \ge 0.$$

Решение. Приведем задачу к канонической форме записи:

$$f = 3\,x_1 + 4\,x_2 \to \max$$

$$\begin{cases} x_1 + x_2 + x_3 = 30, \\ x_1 + 4x_2 + x_4 = 84; \end{cases}$$

$$x_1\,,\, x_2\,,\, x_3\,,\, x_4 \geq 0.$$

Отсюда $n = 4$, $m = 2$, $n - m = 2$. Количество возможных базисных наборов равно $C_4^2 = \dfrac{4!}{2!\;2!} = 6$. Составим табл. 1.2 для решения.

Таблица 1.2

№ базисного набора	Значение переменных				$F\,(x)$
	x_1	x_2	x_3	x_4	
1	0	0	30	84	0
2	0	30	0	–36	—
3	0	21	9	0	84
4	30	0	0	54	90
5	84	0	–54	0	—
6	12	18	0	0	108

Из табл. 1.2 видно, что максимальное значение 108 целевая функция принимает в 6-м наборе, а величина прибыли $F\,(x^*) = 108$ руб. достигается при $x_1^* = 12$, $x_2^* = 18$.

Лекция 1.1.4. Симплексный метод и метод искусственного базиса для нахождения оптимального решения линейных задач исследования операций

Симплексный метод решения задачи ЛП основан на переходе от одного опорного плана к другому, при котором значение целевой функции возрастает (при условии, что данная задача имеет оптимальный план и каждый ее опорный план является невырожденным). Указанный переход возможен, если известен какой-нибудь исходный опорный план.

Для решения симплекс-методом исходная задача должна быть представлена в канонической форме записи:

$$f = c_1 x_1 + c_2 x_2 + \dots + c_n x_n \to \max; \tag{1.4}$$

$$\begin{cases} x_1 + a_{1,m+1}\, x_{m+1} + \ldots + a_{1n}\, x_n = b_1, \\ x_2 + a_{2,m+1}\, x_{m+1} + \ldots + a_{2n}\, x_n = b_2, \\ \vdots \\ x_m + a_{m,m+1}\, x_{m+1} + \ldots + a_{mn}\, x_n = b_m; \end{cases} \tag{1.5}$$

$$x_j \ge 0 \ \ (j = \overline{1,n}). \tag{1.6}$$

В данной задаче количество ограничений m меньше количества переменных n; величины a_{ij}, b_i, c_j $(i = \overline{1,m};\, j = \overline{1,n})$ заданы по условию. Задача (1.4)—(1.6) имеет исходный опорный план $x^{\text{т}}=(b_1; b_2; b_3;\ldots b_m; 0;\ldots; 0)$, где первые m компонент равны координатам вектора-столбца ограничений, а последние $n - m$ компонент нулевые.

Задачу (1.4)—(1.6) для решения симплексным методом удобно представить в векторной форме:

$$F = \sum_{j=1}^{n} c_j\, x_j \to \max \tag{1.7}$$

$$x_1 P_1 + x_2 P_2 + \ldots + x_n P_n = P_0, \tag{1.8}$$

$$x_j \ge 0 \quad (j = \overline{1,n}), \tag{1.9}$$

где

$$P_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix};\ P_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix};\ldots;\ P_m = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},\ P_{m+1} = \begin{pmatrix} a_{1,m+1} \\ a_{2,m+1} \\ \vdots \\ a_{m,m+1} \end{pmatrix};\ldots;\ P_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix};\ P_0 = \begin{pmatrix} в_1 \\ в_2 \\ \vdots \\ в_m \end{pmatrix}.$$

Исходный опорный план задачи определяется системой единичных векторов P_1, $P_2,\ldots$, P_m, которые образуют базис m-мерного пространства. Поэтому каждый из векторов P_{m+1}, P_{m+2}, P_m, P_0 может быть представлен в виде линейной комбинации векторов данного базиса.

Далее для каждого $j\ (j = \overline{1,n})$ вычисляются оценки

$$\Delta_j = z_j - c_j = \sum_{i=1}^{m} c_i\, a_{ij} - c_j\,. \tag{1.10}$$

Значения Δ_j позволяют определить, оптимален данный план или нет. Если все $\Delta_j \geq 0$, то процесс решения задачи окончен, рассматриваемое опорное решение оптимально. Оптимум целевой функции равен $F_{\max} = \sum_{i=1}^{m} c_i\, в_i$.

Если существуют оценки $\Delta_j < 0$, то среди них выбирают максимальную по модулю (наименьшую отрицательную), в дальнейшем столбец с этой оценкой будет называться направляющим и войдет в базис. Для определения вектора, выводимого из базиса, вычисляют соотношение $\lambda_i = \dfrac{в_i}{a_{ik}}$, где k — индекс направляющего столбца, из найденных значений λ_i всегда выбирается наименьшее положительное, т. е. $\min\limits_i \lambda_i = \min\limits_i \left(\dfrac{в_i}{a_{ik}} \right)$ для всех $a_{ik} > 0$. Пусть этот минимум достигается при $i = r$, тогда из базиса исключается вектор P_r, а на его место встает вектор P_k. Строка с индексом r называется направляющей, а элемент, находящийся на пересечении k-го столбца и r-ой строки (a_{rk}), называется разрешающим.

Если в столбце с элементами a_{ik} нет положительных, т.е. все $a_{ir} \le 0$, то целевая функция не ограничена сверху на множестве допустимых планов.

После выделения направляющего столбца и направляющей строки находят новый опорный план и коэффициенты разложения векторов P_j по векторам нового базиса. Компоненты нового опорного плана вычисляют (по методу Жордана — Гаусса) следующим образом:

$$в_i' = \begin{cases} в_i - (в_r / a_{rk})\, a_{ik} & \text{при } i \ne r, \\ в_r / a_{rk} & \text{при } i = r, \end{cases} \tag{1.11}$$

а коэффициенты разложения векторов по векторам нового базиса, соответствующего новому опорному плану, по формулам

$$a_{ij}' = \begin{cases} a_{ij} - (a_{rj} / a_{rk}) a_{ik} & \text{при } i \ne r, \\ a_{rj} / a_{rk} & \text{при } i = r. \end{cases} \tag{1.12}$$

После вычисления всех элементов, соответствующих новому опорному плану, вновь рассчитывают величины Δ_j для всех j и делают вывод:

а) если все $\Delta_j \ge 0$ — план оптимальный;

б) если существует хотя бы одна оценка $\Delta_j < 0$ и для каждого такого j по крайней мере одно из чисел $a_{ij} > 0$, то возможен переход к новому опорному плану, при котором значение целевой функции увеличивается;

в) если существует хотя бы одна оценка $\Delta_j < 0$, но при этом все числа $a_{ij} \le 0$, то целевая функция не ограничена.

Для удобства решения задачи симплексным методом составляют таблицу (табл. 1.2). Первый столбец таблицы («i») означает номер строки; второй («базис») — содержит вектора базиса, записанные в порядке возрастания номера единичной координаты; третий («$C_{баз}$») — коэффициенты целевой функции, соответствующие базисным векторам; четвертый («P_0») — это вектор-столбец правых частей системы, записанных в том же порядке, что и вектора базиса; последующие столбцы таблицы («P_1», …, «P_n») заполняются по матрице условий задачи.

Таблица 1.2

i	Базис	$\boldsymbol{C}_{баз}$	P_0	C_1	…	C_r	…	C_m	C_{m+1}	…	C_k	…	C_n
				P_1	…	P_r	…	P_m	P_{m+1}	…	P_k	…	P_n
1	P_1	C_1	$в_1$	1	…	0	…	0	$a_{1,m+1}$	…	a_{1k}	…	a_{1n}
2	P_2	C_2	$в_2$	0	…	0	…	0	$a_{2,m+1}$	…	a_{2k}	…	a_{2n}
⋮	⋮	⋮	⋮	⋮	…	⋮	…	⋮	⋮	…	⋮	…	⋮
r	P_{r}	C_r	$в_{\text{r}}$	0	…	1	…	0	$a_{\text{r},m+1}$	…	a_{rk}	…	a_{rn}
⋮	⋮	⋮	⋮	⋮	…	⋮	…	⋮	⋮	…	⋮	…	⋮
m	P_{m}	C_m	$в_{\text{m}}$	0	…	0	…	1	$a_{\text{m},m+1}$	…	a_{mk}	…	a_{mn}
m+1			F_0	0	…	0	…	0	Δ_{m+1}	…	Δ_k	…	Δ_n

Строка m+1 является оценочной строкой, первый ее элемент F_0 (значение целевой функции) расположен в столбце P_0 и рассчитывается по формуле $F_0 = \sum_{i=1}^{m} c_i\, в_i$. Остальные элементы строки m+1 рассчитывают по формуле (1.10). После пересчета элементов нового плана заполнение таблицы продолжают до тех пор, пока все Δ_j не станут неотрицательными.

Алгоритм симплекс-метода в виде блок-схемы может быть представлен следующим образом (рис. 1.2).

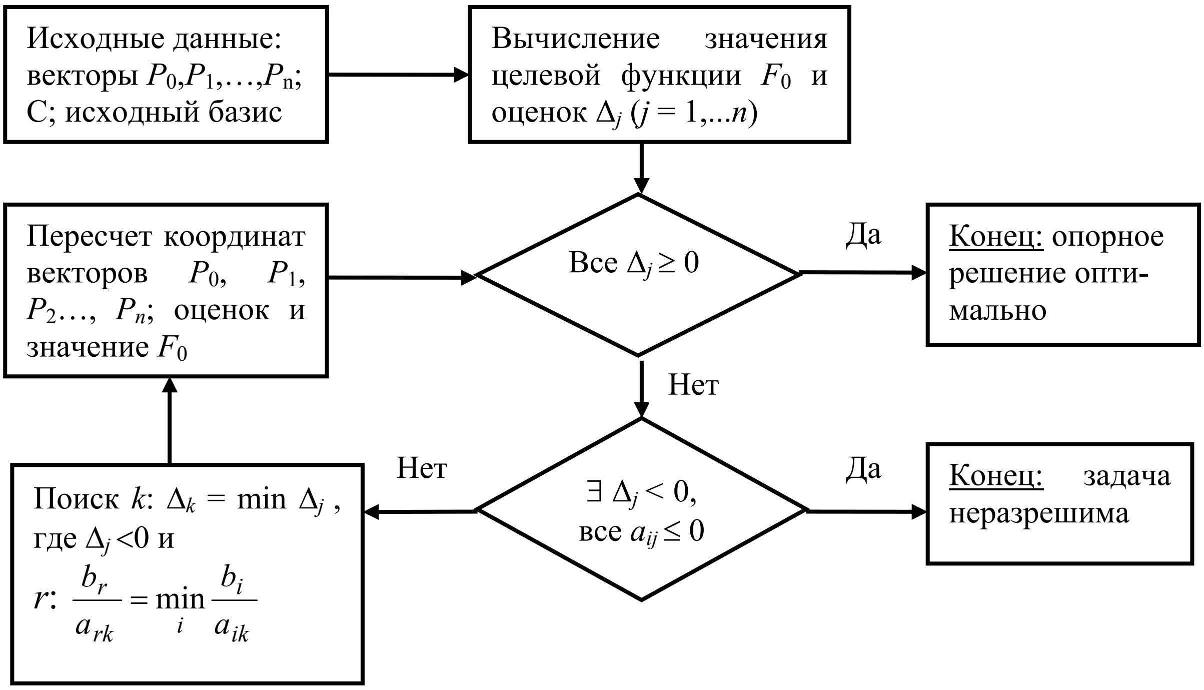


Рис. 1.2. Блок-схема алгоритма симплекс-метода

Пример. *Пусть экономико-математическая модель задачи записана:*

$$F = 3\,x_1 + 4\,x_2 \to max;$$

$$\begin{cases} x_1 + x_2 \leq 30, \\ x_1 + 4x_2 \leq 84; \end{cases}$$

$$x_1 \geq 0,\ x_2 \geq 0.$$

Найти решение задачи симплекс-методом.

Решение. Чтобы воспользоваться алгоритмом симплекс-метода, задачу следует привести к канонической форме, т. е.:

$$F = 3\,x_1 + 4\,x_2 \to \max;$$

$$\begin{cases} x_1 + x_2 + x_3 = 30, \\ x_1 + 4x_2 + x_4 = 84; \end{cases}$$

$$x_j \geq 0,\ j = \overline{1,4}.$$

Запишем векторную форму задачи:

$$x_1 P_1 + x_2 P_2 + \ldots + x_n P_n = P_0,$$

где

$$P_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix};\ P_2 = \begin{pmatrix} 1 \\ 4 \end{pmatrix};\ P_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix};\ P_4 = \begin{pmatrix} 0 \\ 1 \end{pmatrix};\ P_0 = \begin{pmatrix} 30 \\ 84 \end{pmatrix}.$$

Система имеет два ограничения и четыре переменные, в базис исходного опорного плана войдут два вектора — это единичные векторы P_3 и P_4. Составим симплекс-таблицу (табл. 1.3).

Таблица 1.3

i	Базис	$C_{баз}$	P_0	3	4	0	0	λ_j
				P_1	P_2	P_3	P_4	
1	P_3	0	30	30	1	1	0	30
2	P_4	0	84	84	4	0	1	**21**
3	Δ_j	—	0	0	**–4**	0	0	—
1	P_3	0	9	3/4	0	1	–1/4	**12**
2	P_2	4	21	1/4	1	0	1/4	84
3	Δ_j	—	84	**–2**	0	0	1	—
1	P_1	3	12	1	0	4/3	–1/3	—
2	P_2	4	18	0	1	–2/3	1/3	—
3	Δ_j	—	108	0	0	4	2	—

Исходный опорный план имеет две отрицательных оценки: $\Delta_1 = –3$ и $\Delta_2 = –4$. Выбираем Δ_2, отсюда $k = 2$. Ищем $\min\limits_i \frac{в_i}{a_{i2}} = \frac{в_4}{a_{22}} = \frac{в_4}{4} = \frac{84}{4} = 21$ ($a_{i2} > 0$), отсюда $r = 2$. Разрешающий элемент таблицы $a_{rk} = a_{22} = 4$.

Осуществляем пересчет табл. 1.3 по формулам (1.11)—(1.12), получаем новый план, где базисными являются вектора P_3 и P_2. Значение целевой функции F_0 увеличивается с 0 до 84. Оценочная строка табл. 1.3 рассчитывается по формуле (1.10).

В новом опорном плане присутствует оценка $\Delta_1 = -2$, значит, этот план оптимальным не будет.

Поскольку отрицательная оценка одна, то $k = 1$, а r выбираем: $\min\limits_i \frac{в_i}{a_{i1}} = \frac{в_1}{a_{11}} = \frac{9}{3/4} = 12$ $(a_{i1} > 0)$. Отсюда $r = 1$, разрешающий элемент $a_{11} = 3/4$. Вновь пересчитываем таблицу по формулам (1.11)—(1.12) для нового базиса P_1, P_2. Данный опорный план имеет все $\Delta_j \geq 0$, следовательно, он оптимальный. Выпишем теперь ответ из последней симплекс-таблицы. Так как в базисе содержатся два элемента P_1 и P_2, то соответствующие им значения x_1 и x_2 находим на пересечении строки P_1 (P_2) и столбца P_0, т. е. $x_1 = 12$ $x_2 = 18$. Значение функции находится на пересечения строки Δ_j и столбца P_0, т. е. $F^*_{\max} = 108$.

Ответ: $x^* = (12; 18)$, $F^*_{\max} = 108$.

Метод искусственного базиса

Предыдущий алгоритм симплекс-метода рассматривался для случаев, когда в системе присутствует полный единичный базис, которому соответствует допустимое (а, значит, и опорное) решение. Однако так бывает далеко не всегда. Для того чтобы использовать симплекс-метод для решения задачи в этих случаях, приходится прибегать к приему, который называется введением искусственного базиса. Суть его заключается в том, что в ограничения задачи искусственно вводят несколько новых переменных с таким расчетом, чтобы полученная новая система уравнений-ограничений уже имела полный единичный базис, которому соответствует опорное решение новой системы ограничений. Затем решают задачу ЛП с новыми ограничениями и со специально построенной целевой функцией (в дальнейшем будем называть предложенную задачу расширенной по отношению к исходной).

Пусть требуется найти максимум функции

$$F = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \qquad (1.13)$$

при условиях:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = в_1, \\ \qquad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = в_m; \end{cases} \tag{1.14}$$

$$x_j \geq 0 \ (j = \overline{1,n}), \tag{1.15}$$

где $в_i \geq 0 \ (i = \overline{1,m})$, $m < n$ и среди векторов

$$P_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}; \ P_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}; \ \ldots; \ P_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

нет m единичных.

Составим к исходной задаче расширенную: требуется определить максимальное значение функции

$$F = c_1 x_1 + c_2 x_2 + ... + c_n x_n - M x_{n+1} - \ ... \ - M x_{n+m} \tag{1.16}$$

при условии

$$\begin{cases} a_{11} \ x_1 + a_{12} \ x_2 + \ldots + a_{1n} \ x_n + x_{n+1} = в_1, \\ \qquad\vdots \\ a_{m1} \ x_1 + a_{m2} \ x_2 + \ldots + a_{mn} \ x_n + x_{n+m} = в_m, \end{cases} \tag{1.17}$$

$$x_j \geq 0 \ (j = 1, \overline{n+m}), \tag{1.18}$$

где M — некоторое достаточно большое положительное число.

Переменные $x_{n+1}, \ldots, x_{n+m}$ называются искусственными (как и векторы $P_{n+1}, \ldots, P_{n+m}$). Количество искусственных векторов может быть от 1 до m в зависимости от наличия в исходной задаче единичных векторов.

Расширенная задача имеет опорный план $x^{\mathrm{T}} = (0; 0; \ldots; 0; в_1; \ldots; в_m)$, в котором n-m нулевых элементов, поэтому ее решение может быть найдено симплекс-методом. В процессе решения искусственные переменные выводятся из базиса и, следовательно, принимают нулевые значения в оптимальном плане. Значение целевой функции расширенной задачи в оптимальном плане (т.е. при нулевых значениях искусственных переменных) совпадает со значением целевой функции исходной задачи (1.13)—(1.15).

При опорном плане $x^{\text{т}} = (0; 0; ...; 0; в_1, ..., в_m)$ расширенной задачи значение линейной формы есть $F_0 = -M\sum_{i=1}^{m} в_i$, а значения оценок

$$\Delta_j = z_j - c_j = -M\sum_{i=1}^{m} a_{ij} - c_j .$$

Таким образом, F_0 и разности $z_j - c_j$ состоят из двух частей, одна из которых зависит от M.

В процессе решения расширенной задачи составляют симплекс-таблицу, в которой после обычной (m+1)-й строки оценок, где записываются слагаемые, не содержащие M, помещают (m+2)-ю строку, где записывают коэффициенты при M.

При переходе от одного опорного плана к другому в базис вводят вектор, соответствующий наибольшему по абсолютной величине отрицательному числу (m+2)-й строки. Искусственный вектор, исключенный из базиса в результате некоторой итерации, в дальнейшем не имеет смысла вводить ни в один из последующих базисов, и преобразования столбцов этого вектора излишни. Пересчет симплекс-таблицы при переходе от одного опорного плана к другому производят по общим правилам симплексного метода по (m+2)-й строке до тех пор, пока:

а) все искусственные вектора не будут исключены из базиса. После этого определение оптимального плана продолжают по (m+1)-й строке;
б) не все искусственные вектора исключены из базиса, но при этом в (m+2)-й строке нет больше отрицательных значений.

Тогда, если элемент, стоящий в (m+2)-й строке столбца P_0, отрицателен, то задача не имеет решения; если он равен нулю, то найденный опорный план исходной задачи является вырожденным и базис содержит, по крайней мере, один из векторов искусственного базиса.

Пример. Найти решение задачи методом искусственного базиса

$$F = x_1 + 4\,x_2 + x_3 \to \max;$$

$$\begin{cases} x_1 - x_2 + x_3 = 8, \\ 2x_1 - 5x_2 - x_3 = 0; \end{cases}$$

$$x_j \geq 0,\ \ j = \overline{1,3}.$$

Решение. Запишем данную задачу в векторной форме:

$$x_1 P_1 + x_2 P_2 + x_3 P_3 = P_0,$$

где $P_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}; P_2 = \begin{pmatrix} -1 \\ -5 \end{pmatrix};\ P_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix};\ P_0 = \begin{pmatrix} 3 \\ 0 \end{pmatrix}.$

В базисе должны присутствовать два единичных вектора, однако среди векторов P_1, P_2, P_3 нет единичных. Поэтому в ограничения задачи добавляют искусственные переменные x_4 и x_5 (координаты векторов P_4 и P_5, соответственно, равны $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ и $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$).

Составляем целевую функцию расширенной задачи, куда переменные x_4 и x_5 войдут с коэффициентом ($-M$).

$$F = x_1 + 4\,x_2 + x_3 - M\,x_4 - M\,x_5 \to \max,$$

а ограничения преобразуются так:

$$\begin{cases} x_1 - x_2 + x_3 + x_4 = 3 \\ 2x_1 - 5x_2 - x_3 + x_5 = 0 \end{cases},$$

$$x_j \geq 0,\ \ j = \overline{1,5}.$$

Составим симплекс-таблицу для решения (табл. 1.4).

В первой части таблицы среди оценок (m+2)-й строки всего одна отрицательная — в столбце P_1. Значит, в следующий базис войдет вектор P_1, а вектор P_5 будет выведен (так как $\min\limits_{a_i k > 0}\left(\frac{3}{1};\frac{0}{2}\right) = 0$).

Таблица 1.4

i	Базис	$C_{баз}$	P_0	1	4	1	–М	$-M$	λ_j
				P_1	P_2	P_3	P_4	P_5	
1	P_4	–М	3	1	–1	1	1	0	—
2	P_5	–М	0	2	–5	–1	0	1	3
3	—	Δ_j	0	–1	–4	–1	0	**0**	**0**
4	—	Δ_M	–3	**–3**	**0**	0	0	0	—
1	P_4	–М	3	0	3/2	3/2	1	—	—
2	P_1	1	0	1	–5/2	–1/2	0	—	**2**
3		Δ_j	0	0	**–13/2**	–3/2	0	—	—
4		Δ_M	−3	0	**–3/2**	–3/2	0	—	—
1	P_2	4	2	0	1	1	—	—	—
2	P_1	1	5	1	0	2	—	—	—
3	—	Δ_j	13	0	0	5	—	—	—

Пересчет элементов таблицы проводят по формулам (1.11)—(1.12). Во второй части таблицы среди оценок (m+2)-й строки две одинаковых отрицательных оценки в столбцах P_2 и P_3 . Для введения в следующий базис нужно выбрать вектор P_2, так как он имеет больший коэффициент в целевой функции. Вектор P_2 заменит вектор P_4, так как $\min\limits_{a_ik>0}\left(\frac{3}{3/2};\frac{0}{-5/2}\right)=2$, поскольку $a_{ik} = a_{22} = -5/2 < 0$. В завершающей части таблицы оценка плана на оптимальность идет уже по (m+1)-й строке, так как искусственные векторы выведены. В этой строке на данном этапе нет отрицательных оценок, а значение целевой функции равно 13. Таким образом, найденный опорный план является оптимальным с координатами $x^{т*} = (5; 2; 0)$, значение $F^*_{\max} = 13$.

Решение симплекс-методом и методом искусственного базиса некоторых других типов линейных задач

В предложенных примерах решения задач были использованы линейные модели производства, для которых наиболее эффективным методом решения служит симплекс-метод. Однако он может быть применен и для решения задач транспортного типа, для которых существуют другие, более эффективные, приемы решения (например, метод потенциалов). Первоначальный опорный план транспортной задачи также может быть найден более простыми способами (например, методом северо-западного угла, методом минимального элемента и др.).

Решение задачи транспортного типа симплекс-методом (либо методом искусственного базиса) более длительно и громоздко, чем решение ее специальными методами, но, в принципе, реализуемо.

ТЕМА 1.2. ТЕОРИЯ ДВОЙСТВЕННОСТИ В ОПЕРАЦИОННОМ АНАЛИЗЕ ЭКОНОМИЧЕСКИХ СИСТЕМ

Лекция 1.2.1. Формальная теория двойственности: основные понятия, определения, теоремы. Прямая и двойственная задачи, взаимосвязь их решений, правила построения

Основные результаты формальной теории двойственности

Каждой задаче линейного программирования можно определенным образом сопоставить некоторую другую задачу (линейного программирования), называемую двойственной или сопряженной по отношению к исходной или прямой. Дадим определение двойственной задачи по отношению к общей задаче линейного программирования, состоящей в нахождении максимального значения функции $F = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$ при условиях

$$\begin{cases} a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n \le b_1, \\ a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n \le b_2, \\ \vdots \\ a_{k1} x_1 + a_{k2} x_2 + \ldots + a_{kn} x_n \le b_k, \\ a_{k+1,1} x_1 + a_{k+12} x_2 + \ldots + a_{k+1n} x_n \le b_{k+1}, \\ \vdots \\ a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n \le b_m, \end{cases}$$

$$x_j \ge 0,\ j = \overline{1,n}.$$

Задача, состоящая в нахождении минимального значения функции $F^* = b_1 y_1 + b_2 y_2 + \ldots + b_m y_m$ *при условиях*

$$\begin{cases} a_{11}\,y_1 + a_{21}\,y_2 + \ldots + a_{m1}\,y_m \ge c_1, \\ a_{12}\,y_1 + a_{22}\,y_2 + \ldots + a_{m2}\,y_m \ge c_2, \\ \vdots \\ a_{1l}\,y_1 + a_{2l}\,y_2 + \ldots + a_{ml}\,y_m \ge c_l, \\ a_{1,l+1}\,y_1 + a_{2,l+1}\,y_2 + \ldots + a_{m,l+1}\,y_m \ge c_{l+1}, \\ \vdots \\ a_{1n}\,y_1 + a_{2n}\,y_2 + \ldots + a_{mn}\,y_n \ge c_n, \end{cases},$$

$$y_i \ge 0, i = \overline{1,m},$$

*называется **двойственной** по отношению к исходной задаче.*

Тогда исходная задача и двойственная к ней образуют пару задач, называемую в линейном программировании двойственной парой.

Таким образом, взаимно двойственные задачи могут быть записаны в виде:

прямая задача	**двойственная задача**
$\max\,(c, x) = \max\, c^{\mathrm{T}}x$	$\min\,(в, y) = \min\, в^{\mathrm{T}}y$
$Ax \le в$	$A^{\mathrm{T}}y \ge c$
$x \ge 0.$	$y \ge 0.$

Сравнивая две сформулированные задачи, видим, что двойственная задача по отношению к исходной составляется согласно следующим правилам.

1. Целевая функция исходной задачи задается на максимум, а целевая функция двойственной — на минимум.
2. Матрица

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix},$$

составленная из коэффициентов при неизвестных в системе ограничений исходной задачи, и аналогичная матрица

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

в двойственной задаче получаются друг из друга транспонированием (т. е. заменой строк столбцами, а столбцов — строками).

3. Число переменных в двойственной задаче равно числу ограничений в системе исходной задачи, а число ограничений в системе двойственной задачи — числу переменных в исходной задаче.
4. Коэффициентами при неизвестных в целевой функции двойственной задачи являются свободные члены из системы ограничений исходной задачи, а правыми частями в соотношениях системы двойственной задачи — коэффициенты при неизвестных в целевой функции исходной задачи.
5. Если переменная x_j исходной задачи может принимать только лишь положительные значения, то j-е условие в системе двойственной задачи является неравенством вида «$\geq$». Если же переменная x_j может принимать как положительные, так и отрицательные значения, то j-е соотношение в системе двойственной задачи представляет собой уравнение. Аналогичные связи имеют место между ограничениями исходной задачи и переменными двойственной задачи. Если i-ое соотношение в системе исходной задачи — неравенство, то i-я переменная двойственной задачи $y_i \geq 0$. В противном случае переменная y_i может принимать как положительные, так и отрицательные значения.

Двойственные пары задач обычно подразделяют на симметричные и несимметричные. В симметричной паре двойственных задач ограничения прямой задачи и соотношения двойственной задачи являются неравенствами вида «$\leq$». Таким образом, переменные обеих задач могут принимать только неотрицательные значения.

В терминах двух взаимосвязанных задач формулируются основные теоретические результаты. Приведем основные теоремы, составляющие формальную (математическую) часть линейного программирования.

ЛЕММА 1. *Двойственная к двойственной задаче в точности совпадает с исходной.*

ЛЕММА 2. *Если х и у соответственно допустимые решения прямой и двойственной задач, то (с, х)* $\leq$ *(в, у).*

ТЕОРЕМА *(критерий оптимальности Канторовича)***.** *Пусть* $\bar{x}$ *и* $\bar{y}$ *соответственно допустимые решения прямой и двойственной задач и* $(c, \bar{x}) = (в, \bar{y})$. *Тогда* $\bar{x}$ *и* $\bar{y}$ *являются оптимальными решениями соответствующих задач.*

ТЕОРЕМА (*основная теорема двойственности*). *Если одна из двух взаимно двойственных задач имеет оптимальное решение, то и другая имеет оптимальное решение, причем экстремальные значения целевых функций равны. Если одна из двойственных задач не разрешима вследствие неограниченности целевой функции на множестве допустимых решений, то система ограничений другой задачи противоречива.*

Пример. *Составить двойственную задачу по отношению к задаче, состоящей в максимизации функции* $F = 2x_1 + x_2 + 3x_3$ *при условиях*

$$\begin{cases} -x_1 + 3x_2 - 5x_3 = 12, \\ 2x_1 - x_2 + 4x_3 = 24, \\ 3x_1 + x_2 + x_3 = 18, \end{cases}$$

$$x_1, x_2, x_3 \geq 0.$$

Решение. Число переменных в двойственной задаче равно числу уравнений в системе ограничений, т. е. трем. Коэффициентами в целевой функции двойственной задачи служат свободные члены системы уравнений ограничений, т. е. числа 12, 24, 18. Для данной задачи

$$A = \begin{pmatrix} -1 & 3 & -5 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{pmatrix} \text{ и } A^T = \begin{pmatrix} -1 & 2 & 3 \\ 3 & -1 & 1 \\ -5 & 4 & 1 \end{pmatrix}.$$

Целевая функция исходной задачи исследуется на максимум, а система условий содержит только уравнения. Поэтому в двойственной задаче целевая функция исследуется на минимум, а ее переменные могут принимать любые значения, в том числе и отрицательные. Поскольку все три переменные исходной задачи принимают только лишь неотрицательные значения, то в системе условий двойственной задачи должны быть три неравенства вида «$\geq$». Следовательно, для сформулированной исходной задачи двойственная задача такова: найти минимум функции $F^* = 12y_1 + 24y_2 + 18y_3$ при условиях

$$\begin{cases} -y_1 + 2y_2 + 3y_3 \geq 2, \\ 3y_1 - y_2 + y_3 \geq 1, \\ -5y_1 + 4y_2 + y_3 \geq 3. \end{cases}$$

Составление и анализ двойственных задач для других типов линейных моделей исследования операций

Формальное составление двойственных к линейным задачам исследования операций обычно не вызывает трудностей, за исключением транспортной задачи и ее модификаций. Рассмотрим частный случай задачи, когда

$$\sum_i a_i = \sum_j в_j .$$

В этих условиях задача имеет вид:

$$\sum_i^m \sum_j^n c_{ij} x_{ij} \to \min,$$

$$\sum_i^n x_{ij} = в_j, \; j = \overline{1, n}, \tag{1.19}$$

$$\sum_{j}^{m} x_{ij} = a_i, i = \overline{1,m},$$

$$x_{ij} \geq 0,\ i = \overline{1,m};\ j = \overline{1,n}.$$

Рассмотрим построение двойственной задачи. Возьмем для простоты $m = 2$, $n = 3$ (два пункта производства, три пункта потребления). Тогда имеем:

$$c_{11}x_{11} + c_{12}x_{12} + c_{13}x_{13} + c_{21}x_{21} + c_{22}x_{22} + c_{23}x_{23} \to \min,$$

$$\begin{cases} x_{11} + x_{12} + x_{13} = a_1 \\ x_{21} + x_{22} + x_{23} = a_2 \\ x_{11} + x_{21} = в_1 \\ x_{12} + x_{22} = в_2 \\ x_{13} + x_{23} = в_3\ , \end{cases} \qquad (1.19')$$

$$x_{ij} \geq 0,\ i = \overline{1,2};\ j = \overline{1,3}.$$

Введем векторы $\vec{c}$ и $\vec{x}$. Компонентами вектора $\vec{c}$ будут значения матрицы перевозок $[c_{ij}]$, компонентами вектора $\vec{x}$ — значения неизвестных x_{ij}, для всех $j = \overline{1,n}$; $i = \overline{1,m}$. Получим:

$$\vec{c}^{\,T} = (c_{11}, c_{12}, c_{13}, c_{21}, c_{22}, c_{23});$$
$$\vec{x}^{\,T} = (x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}).$$

Запишем задачу, причем каждой компоненте вектора $\vec{c}$ присвоим значение c_i с номером компоненты ($c_1 = c_{11},\ldots, c_6 = c_{23}$), каждой компоненте вектора $\vec{x}$ — значение x_i с номером компоненты ($x_1 = x_{11},\ldots, x_6 = x_{23}$), получим:

$$c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 + c_5x_5 + c_6x_6 \to \min,$$

$$x_1 + x_2 + x_3 \qquad\qquad = a_{1,}$$

$$x_4 + x_5 + x_6 = a_{2,}$$

$$x_1 + x_4 = в_{1,} \qquad (1.20)$$

$$x_2 + x_5 = в_2,$$

$$x_3 + x_6 = в_3.$$

Записанная выше задача является задачей линейного программирования в канонической форме. В соответствие с правилом построения двойственных задач, двойственная к ней задача будет записана следующим образом:

$$a_1 y_1 + a_2 y_2 + в_1 y_3 + в_2 x_4 + в_3 x_5 \to \max$$

$$y_1 + y_3 \le c_1,$$

$$y_1 + y_4 \le c_2,$$

$$y_1 + y_5 \le c_3,$$

$$y_2 + y_3 \le c_4, \tag{1.21}$$

$$y_2 + y_4 \le c_5,$$

$$y_2 + y_5 \le c_6,$$

y_i не ограничен в знаке.

Введем обозначения

$$y_1 = -u_1;\ y_2 = -u_2\ ;\ y_3 = v_1\ ;\ y_4 = v_2\ ;\ y_5 = v_3$$

и вернемся от номера компоненты вектора $\vec{c}$ к значению $c_{ij}\,(i = \overline{1,2};\ j = \overline{1,3})$.

Задачу (1.21) запишем в виде

$$\sum_j в_j v_j - \sum_i a_i\, u_i \to \max$$

$$v_j - u_i \le c_{ij}\,;\ i = \overline{1,m};\ j = \overline{1,n}$$

$v_j\,;\,u_i$ не ограничены в знаке.

Экономический смысл оптимальных двойственных оценок v_j, u_i легко определить из равенства оптимальных значений целевых функций прямой и двойственной задач. Здесь v_j — оценка потребителя, а u_i — поставщика:

$$\bar{c} = \sum_i \sum_j c_{ij} x_{ij}^{\ *} = \sum_j^n в_j v_j^* - \sum_i^m a_i u_i^* \tag{1.22}$$

Если оценки потребителей v_j^* строго положительны, то местоположение этого потребителя выгодно с точки зрения транспортных затрат, выгодно для подвоза продукции. При увеличении потребности такого потребителя на малую единицу общий объем перевозки увеличивается и минимальные издержки возрастают на v_j^*. Если оценка $u_i^*>0$, то продукцию удобно доставлять потребителю, и если объем производства в этом пункте увеличится на малую единицу, то потребителям выгоднее сократить заказы у других и увеличить их у *i*-го поставщика, при этом общие транспортные затраты снизятся на величину u_i^*.

Поскольку рассматриваемая модель является транспортной моделью закрытого типа, то, значит, изменение правых частей задачи должны компенсировать друг друга. Из равенства значений целевых функций прямой и двойственной задач для оптимальных решений следует, что если увеличится выпуск в *i*-м пункте на малую единицу и увеличится потребность в *j*-м пункте на ту же единицу (так как модель закрытая), то оптимальные затраты увеличатся на величину $\Delta C^* = \mathrm{v}_j^* - u_i^*$. Если в оптимальном плане $x_{ij}^* \geq 0$, (т. е. прямой маршрут (*i, j*) — из *i* пункта в *j*, существовал в оптимальном решении), то по этому маршруту добавляется перевозка единицы продукции

$$\Delta C^* = v_j^* - u_i^* < C_{ij}$$

в силу условий двойственной задачи.

Если перевозка (*i, j*) не входила в оптимальный план, то

$$\Delta C^* = v_j^* - u_i^* < C_{ij}$$

и переброска малой единицы произойдет по окружному, но более дешевому пути.

Таким образом, в отдельности каждая двойственная оценка оптимального решения не имеет определенного экономического смысла, экономический смысл оценки эффективности маршрута имеет лишь разность оптимальных оценок $\Delta C^{*} = v_j^{*} - u_i^{*}$, где ΔC — дополнительные затраты на перевозку дополнительной малой единицы в условиях оптимального плана перевозок.

Лекция 1.2.2. Постоптимальный анализ решения линейных моделей с использованием двойственных оценок: геометрическая интерпретация

Рассмотрим процедуру ситуационного анализа на конкретном примере линейной модели производства.

Пример. *Определить суточную производственную программу небольшого цеха по пошиву женской одежды. Для весенне-летнего сезона модельеры цеха разработали новые модели женских брюк и юбок; известны затраты на пошив этих изделий и цена их реализации на рынке. Требуется установить количество брюк и юбок, которые нужно сшить за сутки. Цифровая информация по данной ситуации приведена в табл.* 1.5.

Таблица 1.5

Производственные факторы	Расходы на одно готовое изделие		Максимально возможный суточный запас
	брюки	юбки	
Ткань, м	1,5	2	42
Трудоемкость, чел.-ч	3	2	60
Фурнитура, долл.	5	5	200
Цена одного изделия, долл.	60	50	—

Изучение рынка сбыта показало, что суточный спрос на брюки никогда не превышает 18 шт. Спрос на юбки обеспечен. Какое количество брюк и юбок в сутки должен сшить цех, чтобы доход от реализации продукции был максимальным?

Решение. Построим математическую модель. Поскольку требуется определить объемы производства, то переменными в модели являются:

x_1 — объем производства брюк в сутки, шт.

x_2 — объем производства юбок в сутки, шт.

При решении рассматриваемой задачи должны быть учтены ограничения на расход производственных факторов (ткани, труда и фурнитуры), а также спрос на готовую продукцию. Это приводит к следующим четырем ограничениям:

1) $1{,}5\,x_1 + 2\,x_2 \le 42$;
2) $3\,x_1 + 2\,x_2 \le 60$;
3) $5\,x_1 + 5\,x_2 \le 200$;
4) $x_1 \le 18$.

Объемы производства продукции не могут принимать отрицательные значения, т. е. $x_1 \ge 0$ и $x_2 \ge 0$. Цель нашего анализа заключается в максимизации дохода, количественным выражением которого является выражение: $60\,x_1 + 50\,x_2 \to \max$. Итак, имеем задачу линейного программирования:

$$F(x_1, x_2) = 60\,x_1 + 50\,x_2 \to \max,$$

$$\begin{cases} 1{,}5x_1 + 2x_2 \le 42 \\ 3x_1 + 2x_2 \le 60 \\ 5x_1 + 5x_2 \le 200 \\ x_1 \le 18 \end{cases},$$

$$x_1, x_2 \ge 0.$$

Решение можно получить графическим способом (рис. 1.3).

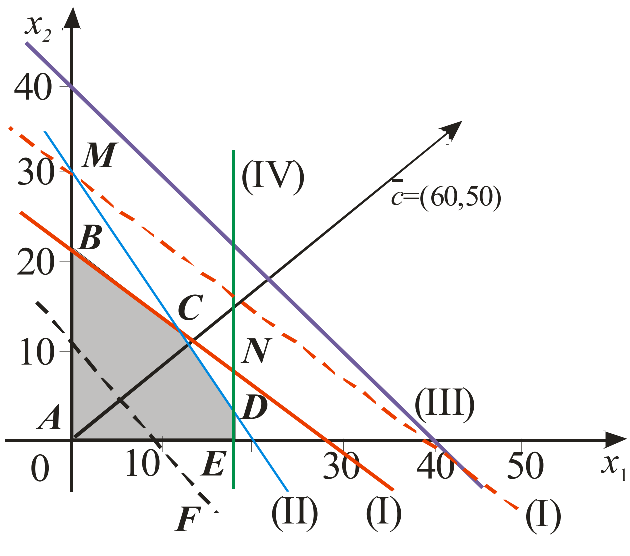


Рис. 1.3. Графическое изображение пространства решений задачи

Искомым пространством решений, в котором одновременно выполняются все ограничения модели, является многоугольник *ABCDE*. Для того чтобы найти оптимальное решение, следует перемещать прямую, характеризующую доход (прямая *F* на рис. 1.3), в направлении возрастания целевой функции до тех пор, пока она не сместится в область недопустимых решений. На рис. 1.3 видно, что оптимальному решению соответствует точка *C*, служащая точкой пересечения прямых ограничений (I) и (II). Определим координаты точки *C*, решив следующую систему:

$$\begin{cases} 1{,}5x_1 + 2x_2 = 42, \\ 3x_1 + 2x_2 = 60. \end{cases}$$

Решение указанной системы уравнений дает $x_1 = 12$, $x_2 = 12$. Полученное решение означает, что цех должен в сутки производить по 12 брюк и юбок. Доход в этом случае равен:

$$F\,(12;\,12) = 60 \cdot 12 + 50 \cdot 12 = 1\,320 \text{ долл.}$$

Теперь, когда оптимальное решение задачи получено, займемся собственно анализом на чувствительность. В рамках такого анализа выявляется чувствительность оптимального решения к определённым изменениям исходной модели. В задаче о швейном цехе, например, может представлять интерес вопрос о том, как повлияют на оптимальное решение увеличение или уменьшение спроса и изменение запасов исходных производственных факторов. Вместе с тем целесообразно также определить влияние на оптимальное решение изменения рыночных цен.

При таком анализе всегда рассматривается некоторая совокупность линейных оптимизационных моделей, т. е., по существу, некоторая модель исследования операций. Это придает модели определенную динамичность, позволяющую исследователю проанализировать влияние возможных изменений исходных условий на полученное ранее оптимальное решение. Динамические характеристики моделей фактически отображают аналогичные характеристики, свойственные реальным процессам. Отсутствие методов, позволяющих выявить влияние возможных изменений параметров модели на оптимальное решение, может привести к тому, что полученное (статическое) решение устареет еще до своей реализации.

В данном учебном пособии для проведения анализа на чувствительность первоначально используются графические методы, поэтому применяемые приемы достаточно просты. Тем не менее нам удастся получить результаты, на которых основываются весьма эффективные методы анализа моделей на чувствительность.

Первая задача анализа на чувствительность: на сколько сократить или увеличить запасы ресурсов

После нахождения оптимального решения представляется логичным выяснить, как отразится на оптимальном решении изменение запасов производственных факторов. Особенно важно проанализировать следующие аспекты:

1. На сколько можно увеличить запас некоторого ресурса для улучшения полученного оптимального значения целевой функции F?
2. На сколько можно снизить запас некоторого ресурса при сохранении полученного оптимального значения целевой функции?

Так как величина запаса каждого из ресурсов фиксируется в правых частях ограничений, то этот вид анализа обычно идентифицируется как анализ модели на чувствительность к правой части ограничений.

Прежде чем ответить на поставленные вопросы, классифицируем ограничения линейной модели как **связывающие** (активные) и **несвязывающие** (неактивные). Прямая, представляющая связывающее ограничение, должна проходить через оптимальную точку. В противном случае соответствующее ограничение будет несвязывающим. На рис. 1.3 связывающими являются только ограничения (I) и (II), т. е. те, которые лимитируют запас ткани и фонд рабочего времени.

Если некоторое ограничение является связывающим, то соответствующий ему ресурс будем называть ***дефицитным****, т. е. он используется полностью.*

Ресурс, с которым ассоциировано несвязывающее ограничение, следует отнести к разряду ***недефицитных*** *(т. е. имеющихся в некотором избытке).*

*Те ограничения, которые в данной ситуации напрямую даже не участвуют в формировании пространства допустимых решений, будем называть **избыточными**.*

Таким образом, при анализе модели на чувствительность к правым частям ограничений относятся:

1. предельно допустимое увеличение запаса дефицитного ресурса, позволяющее улучшить найденное оптимальное решение;
2. предельно допустимое снижение запаса недефицитного ресурса, не изменяющее найденного ранее оптимального значения целевой функции. Информация, полученная в последнем случае, особенно полезна в тех ситуациях, когда излишки недефицитного ресурса могут быть использованы для других целей.

Лекция 1.2.3. Постоптимальный анализ решения линейных моделей с использованием двойственных оценок: геометрическая интерпретация (продолжение)

Рассмотрим первую задачу анализа на чувствительность на примере ситуации со швейным цехом. Дефицитными ресурсами здесь являются ткань (использование которой описывается ограничением I) и фонд рабочего времени (использование которого описывается ограничением II).

Изменение суточного запаса ткани графически будет выражаться в перемещении прямой ограничения (I) параллельно самой себе до точки *M* (дальнейшее увеличение запаса нецелесообразно, т.к. тогда ресурс станет недефицитным). В результате перемещения прямой ограничения (I) пространство допустимых решений увеличится на треугольник *BMC*, оптимальному решению при этом соответствует точка *M*. Таким образом, запас ткани не следует увеличивать сверх того предела, когда соответствующее ограничение (I) становится избыточным и уже не влияет ни на пространство решений, ни на оптимальное решение. Предельный уровень изменения запаса ткани определяется следующим образом: устанавливаются координаты точки *M* (она образована пересечением прямой ограничения (II) и осью $x_1 = 0$). В результате получаем $x_1 = 0$, $x_2 = 30$. Затем путем подстановки координат точки *M* в левую часть ограничения (I) определяется максимально допустимый суточный запас ткани: $1{,}5 \cdot 0 + 2 \cdot 30 = 60$ м. Таким образом, целесообразно увеличить суточный запас ткани на 60 – 42 = 18 м. Величина прироста дохода от реализации в этом случае составит 1500 – 1320 = 180 долл.

Далее рассмотрим вопрос о целесообразности увеличения второго дефицитного ресурса (фонда рабочего времени). В этом случае увеличение суточного фонда времени графически выражается в параллельном перемещении прямой ограничения (II) до точки *N* (см. рис. 1.3). Дальнейшее увеличение запаса данного ресурса нецелесообразно, так как он станет избыточным. В результате перемещения прямой ограничения (П) новым пространством допустимых значений станет многоугольник *ABNE*, а новой оптимальной точкой — точка *N*. Установим ее координаты, решив систему:

$$\begin{cases} 1{,}5x_1 + 2x_2 = 42, \\ x_1 = 18, \end{cases}$$

отсюда x_1=18; x_2=7,5. Подставим координаты точки N в левую часть ограничения (II) и получим максимально допустимый суточный фонд рабочего времени: 3·18 + 2·7,5=69 чел.-ч. Таким образом, изменить суточный запас рабочего времени нужно на 69-60=9 чел.-ч. Доход же от реализации в этом случае составит 60·18 + 50·7,5 = 1455 долл., т. е. увеличится на 1455 – 1320 = =135 долл.

Рассмотрим теперь вопрос об уменьшении правой части несвязывающих и избыточных ограничений. Ограничение (III) является избыточным, поэтому суточный запас фурнитуры можно уменьшить. Графически это изображается как перемещение прямой ограничения (III) до точки C (ведь мы хотим, чтобы это уменьшение не повлияло на оптимальность ранее полученного решения) параллельно самой себе. Оптимальный план по-прежнему определяется точкой C с координатами x_1=12, x_2=12. Чтобы установить требуемую величину суточного запаса фурнитуры, подставим координаты точки C в ограничение (III). Получим 5·12 + 5·12 = 120. Таким образом, снижение запаса фурнитуры составит 120 – 200 = – 80 долл. Величина дохода в этом случае не меняется. Ограничение (IV) фиксирует предельный уровень спроса на брюки. Не изменяя оптимального плана, прямую (IV) можно сдвигать параллельно самой себе до точки C. Так как точка C имеет координаты x_1=12; x_2 = 12, то снижение спроса на брюки до величины 12 никак не повлияет на оптимальность ранее полученного решения. Снижение спроса в данном случае составляет 6 шт. в сутки. Результаты проведенного анализа можно свести в табл. 1.6.

Таблица 1.6

Ресурсы	Название ресурсов	Тип ресурсов	Максимальное изменение запаса ресурсов	Максимальное изменение дохода от реализации, долл.
1	Ткань	Дефицитный	60 – 42 = 18 м	1500 – 1320 = 180

1	2	3	4	5
2	Затраты труда	Дефицитный	69 – 60 = 9 чел.-ч	1455 – 1320 = 135
3	Фурнитура	Избыточный	120–200= –80 долл.	1320 – 1320 = 0
4	Спрос	Недефицит-ный	12 – 18 = –6 шт.	1320 – 1320 = 0

Вторая задача анализа на чувствительность: увеличение объема какого ресурса наиболее выгодно

В первой задаче анализа на чувствительность мы исследовали влияние на оптимум увеличения объема дефицитных ресурсов. При ограничениях на затраты, связанные с дополнительным привлечением ресурсов, естественно определить, какому из ресурсов следует отдать предпочтение при вложении дополнительных средств. Для этого вводится характеристика ценности каждой дополнительной единицы дефицитного ресурса, выражаемая через соответствующее приращение оптимального значения целевой функции. Обозначим ценность дополнительной единицы ресурса i через y_i. Величина y_i определяется из соотношения:

$$y_i = \frac{\text{Максимальное приращение оптимального значения дохода } F}{\text{Максимально допустимый прирост объема ресурса } i}.$$

По данным табл. 2.2 проведем вычисление ценности единицы каждого ресурса:

$$y_i = \frac{180\ \textit{долл.}}{18\ \textit{м}} = 10\ \textit{долл./м}; \quad \begin{cases} y_i = \dfrac{180\ \textit{долл.}}{18\ \textit{м}} = 10\ \textit{долл./м}; \\ y_2 = \dfrac{135\ \textit{долл.}}{9\ \textit{чел.} - \textit{ч}} = 15\ \textit{долл./чел.} - \textit{ч}; \end{cases}$$

$$y_2 = \frac{135\ \textit{долл}}{9\ \textit{чел.} - \textit{ч.}} = 15\ \textit{долл/чел} \cdot \textit{час};$$

$$y_3 = \frac{0}{-80\ долл.} = 0;$$

$$y_4 = \frac{0}{-6\,шт.} = 0\,.$$

Полученные результаты свидетельствуют, что вложения в первую очередь следует направить на увеличение фонда рабочего времени и лишь затем — на закупку дополнительной ткани. Объем недефицитных ресурсов увеличивать не следует.

Третья задача анализа на чувствительность: в каких пределах допустимо изменение коэффициентов функции

Изменение коэффициентов целевой функции оказывает влияние на наклон прямой, которая представляет эту функцию в системе координат. Определение конкретной угловой точки пространства допустимых решений в качестве оптимума зависит, прежде всего, от наклона этой прямой. Это означает, что вариации коэффициентов целевой функции могут привести к изменениям совокупности связывающих ограничений и, следовательно, статуса того или иного ресурса. Таким образом, в рамках анализа на чувствительность к изменениям коэффициентов целевой функции могут исследоваться вопросы:

— каков диапазон изменения того или иного коэффициента целевой функции, при котором не происходит изменения оптимального решения;
— на сколько следует изменить тот или иной коэффициент целевой функции, чтобы изменить статус некоторого ресурса.

Рассмотрим поставленные вопросы на примере задачи швейного цеха. Обозначим доход от реализации брюк и юбок как c_1 и c_2 соответственно. Тогда целевая функция будет выглядеть следующим образом:

$$F = c_1 x_1 + c_2 x_2 \to \max\,.$$

При увеличении c_1 (или уменьшении c_2) прямая, представляющая целевую функцию F, вращается вокруг точки C по часовой стрелке (рис. 1.3). Если же c_1 уменьшается (или c_2 увеличивается), эта прямая вращается в противоположном направлении. Таким образом, точка C будет оставаться оптимальной до тех пор, пока наклон прямой не выйдет за пределы, определяемые наклонами прямых ограничений (I) и (II). Когда наклон прямой F станет равен наклону прямой для ограничения (I), получим две альтернативные угловые точки B и C. Аналогично, если наклон прямой F станет равным наклону прямой для ограничения (II), будем иметь альтернативные оптимальные угловые точки C и D. Как только наклон прямой F выйдет за пределы указанного интервала, получим некоторое новое оптимальное решение (точки B или D).

Вычислим границы интервалов возможных колебаний c_1 и c_2, при которых точка C останется оптимальной. Зафиксируем c_2=50. Крайние значения коэффициента c_1 можно определить из равенства наклонов прямой целевой функции F и прямых ограничений (I) и (II). Тангенс угла наклона для прямой F равен $c_1/50$, а для прямых (I) и (II), соответственно, 3/4 и 3/2. Минимальное значение c_1 определяем из равенства $c_1/50 = 3/4$, тогда минимальное значение c_1 =37,5 долл., а максимальное значение c_1 находим из равенства c_1/c_2=3/2, где max c_1 =75 долл. Таким образом, интервал изменения c_1, в котором точка C по-прежнему останется единственной оптимальной, определяется неравенством $37{,}5 \le c_1 \le 75$.

Аналогичным образом анализ выполняется для коэффициента c_2. Фиксируем значение c_1=60, тогда тангенс угла наклона прямой целевой функции определяется соотношением $60/c_2$. Применив уже известное равенство тангенсов углов наклона прямой F и прямых (I) и (II), получим:

$$60/c_2 = 3/4, \text{ откуда } \max c_2 = 80;$$

$$60/c_2 = 3/2, \text{ откуда } \min c_2 = 40.$$

Можно заметить, что, как только коэффициент c_1 =37,5 долл., ресурс (II) становится недефицитным. Для швейного цеха это означает, что, если доход от продажи одних брюк станет меньше 37,5 долл., надо пересматривать суточную производственную программу, которая теперь будет определять максимальное количество юбок (т.е. $x_1 = 21$, $x_2 = 0$). Когда значение c_1 превысит 75 долл., по суточной производственной программе выпуск составит 18 брюк и 3 юбок (оптимальный план — точка D).

Соответствующие выводы нужно будет сделать и при отклонении цены одной юбки c_2 за пределы интервала $40 \le c_2 \le 80$.

Этот пример характеризует основные принципы анализа моделей на чувствительность (после нахождения оптимального решения).

Лекция 1.2.4. Постоптимальный анализ решения линейных моделей с использованием двойственных оценок: аналитический подход

После геометрической интерпретации процесса решения линейной модели представляется целесообразным расширить методику анализа на чувствительность ее решения для линейных задач большой размерности (когда количество переменных n и количество ограничений m не исчерпываются возможностями графического метода и предполагается решение с использованием компьютерных средств).

Вернемся к задаче цеха по пошиву женской одежды (лекция 1.2.2). Математическая модель имела вид:

$$60x_1 + 50x_2 \to \max$$
$$\begin{cases} 1{,}5x_1 + 2x_2 \le 42, \\ 3x_1 + 2x_2 \le 60, \\ 5x_1 + 5x_2 \le 200, \\ x_1 \le 18, \end{cases}$$
$$x_1,\ x_2 \ge 0.$$

Здесь переменные x_1 и x_2 представляли собой объемы производства брюк и юбок в сутки (шт.). Приведем задачу к канонической форме записи для нахождения оптимального решения:

$$F = 60x_1 + 50x_2 \to \max$$
$$\begin{cases} 1{,}5x_1 + 2x_2 + x_3 = 42, \\ 3x_1 + 2x_2 + x_4 = 60, \\ 5x_1 + 5x_2 + x_5 = 200, \\ x_1 + x_6 = 18, \end{cases}$$
$$x_1, x_2, \ldots, x_6 \ge 0.$$

В канонической форме модели задачи появляются дополнительные переменные x_3, x_4, x_5 и x_6. Рассмотрим их экономическую интерпретацию, которая нам понадобится для постоптимального анализа. Переменная x_3 представляет собой неиспользованный объем ткани в сутки (м). Причем если $x_3 = 0$ в оптимальном плане, то ресурс «ткань» используется полностью и остатка нет, а если в оптимальном плане задачи x_3 принимает значение больше нуля, то ресурс используется не полностью, а величина его остатка определяется значением переменной x_3. Аналогично интерпретируются переменные x_4 и x_5, которые также служат дополнительными переменными в ограничениях ресурсного типа и отражают использование затрат труда и фурнитуры. Переменная x_6 входит в ограничение, связанное со спросом на продукцию (брюки), который в прямом смысле слова не является ресурсом. Однако если рассматривать спрос на продукцию как некий «ресурс», отражающий объем представительства фирмы, производящей данную продукцию, на рынке, то мы можем дополнительную переменную x_6 в четвертом ограничении также трактовать как дополнительную переменную в ограничении ресурсного типа, где спрос выступает в качестве специфического «ресурса».

Ниже приведена симплекс-таблица с решением указанной задачи (табл. 1.7). Мы не будем здесь останавливаться на технологии решения с использованием симплекс-процедуры, поскольку это подробно изложено в лекции 1.1.4 данного пособия.

Как видно из симплекс-таблицы (табл. 1.7), оптимальным решением данной задачи является суточное производство брюк в количестве 12 штук, юбок — в количестве 12 штук. Суточный доход от реализации этих объемов продукции составит 1 320 долл. Что касается использования ресурсов, то запасы ткани и затрат труда расходуются полностью. Имеется остаток накладных расходов в сумме 80 долл. Кроме того, суточный объем производства брюк меньше предельной величины спроса на рынке на 6 штук.

После характеристики оптимального решения проведем анализ его чувствительности к изменению запасов ресурсов и колебаниям цен готовой продукции. Оговоримся сразу, что этапы анализа на чувствительность с помощью аналитических симплекс-таблиц несколько отличаются от предложенной выше схемы, где использовался графический метод, в силу значительного сокращения расчетных манипуляций.

Таблица 1.7

				60	50	0	0	0	0
i	*базис*	*С баз*	P_0	P_1	P_2	P_3	P_4	P_5	P_6
1	P_3	0	42	1,5	2	1	0	0	0
2	P_4	0	60	3	2	0	1	0	0
3	P_5	0	200	5	5	0	0	1	0
4	P_6	0	18	1	0	0	0	0	1
5			0	–60	–50	0	0	0	0
1	P_3	0	15	0	2	1	0	0	$-\frac{3}{2}$
2	P_4	0	6	0	2	0	1	0	–3
3	P_5	0	110	0	5	0	0	1	–5
4	P_1	60	18	1	0	0	0	0	1
5			1080	0	–50	0	0	0	60
1	P_3	0	9	0	0	1	–1	0	$\frac{3}{2}$
2	P_2	50	3	0	1	0	$\frac{1}{2}$	0	$-\frac{3}{2}$
3	P_5	0	95	0	0	0	$-\frac{5}{2}$	1	$\frac{5}{2}$
4	P_1	60	18	1	0	0	0	0	1
5			1230	0	0	0	25	0	–25
1	P_6	0	6	0	0	$\frac{2}{3}$	$-\frac{2}{3}$	0	1
2	P_2	50	12	0	1	1	$-\frac{1}{2}$	0	0
3	P_5	0	80	0	0	$-\frac{5}{3}$	$-\frac{5}{6}$	1	0
4	P_1	60	12	1	0	$-\frac{2}{3}$	$\frac{2}{3}$	0	0
5			1320	0	0	10	15	0	0

Этап 1. Определение статуса ресурсов

Для получения результатов этого этапа достаточно обратиться к последней (оптимальной) симплекс-таблице (табл. 1.7), переписав данные из нее в следующей форме (табл. 1.8).

Таблица 1.8

№ ресурса	Название ресурса	Значение дополнительной переменной ресурса	Значение двойственной переменной ресурса	Статус ресурсов
1	Ткань	$x_3^* = 0$	$\Delta_3^* = y_1^* = 10$	дефицитный
2	Затраты труда	$x_4^* = 0$	$\Delta_4^* = y_2^* = 15$	дефицитный
3	Фурнитура	$x_5^* = 80$	$\Delta_5^* = y_3^* = 0$	недефицитный
4	Спрос	$x_6^* = 6$	$\Delta_6^* = y_4^* = 0$	недефицитный

В табл. 1.8 результатов анализа по выявлению статуса ресурсов используются значения дополнительных (x_3^*, x_4^*, x_5^*, x_6^*) и двойственных (y_1^*, y_2^*, y_3^*, y_4^*) переменных ресурсов, по которым и были сделаны выводы по статусу (типу) ресурсов, приведенные в табл. 1.8. Содержательно эти выводы состоят в следующем:

— *ресурс «ткань»* используется полностью, это подтверждается значением дополнительной переменной ($x_3^* = 0$) и двойственной переменной ($y_1^* = 10$), которые соответствуют характеристикам дефицитного ресурса;

— *ресурс «затраты труда»* используется полностью, это подтверждается значением дополнительной переменной ($x_4^* = 0$) и двойственной переменной ($y_2^* = 15$), которые соответствуют характеристикам дефицитного ресурса;

— *ресурс «фурнитура»* используется не полностью, это подтверждается значением дополнительной переменной ($x_5^* = 80$) и двойственной переменной ($y_3^* = 0$), которые соответствуют характеристикам недефицитного ресурса;

— *ресурс «спрос»* используется не полностью, это подтверждается значением дополнительной переменной ($x_6^* = 6$) и двойственной переменной ($y_4^* = 0$), которые соответствуют характеристикам недефицитного ресурса.

Лекция 1.2.5. Постоптимальный анализ решения линейных моделей с использованием двойственных оценок: аналитический подход (продолжение)

Этап 2. Определение эффективности приращения запасов ресурсов

Напомним, что ранее был определен статус всех ресурсов, участвующих в производстве брюк и юбок. Дефицитными ресурсами оказались ткань и затраты труда, а недефицитными — фурнитура и спрос. Для дефицитных ресурсов представляет интерес определение величины целесообразного прироста запаса, который повлечет за собой увеличение целевой функции (в нашем случае — дохода от реализации). Для недефицитных ресурсов оценивается величина возможного снижения запаса, при котором исходное оптимальное решение сохраняется, а излишние запасы ресурсов сокращаются.

Обозначим как ε_1 величину изменения запаса по дефицитному ресурсу «ткань». Тогда запас этого ресурса будет определяться выражением $42 + \varepsilon_1$. Нас будет интересовать в первую очередь определение величины $\varepsilon_1 > 0$, поскольку ткань является дефицитным ресурсом. Однако с помощью аналитических таблиц симплекс-процедуры (табл. 1.9) можно выяснить также, на сколько может уменьшиться запас дефицитного ресурса, чтобы производственный процесс мог осуществиться (при возникновении, например, ситуации внезапного срыва поставок ткани).

Вспомним, что элементами столбца P_0 служат значения базисных переменных, которые по условиям исходной ситуации были определены как неотрицательные. Тогда нам нужно выяснить, при каких значениях ε_1 выражения, входящие в столбец P_0 оптимальной симплекс-таблицы, будут неотрицательными. Получаем систему неравенств следующего вида:

$$\begin{cases} 6+\dfrac{2}{3}\varepsilon_1 \ge 0, \\ 12+\varepsilon_1 \ge 0, \\ 80-\dfrac{5}{3}\varepsilon_1 \ge 0, \\ 12-\dfrac{2}{3}\varepsilon_1 \ge 0. \end{cases}$$

Таблица 2.5

				60	50	0	0	0	0
i	*базис*	*С баз*	P_0	P_1	P_2	P_3	P_4	P_5	P_6
1	P_3	0	$42+\varepsilon_1$	1,5	2	1	0	0	0
2	P_4	0	60	3	2	0	1	0	0
3	P_5	0	200	5	5	0	0	1	0
4	P_6	0	18	1	0	0	0	0	1
5			0	–60	–50	0	0	0	0
1	P_3	0	$15+\varepsilon_1$	0	2	1	0	0	$-\frac{3}{2}$
2	P_4	0	6	0	2	0	1	0	–3
3	P_5	0	110	0	5	0	0	1	–5
4	P_1	60	18	1	0	0	0	0	1
5			1080	0	–50	0	0	0	60
1	P_3	0	$9+\varepsilon_1$	0	0	1	–1	0	$\frac{3}{2}$
2	P_2	50	3	0	1	0	$\frac{1}{2}$	0	$-\frac{3}{2}$
3	P_5	0	95	0	0	0	$-\frac{5}{2}$	1	$\frac{5}{2}$
4	P_1	60	18	1	0	0	0	0	1
5			1230	0	0	0	25	0	–25
1	P_6	0	$6+\frac{2}{3}\varepsilon_1$	0	0	$\frac{2}{3}$	$-\frac{2}{3}$	0	1
2	P_2	50	$12+\varepsilon_1$	0	1	1	$-\frac{1}{2}$	0	0
3	P_5	0	$80-\frac{5}{3}\varepsilon_1$	0	0	$-\frac{5}{3}$	$-\frac{5}{6}$	1	0
4	P_1	60	$12-\frac{2}{3}\varepsilon_1$	1	0	$-\frac{2}{3}$	$\frac{2}{3}$	0	0
5			$1320+10\varepsilon_1$	0	0	10	15	0	0

В случае исследования возможного увеличения запаса дефицитного ресурса «ткань» мы должны рассматривать случай, когда $\varepsilon_1 > 0$. Тогда решением указанной системы неравенств будет $\varepsilon_1 \le 18$, т. е. целесообразный прирост ткани не должен превышать 18 м в сутки. Если же прирост превысит 18 м, то ткань из разряда дефицитных ресурсов перейдет в недефицитные, и оптимальное решение будет определяться величинами запасов других ресурсов.

Интервально определить величину целесообразного прироста ткани можно следующим образом: $0 < \varepsilon_1 \le 18$ или, оперируя полным объемом запаса ткани, $42 <$ запас ткани ≤ 60.

Определим, на сколько может снизиться запас дефицитного ресурса «ткань», чтобы производственный процесс мог осуществиться. Рассмотрим случай $\varepsilon_1 < 0$. Тогда решением приведенной выше системы неравенств будет $\varepsilon_1 \ge -9$, т. е. в случае снижения запаса ткани ниже чем до уровня 33 м линейная модель швейного цеха не будет иметь решения (в силу несовместности системы ограничений).

Проведем аналогичные рассуждения по второму ресурсу — «затраты труда», который также является дефицитным. Обозначим изменение запаса этого ресурса как ε_2, тогда исходный запас трудозатрат составит $60 + \varepsilon_2$. Проследим сформированные изменения в ходе решения симплекс-методом в табл. 1.10.

Исходя из элементов столбца P_0 оптимальной симплекс-таблицы, получим систему неравенств:

$$\begin{cases} 6 - \dfrac{2}{3}\varepsilon_2 \ge 0, \\ 12 - \dfrac{1}{2}\varepsilon_2 \ge 0, \\ 80 - \dfrac{5}{6}\varepsilon_2 \ge 0, \\ 12 + \dfrac{2}{3}\varepsilon_2 \ge 0. \end{cases}$$

Таблица 1.10

				60	50	0	0	0	0
i	*базис*	C *баз*	P_0	P_1	P_2	P_3	P_4	P_5	P_6
1	P_3	0	42	1,5	2	1	0	0	0
2	P_4	0	$60 + \varepsilon_2$	3	2	0	1	0	0
3	P_5	0	200	5	5	0	0	1	0
4	P_6	0	18	1	0	0	0	0	1
5			0	–60	–50	0	0	0	0
1	P_3	0	15	0	2	1	0	0	$-\frac{3}{2}$
2	P_4	0	$6 + \varepsilon_2$	0	2	0	1	0	–3
3	P_5	0	110	0	5	0	0	1	–5
4	P_1	60	18	1	0	0	0	0	1
5			1080	0	–50	0	0	0	60
1	P_3	0	9	0	0	1	–1	0	$\frac{3}{2}$
2	P_2	50	$3+\frac{1}{2}\varepsilon_2$	0	1	0	$\frac{1}{2}$	0	$-\frac{3}{2}$
3	P_5	0	95	0	0	0	$-\frac{5}{2}$	1	$\frac{5}{2}$
4	P_1	60	18	1	0	0	0	0	1
5			$1230+25\varepsilon_2$	0	0	0	25	0	-25
1	P_6	0	$6-\frac{2}{3}\varepsilon_2$	0	0	$\frac{2}{3}$	$-\frac{2}{3}$	0	1
2	P_2	50	$12-\frac{1}{2}\varepsilon_2$	0	1	1	$-\frac{1}{2}$	0	0
3	P_5	0	$80-\frac{5}{6}\varepsilon_2$	0	0	$-\frac{5}{3}$	$-\frac{5}{6}$	1	0
4	P_1	60	$12+\frac{2}{3}\varepsilon_2$	1	0	$-\frac{2}{3}$	$\frac{2}{3}$	0	0
5			$1320+15\varepsilon_2$	0	0	10	15	0	0

При определении целесообразного прироста запаса трудозатрат ($\varepsilon_2 > 0$) решением данной системы неравенств будет $\varepsilon_2 \leq 9$, т. е. увеличивать трудозатраты имеет смысл до величины 60 + 9 = 69 чел.-ч. Интервал, в котором находится целесообразное увеличение запаса трудозатрат, выглядит так:

$$60 < \text{запас трудозатрат} \leq 69.$$

Возможное снижение трудозатрат ($\varepsilon_2 < 0$), при котором модель будет иметь решение, определяется четвертым ограничением системы и выглядит как $\varepsilon_2 \geq -18$. Это означает, что объем трудозатрат, меньший чем 42 чел.-ч, не позволит выпускать готовую продукцию.

По приведенным выше расчетам, связанным с анализом дефицитных ресурсов, становится понятным формирование системы неравенств, состоящих из элементов столбца P_0, на которые наложено условие неотрицательности. Элементы столбца P_0 в симплекс-таблице, где рассматриваются возможные изменения запасов ресурсов, состоят из элементов столбца P_0 оптимальной симплекс-таблицы нахождения исходного оптимального плана, к которым добавлено произведение ε на коэффициент столбца матрицы A, соответствующего дополнительной переменной исследуемого ресурса.

Третий ресурс («фурнитура») является недефицитным, поэтому при анализе его изменения нас будет интересовать только величина возможного снижения запаса. Обозначим как ε_3 величину возможного снижения запаса фурнитуры и рассмотрим случай $\varepsilon_3 < 0$. Система неравенств выглядит следующим образом:

$$\begin{cases} 6 + 0 \cdot \varepsilon_3 \geq 0, \\ 12 + 0 \cdot \varepsilon_3 \geq 0, \\ 80 + \varepsilon_3 \geq 0, \\ 60 + 0 \cdot \varepsilon_3 \geq 0. \end{cases}$$

Очевидно, что решение определяется третьим неравенством системы ($\varepsilon_3 \geq -80$) и состоит в том, что запасы фурнитуры можно снизить на 80 долл. в сутки (до уровня 120 долл.), причем это не нанесет ущерба реализации выбранной производственной стратегии (производить 12 брюк и 12 юбок).

При анализе недефицитного ресурса «спрос» (обозначим величину его целесообразного снижения как ε_4, причем $\varepsilon_4 < 0$) решение будет определяться ограничением-неравенством вида $6 + \varepsilon_4 \geq 0$, откуда $\varepsilon_4 \geq -6$. Таким образом, величина спроса на брюки, даже снизившись до уровня 18 – 6 = 12 штук в сутки, не отразится на выбранной швейной фабрикой стратегии.

Следует заметить, что проведение анализа по выявлению величин целесообразного снижения запасов недефицитных ресурсов, по сути дела, повторяет сделанные ранее выводы, поскольку найденные величины являются одновременно значениями дополнительных переменных x_5 и x_6, которые установлены в ходе выполнения этапа 1.

Этап 3. Определение чувствительности оптимального решения к колебаниям цен готовой продукции

Как было показано выше, оптимальное решение производственной задачи швейного цеха — выпуск 12 брюк и 12 юбок в сутки, что обеспечивало цеху получение дохода равного 1320 долл. Рассмотрим, как будет меняться это оптимальное решение, если на рынке будут происходить колебания цен на юбки и брюки.

Обозначим как τ_1 величину изменения цены на брюки, причем τ_1 может быть как ≤ 0, так и ≥ 0, поскольку рассматривается как возможное повышение, так и понижение цены на рынке. Определим интервал колебания цены на брюки, при котором найденный ранее оптимальный план сохранит свою устойчивость. Окончательное выражение для цены на брюки с учетом возможных колебаний примет вид: $c_1 + \tau_1$ или $60 + \tau_1$. Проследим, как это изменение отразится на таблице симплекс-процедуры (табл. 1.11).

Поскольку цены готовой продукции участвуют в формировании величин оценочной строки, то для сохранения устойчивости исходного оптимального решения необходимо, чтобы все значения оценок Δ_j были неотрицательны (по критерию оптимальности симплекс-метода).

Таблица 1.11

				60	50	0	0	0	0
i	*базис*	*С баз*	P_0	P_1	P_2	P_3	P_4	P_5	P_6
1	P_6	0	6	0	0	$\frac{2}{3}$	$-\frac{2}{3}$	0	1
2	P_2	50	12	0	1	1	$-\frac{1}{2}$	0	0
3	P_5	0	80	0	0	$-\frac{5}{3}$	$-\frac{5}{6}$	1	0
4	P_1	$60+\tau_1$	12	1	0	$-\frac{2}{3}$	$\frac{2}{3}$	0	0
5			1320	0	0	$10-\frac{2}{3}\tau_1$	$15+\frac{2}{3}\tau_1$	0	0

Оценки Δ_1, Δ_2, Δ_5, Δ_6 равны нулю, поэтому нас будут интересовать только Δ_3 и Δ_4. Получим систему неравенств

$$\begin{cases} 10-\frac{2}{3}\tau_1 \geq 0, \\ 15+\frac{2}{3}\tau_1 \geq 0. \end{cases}$$

Рассмотрим, когда цена на брюки повышается ($\tau_1 > 0$). В этом случае решением указанной системы неравенств будет $\tau_1 \leq 15$. В случае понижения цены на брюки ($\tau_1 < 0$) решением системы неравенств будет $\tau_1 \geq -22{,}5$. Окончательный интервал возможного колебания цен на брюки выглядит следующим образом: $-22{,}5 \leq \tau_1 \leq 15$ или $37{,}5 \leq c_1 \leq 75$. Интерпретация найденного интервала колебания цен на брюки состоит в том, что если рыночная цена на брюки колеблется в указанных пределах, то швейный цех может по-прежнему производить 12 брюк и 12 юбок, что обеспечит ему максимальный суточный доход (величина которого будет отличаться от исходной, но будет наибольшей среди возможных). Как только цена на брюки станет ниже чем 37,5 долл. или выше чем 75 долл., швейному цеху необходимо будет пересмотреть суточную производственную программу, поскольку прежняя уже не будет обеспечивать максимального дохода. Следует пояснить, что интервальное неравенство рассматривается именно как нестрогое, поскольку в случае равенства цены на брюки в 37,5 долл. или в 75 долл. возникает ситуация неединственности решения (в геометрической интерпретации это одна из граней многоугольника решений), когда исходный план (производить 12 юбок и 12 брюк) сохраняет свой статус как один из возможных оптимальных.

Аналогичным образом определим интервал возможного колебания рыночной цены на юбки. Обозначим как τ_2 величину возможного изменения рыночной цены, причем τ_2 может быть как ≤ 0, так и ≥ 0. Тогда выражение для цены на юбки будет иметь вид: $c_2 + \tau_2$ или $50+\tau_2$.

Рассмотрим изменения в последней (оптимальной) симплекс-таблице, если цена на юбки определяется как $50+\tau_2$ (табл. 1.12).

Следуя приведенным выше рассуждениям, когда анализировалась рыночная цена брюк, построим систему неравенств:

$$\begin{cases} 10 + \tau_2 \geq 0, \\ 15 - \dfrac{1}{2}\tau_2 \geq 0. \end{cases}$$

Рассмотрим, когда цена на юбки повышается ($\tau_2 > 0$). В этом случае решением указанной системы неравенств будет $\tau_2 \leq 30$. В случае понижения цены на юбки ($\tau_2 < 0$) решением системы неравенств будет $\tau_2 \geq -10$. Окончательный интервал возможного колебания цен на юбки выглядит следующим образом: $-10 \leq \tau_2 \leq 30$ или $40 \leq c_2 \leq 80$. Интерпретация найденного интервала колебания цен на юбки аналогична предыдущим рассуждениям: если рыночная цена на юбки колеблется в указанных пределах, то швейный цех может по-прежнему производить 12 брюк и 12 юбок, что обеспечит ему максимальный суточный доход (величина которого будет отличаться от исходной, но будет наибольшей среди возможных). Как только цена на юбки станет ниже чем 40 долл. или выше чем 80 долл., швейному цеху необходимо будет пересмотреть суточную производственную программу, поскольку прежняя уже не будет обеспечивать максимального дохода. Здесь также интервальное неравенство рассматривается как нестрогое, поскольку в случае равенства цены на юбки крайним значениям интервала исходный план (производить 12 юбок и 12 брюк) сохраняет свой статус как один из возможных оптимальных.

Таблица 1.12

				60	50	0	0	0	0
i	*базис*	*С баз*	P_0	P_1	P_2	P_3	P_4	P_5	P_6
1	P_6	0	6	0	0	$\frac{2}{3}$	$-\frac{2}{3}$	0	1
2	P_2	$50+\tau_2$	12	0	1	1	$-\frac{1}{2}$	0	0
3	P_5	0	80	0	0	$-\frac{5}{3}$	$-\frac{5}{6}$	1	0
4	P_1	60	12	1	0	$-\frac{2}{3}$	$\frac{2}{3}$	0	0
5			1320	0	0	$10+\tau_2$	$15-\frac{1}{2}\tau_2$	0	0

Следует обратить внимание на то, что в построении системы неравенств каждый раз использовались математические выражения оценок, где неизменной оставалась составляющая, не связанная с τ_1 (или τ_2), а добавлялась компонента, состоящая из произведения τ_1 (или τ_2) на коэффициент, соответствующий данному виду производственной деятельности (пошив брюк или юбок) в матрице *A* оптимальной симплекс-таблицы. При понимании этого факта единственно необходимой информацией для проведения постоптимального анализа является только и единственно расчетная таблица симплекс-процедуры.

Раздел 2. Нелинейные и специальные модели исследования операций

ТЕМА 2.1. НЕЛИНЕЙНОСТЬ В ЭКОНОМИЧЕСКИХ ПРОЦЕССАХ

Лекция 2.1.1. Постановка задачи и методы решения для моделей нелинейного программирования

Во многих экономических исследованиях операций зависимости между постоянными и переменными факторами при более детальном рассмотрении оказываются нелинейными. Как правило, в теории управления такие показатели, как прибыль, себестоимость, совокупные транспортные затраты, капитальные затраты на производство, зависят от объема производства (расходы ресурсов, объема перевозок и др.) нелинейно. В этом случае возникает задача нелинейного программирования, математическая модель которой в векторной форме может быть представлена как определение максимального (минимального) значения функции $f(x_1, x_2, \ldots x_n)$:

$$\max f(x) = \max f(x_1, x_2, \ldots x_n) \quad \text{или} \quad \min f(x) = \min f(x_1, x_2, \ldots x_n) \qquad (2.1)$$

при условии, что её переменные удовлетворяют соотношениям

$$g_i(x_1, x_2, \ldots x_n) \le b_i \qquad (i = 1, \ldots, k), \qquad (2.2)$$

$$g_i(x_1, x_2, \ldots x_n) = b_i \qquad (i = k+1, \ldots, m), \qquad (2.3)$$

где f и g_i — некоторые известные функции n переменных, а b_i — заданные числа. Функции f и $g_i(x)$ — нелинейные.

Сведение задачи условной оптимизации к безусловной

Данный метод применим для случая, если задача имеет ограничения только типа равенств, т. е. (2.3). Суть этого метода состоит в том, что за счет ограничений-равенств в задачи уменьшается число переменных.

Будем считать, что система из m ограничений относительно n переменных

$$g_i(x) - b_i = 0, \quad i = 1,2,...,m \tag{2.4}$$

может быть разрешима относительно части своих переменных. Не нарушая общности, можно считать, что в системе (2.4) переменные $x_1, x_2, ..., x_m$ — зависимые, а $x_m, x_{m+1}, ..., x_n$ — независимые. В этом случае из системы (2.4) получим выражения для зависимых переменных через независимые:

$$x_j = \varphi_j(x_{m+1}, x_{m+2}, ..., x_n), \quad j = 1, 2, ..., m. \tag{2.5}$$

Подставляя (2.4) в (2.5), получим новую задачу:

$$\max f(\varphi_1(x_{m+1}, ..., x_n), ..., \varphi_m(x_{m+1}, ..., x_n), x_{m+1}, ..., x_n), \tag{2.6}$$

в которой оптимизируемая функция уже зависит от $n – m$ переменных.

Пример. *Найти минимум функции* $x_1^2 + x_2^2 + x_1 - x_2 + x_3 - 2x_4 + x_5$, *если*

$$x_1 + x_2 + x_3 = 5,$$

$$x_1 + 2x_2 + x_3 + x_4 = 10,$$

$$5x_1 + 5x_2 + 3x_3 + x_4 + x_5 = 25,$$

$$x_1, ..., x_5 \geq 0.$$

Решение. В задаче требуется найти минимум функции относительно пяти переменных, определенных на множестве с тремя ограничениями. Следовательно, две переменные будут независимыми, а три — зависимыми. Пусть зависимыми переменными будут x_3, x_4 и x_5. Из ограничений найдем

$$\begin{cases} x_3 = 5 - x_1 - x_2 \geq 0 \\ x_4 = 5 - x_2 \geq 0 \\ x_5 = 5 + 8x_1 - x_2 \geq 0 \end{cases}.$$

Подставив найденные выражения для x_3, x_4 и x_5 в функцию $f(x)$, получим

$$\min_{x_1, x_2 \geq 0} f(x) = \min_{x_1, x_2 \geq 0} \left(x_1^2 + x_2^2 + x_1 - x_2 + (5 - x_1 - x_2) - 2(5 - x_2) + (5 + 8x_1 - x_2) \right) =$$

$$= \min_{x_1, x_2 \geq 0} \left(x_1^2 + x_2^2 + 8x_1 - x_2 \right)$$

при ограничениях

$$\begin{aligned} & 5 - x_1 - x_2 \geq 0 \\ & 5 - x_2 \geq 0 \\ & 5 + 8x_1 - x_2 \geq 0 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}.$$

Далее эту задачу можно решить графически (рис. 2.1).

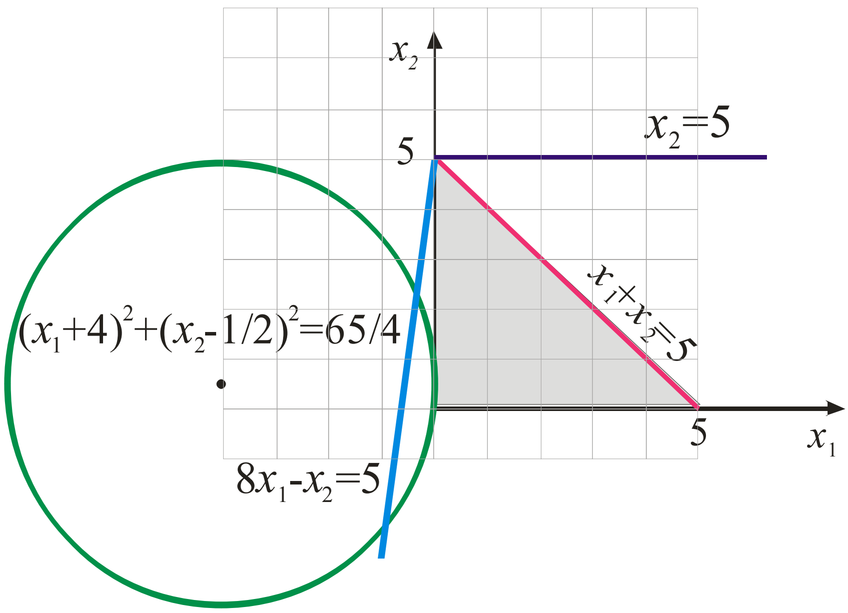


Рис. 2.1. Графическое решение задачи

Т. к. минимизируемая функция после преобразования стала квадратичной относительно двух переменных: $f(x_1, x_2) = (x_1 + 4)^2 + \left(x_2 - \frac{1}{2}\right)^2 - \frac{65}{4}$, то из ее вида следует, что она будет принимать наименьшее значение при x_1=0 и x_2=1/2. Возвращаясь к исходной задаче, получим ответ:

$$x_1^* = 0, \quad x_2^* = \frac{1}{2}, \quad x_3^* = 5 - 0 - \frac{1}{2} = \frac{9}{2}, \quad x_4^* = 5 - \frac{1}{2} = \frac{9}{2}, \quad x_5^* = 5 + 0 - \frac{1}{2} = \frac{9}{2},$$

$$f^* = \frac{65}{4}.$$

Метод множителей Лагранжа

Рассмотрим частный случай общей задачи нелинейного программирования, предполагая, что система ограничений содержит только уравнения, отсутствуют условия неотрицательности переменных и

$$f(x_1, x_2, \ldots x_n) \to \max(\min);$$

$$g_i(x_1, x_2, \ldots x_n) = b_i (i = 1, \ldots, m).$$

В курсе математического анализа данную задачу называют задачей на условный экстремум или классической задачей оптимизации. Чтобы найти решение этой задачи, вводят набор переменных $\lambda_1, \lambda_2, \ldots, \lambda_m$, называемых множителями Лагранжа, и составляют функцию Лагранжа

$$F(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m) = f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{m} \lambda_i \left[b_i - g_i(x_1, x_2, \ldots, x_n)\right]. \quad (2.7)$$

Далее находят частные производные

$$\frac{\partial F}{\partial x_j} \ (j = 1, \ldots, n) \text{ и } \frac{\partial F}{\partial \lambda_i} \ (i = 1, \ldots, m)$$

и рассматривают систему $n+m$ уравнений:

$$\begin{cases} \dfrac{\partial F}{\partial x_j} = \dfrac{\partial f}{\partial x_j} - \sum\limits_{i=1}^{m} \lambda_i \dfrac{\partial g_i}{\partial x_j} = 0 \\ \\ \dfrac{\partial F}{\partial \lambda_i} = b_i - g_i(x_1, x_2, ..., x_n) = 0 \end{cases}$$

с $n+m$ неизвестными $x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m$. Всякое решение системы уравнений определяет точку $X^{\mathrm{T}} = (x_1, x_2, ..., x_n)$, в которой может иметь место экстремум функции $f(x_1, x_2, ..., x_n)$. Следовательно, решив данную систему уравнений, получают все точки, в которых указанная функция может иметь экстремальные значения. Дальнейшее исследование найденных точек проводят так же, как и в случае безусловного экстремума.

Таким образом, определение экстремальных точек задачи методом множителей Лагранжа включает следующие этапы:

— составляют функцию Лагранжа;
— находят частные производные от функции Лагранжа по переменным x_j и λ_i, приравнивают их к нулю;
— решая данную систему уравнений, находят точки, в которых целевая функция задачи может иметь экстремум;
— среди точек, подозрительных на экстремум, находят такие, в которых достигается экстремум, и вычисляют значение функции в этих точках.

Лекция 2.1.2. Постановка задачи и методы решения для моделей выпуклого программирования

Теперь рассмотрим задачу нелинейного программирования с условиями неотрицательности переменных:

$$f(x_1, x_2, ..., x_n) \to \max$$

$$g_i(x_1, x_2, ..., x_n) - b_i \geq 0 \quad (i = 1, ..., m) \tag{2.8}$$

$$x_j \geq 0 \quad (j = 1, ..., n),$$

где f и g_i — некоторые нелинейные функции n переменных $x_1, x_2, ..., x_n$.

Для решения сформулированной задачи в такой общей постановке не существует универсальных методов. Однако для отдельных классов задач, в которых сделаны дополнительные ограничения относительно свойств функций f и g_i, есть эффективные методы решения. В частности, ряд таких методов имеется для решения задач нелинейного программирования при условии, что f — вогнутая (выпуклая) функция и область допустимых решений, определяемая ограничениями, выпуклая.

Множество $U \in E^n$ является ***выпуклым,*** *если вместе с любыми двумя точками $x^{(1)}$ и $x^{(2)}$ из этого множества U, ему принадлежит и отрезок, их соединяющий:*

$$\alpha x^{(1)} + (1-\alpha)x^{(2)} \in U, \quad \alpha \in [0;1]$$

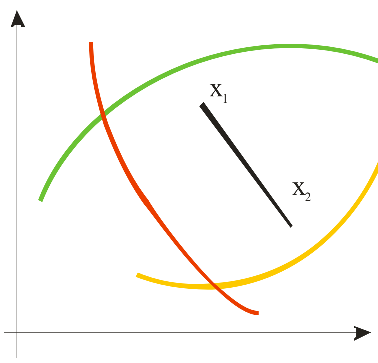


Рис. 2.2. Пример выпуклого множества

Функция $f(x)$ *является* ***выпуклой*** *на выпуклом множестве* $U \subseteq E^n$*, если для любых двух точек* $x^{(1)},\ x^{(2)} \in U$ *выполняется неравенство Иенсена:*

$$f\left(\alpha x^{(1)} + (1-\alpha)x^{(2)}\right) \le \alpha f\left(x^{(1)}\right) + (1-\alpha) f\left(x^{(2)}\right) \qquad 0 < \alpha < 1$$

Для дифференцируемых функций удобней использовать другое определение выпуклой функции.

Функция $f(x)$ *является* ***выпуклой в точке*** x^**, если матрица ее вторых частных производных (матрица Гессе)*

$$H\left(x^*\right) = \begin{pmatrix} \dfrac{\partial^2 f\left(x^*\right)}{\partial x_1 \partial x_1} & \dfrac{\partial^2 f\left(x^*\right)}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f\left(x^*\right)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f\left(x^*\right)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f\left(x^*\right)}{\partial x_2 \partial x_2} & \cdots & \dfrac{\partial^2 f\left(x^*\right)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f\left(x^*\right)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f\left(x^*\right)}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f\left(x^*\right)}{\partial x_n \partial x_n} \end{pmatrix}$$

положительно определена.

Для функции одной переменной это означает, что функция $f(x)$ лежит ниже хорды, соединяющей любые две точки ее графика.

Функцией Лагранжа *задачи выпуклого программирования называется функция*

$$L(x_1, x_2, ..., x_n, y_1, y_2, ..., y_m) = f(x_1, x_2, ..., x_n) + \sum_{i=1}^{m} y_i \cdot \left[b_i - g_i(x_1, x_2, ..., x_n)\right]$$

где $y_1, y_2, ..., y_m$ — ***множители Лагранжа***.

Точка $(X_0; Y_0)^{\mathrm{T}} = (x_1^0, x_2^0, ..., x_n^0; y_1^0, y_2^0, ..., y_m^0)$ *называется* ***седловой точкой*** *функции Лагранжа, если*

$$L(x_1, ..., x_n, y_1^0, ..., y_m^0) \le L(x_1^0, ..., x_n^0; y_1^0, ..., y_m^0) \le L(x_1^0, ..., x_n^0; y_1, ..., y_m) \qquad (2.9)$$

для всех $x_j \ge 0\ \ (j = 1, ..., n)$ и $y_i \ge 0\ \ (i = 1, ..., m)$.

Говорят, что выполняется ***условие регулярности,*** *если существует, по крайней мере, одна точка* $X_0^{\mathrm{T}} = (x_1^0, x_2^0, ..., x_n^0)$ *, для которой* $g_i(X_0^{\mathrm{T}}) > 0$ *для всех i.*

ТЕОРЕМА (*теорема Куна — Таккера*). *Для задачи выпуклого программирования, множество допустимых решений которой обладает свойством регулярности,* $X_0 = (x_1^0, x_2^0, ..., x_n^0)$ *является оптимальным планом для* (3.8) *тогда и только тогда, когда существует такой вектор* $Y_0 = (y_1^0, y_2^0, ..., y_m^0),\ (y_i^0 \geq 0, i = 1, ..., m)$, *что* $(X_0; Y_0)^{\mathrm{T}}$ *— седловая точка функции Лагранжа.*

Если предположить, что целевая функция f и функции g_i непрерывно дифференцируемы, то теорема Куна — Таккера может быть дополнена аналитическими выражениями, определяющими необходимые и достаточные условия того, чтобы точка $(X_0; Y_0)$ была седловой точкой функции Лагранжа, т. е. являлась решением задачи выпуклого программирования. Эти выражения имеют следующий вид (условия для задачи на максимум):

$$\frac{\partial L_0}{\partial x_j} \leq 0 \qquad (j = 1, ..., n),$$

$$x_j^{0\,\mathrm{T}} \frac{\partial L_0}{\partial x_j} = 0 \quad (j = 1, ..., n),$$

$$x_j^0 \geq 0 \qquad (j = 1, ..., n),$$

$$\frac{\partial L_0}{\partial y_i} \geq 0 \quad (i = 1, ..., m), \tag{2.10}$$

$$y_i^{0\,\mathrm{T}} \frac{\partial L_0}{\partial y_i} = 0 \quad (i = 1, \ldots, m),$$

$$y_i^0 \geq 0 \quad (i = 1, \ldots, m),$$

где $\frac{\partial L_0}{\partial x_j}$ и $\frac{\partial L_0}{\partial y_i}$ — значения соответствующих частных производных функций Лагранжа, вычисленных в седловой точке.

Итак, процесс нахождения решения задачи выпуклого программирования включает следующие этапы:

— проверка на принадлежность задачи выпуклому программированию;
— составляют функцию Лагранжа;
— записывают необходимые и достаточные условия существования седловой точки для функции Лагранжа;
— находят координаты седловой точки функции Лагранжа (проверяя, будет ли найденная точка являться точкой максимума), либо устанавливают ее отсутствие;
— записывают оптимальное решение и находят значение целевой функции.

Пример. *Найти* $\min\left(x_1^2 - x_2\right)$ *при ограничениях*

$$x_1 \geq 1,\ x_1^2 + x_2^2 \leq 10,\ x_2 \geq 0 \text{ (рис. 2.3).}$$

Решение. Графическое решение задачи представлено на рис. 2.3.

Рассмотрим теперь аналитическое решение. Составим функцию Лагранжа.

$$g_1(x) = 1 - x_1,$$
$$g_2(x) = -10 + x_1^2 + x_2^2,$$
$$g_3(x) = -x_2,$$
$$L(x, \lambda) = x_1^2 - x_2 + \lambda_1(1 - x_1) + \lambda_2\left(-10 + x_1^2 + x_2^2\right) - \lambda_3 x_2.$$

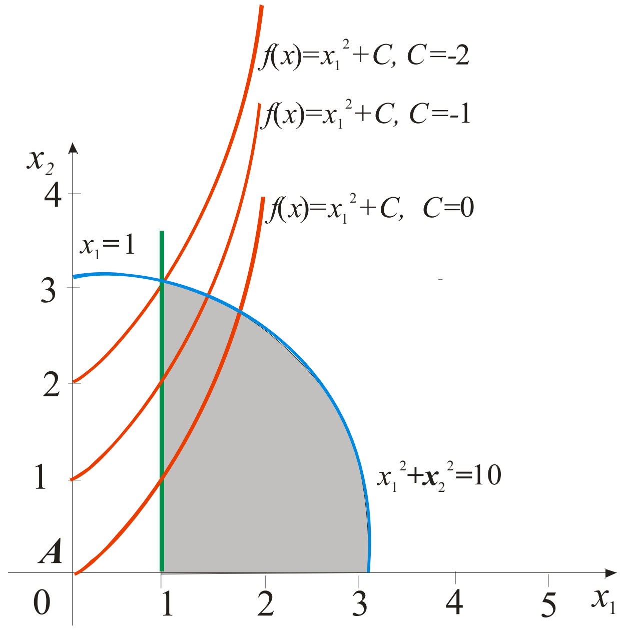


Рис. 2.3. Графическое решение задачи

Выписываем систему необходимых и достаточных условий существования седловой точки для функции Лагранжа.

$$\frac{\partial L_0}{\partial x_j} \geq 0 : \begin{cases} 2x_1 + 2\lambda_2 x_1 - \lambda_1 \geq 0 \\ -1 + 2\lambda_2 x_2 - \lambda_3 \geq 0 \end{cases},$$

$$\frac{\partial L_0}{\partial \lambda_i} \leq 0 : \begin{cases} 1 - x_1 \leq 0 \\ -10 + x_1^2 + x_2^2 \leq 0 \\ -x_2 \leq 0 \end{cases},$$

$$x_j^0 \frac{\partial L_0}{\partial x_j} = 0 : \begin{cases} x_1(2x_1 + 2\lambda_2 x_1 - \lambda_1) = 0 \\ x_2(-1 + 2\lambda_2 x_2 - \lambda_3) = 0 \end{cases},$$

$$\lambda_i^0 \frac{\partial L_0}{\partial \lambda_i} = 0 : \begin{cases} \lambda_1(1 - x_1) = 0 \\ \lambda_2(-10 + x_1^2 + x_2^2) = 0 \\ \lambda_3 x_2 = 0 \end{cases},$$

$$x_j^0 \geq 0\,, \quad \lambda_i^0 \geq 0 .$$

Предположим, что $1 - x_1=0$ — активное ограничение, т. е. оптимальное решение лежит на этой прямой. Тогда $\lambda_1 \neq 0$. Подставив $x_1 = 1$ в систему, получим:

$$\begin{aligned}
&2+2\lambda_2-\lambda_1 \geq 0,\\
&-1+2\lambda_2 x_2-\lambda_3 \geq 0,\\
&-10+1+x_2^2 \leq 0,\\
&-x_2 \leq 0,
\end{aligned}$$

$$\begin{aligned}
&2+2\lambda_2-\lambda_1=0,\\
&x_2\left(-1+2\lambda_2 x_2-\lambda_3\right)=0,\\
&\lambda_2\left(-10+1+x_2^2\right)=0,\\
&\lambda_3 x_2=0, \qquad\qquad\qquad (*)\\
&\lambda_1 \neq 0, \quad \lambda_2 \geq 0, \quad \lambda_3 \geq 0,\\
&x_1=1, \quad x_2 \geq 0.
\end{aligned}$$

Если предположить, что $x_2 = 0$, а $\lambda_3 \neq 0$ (в системе это условие помечено звездочкой — *), то система упростится еще:

$$\begin{aligned}
&2-\lambda_1 \geq 0 \qquad \Rightarrow \lambda_1 \leq 2,\\
&-1-\lambda_3 \geq 0 \quad\ \Rightarrow \lambda_3 \leq -1, \qquad (*)\\
&2-\lambda_1=0, \qquad \Rightarrow \lambda_1=2\\
&\lambda_1 \neq 0, \quad \lambda_2=0, \quad \lambda_3 \neq 0,\\
&x_1=1, \quad x_2=0.
\end{aligned}$$

Но в этом случае нарушается условие неотрицательности для множителей $\lambda_i \geq 0$, т.к. $\lambda_3 \leq -1$ (условие со звездочкой). Следовательно, предположение, что $x_2 = 0$ не верно. Пусть теперь $x_2 \neq 0$, т. е. данное ограничение пассивно. Тогда $\lambda_3 = 0$.

$$\begin{aligned}
&2+2\lambda_2-\lambda_1 \geq 0,\\
&-1+2\lambda_2 x_2 \geq 0,\\
&-10+1+x_2^2 \leq 0,\\
&2+2\lambda_2-\lambda_1=0,\\
&-1+2\lambda_2 x_2=0,\\
&\lambda_2\left(-9+x_2^2\right)=0, \qquad\qquad (*)\\
&\lambda_1 \neq 0, \quad \lambda_2 \geq 0, \quad \lambda_3 \neq 0,\\
&x_1=1, \quad x_2>0.
\end{aligned}$$

Предположим, что $x_2 = 3$, а $\lambda_2 = 0$, тогда решением системы будет

$$\begin{aligned} &\lambda_1 \neq 0, \quad \lambda_2 = 0, \quad \lambda_3 \neq 0, \\ &x_1 = 1, \\ &x_2 = 3. \end{aligned}$$

Полученное решение не противоречит условиям теоремы и, следовательно, является искомым, т. к. в задачах выпуклого программирования решение единственно. Ответ: $x^{*\text{т}} = (1, 3)$, $f^*(x) = -2$.

Ограничения типа «неравенств»

В общем виде задача условной оптимизации имеет вид:

$$\begin{aligned} &\max_{x \in U} f(x) \ \text{ или } \ \min_{x \in U} f(x), \\ &\textit{при условии } U = \left\{ x \in E^n \,\middle|\, g(x) \geq 0,\ h(x) = 0 \right\}, \end{aligned} \tag{2.11}$$

где $g(x) = (g_1(x), \cdots, g_m(x))$ — m-мерная вектор-функция,

$h(x) = (h_1(x), \cdots, h_k(x))$ — k-мерная вектор-функция, $m,\ k < n$.

Задача состоит в нахождении точки x^*, которая будет доставлять максимум (минимум) функции $f(x)$ на множестве точек U.

Рассмотрим вначале случай ограничений «типа неравенств». Тогда общий вид задачи условной оптимизации с ограничениями смешанного типа (2.11) можно переформулировать:

$$\max_{x \in U} f(x) \ \left(\min_{x \in U} f(x) \right), \text{ при условии } U = \left\{ x \in E^n \,\middle|\, g(x) \leq 0 \right\}. \tag{2.12}$$

Ограничения $g_s(x) \geq 0$ $\left(g_s(x) \leq 0 \right)$ из (3.10) и (3.12) для некоторого значения индекса $s,\ 1 \leq s \leq m$ в допустимой точке $\bar{x} \in U$ называют ***активными,*** *если $g_s(\bar{x}) = 0$, в противном случае ограничения называют* ***пассивными.***

ТЕОРЕМА (*Фрица — Джона*). *Пусть функции* $f(x)$, $g_i(x)$, $i = 1,2,...m$ *обладают непрерывными частными производными на некотором множестве* $U \subset E^n$, *содержащим точку* $x^* \in U$. *Если* x^* *— точка минимума для f (x) при условии* $g_i(x) \le 0 \quad i = 1,2,...m$, *то существуют* $\lambda_i \quad i = 1,2,...m$, *не все равные нулю, удовлетворяющие следующим соотношениям:*

$$\begin{gathered} \lambda_0 \nabla f\left(x^*\right) + \sum_{i=1}^{m} \lambda_i \nabla g_i\left(x^*\right) = 0, \\ \lambda_i g_i\left(x^*\right) = 0, \quad \lambda \ge 0. \end{gathered} \tag{2.13}$$

Ограничения «смешанного типа»

Необходимые условия оптимальности для задачи условной оптимизации с ограничениями смешанного типа представимы в виде:

$$\begin{gathered} \lambda_0 \nabla f\left(x^*\right) + \sum_{i=1}^{m} \lambda_i \nabla g_i\left(x^*\right) + \sum_{j=1}^{k} \lambda^+{}_i \nabla h_j\left(x^*\right) + \sum_{j=1}^{k} \lambda^-{}_i \nabla h_j\left(x^*\right) = 0, \\ \begin{pmatrix} \lambda_0 \\ \lambda^{\pm} \end{pmatrix} \ge 0, \quad \lambda \ne 0. \end{gathered} \tag{2.14}$$

или

$$\begin{gathered} \lambda_0 \nabla f\left(x^*\right) + \sum_{i=1}^{m} \lambda_i \nabla g_i\left(x^*\right) + \sum_{j=1}^{k} v_j \nabla h_j\left(x^*\right) = 0, \\ \lambda_i g_i\left(x^*\right) = 0, \quad h_j\left(x\right) = 0, \\ \begin{pmatrix} \lambda_0 \\ \lambda \end{pmatrix} \ge 0, \quad \lambda \ne 0, \quad v \text{ — любого знака.} \end{gathered} \tag{2.15}$$

Пример. *Найти* $\min\left(x_1^2 - x_2\right)$ *при ограничениях*

$$x_1 \ge 1,\ x_1^2 + x_2 \le 5,\ \ x_1 + x_2 = 3 \ \textit{(рис. 3.4)}.$$

Решение. Формализуем условия задачи:

$$g_1(x) = 1 - x_1, \quad g_2(x) = x_1^2 + x_2 - 5, \quad h(x) = x_1 + x_2 - 3.$$

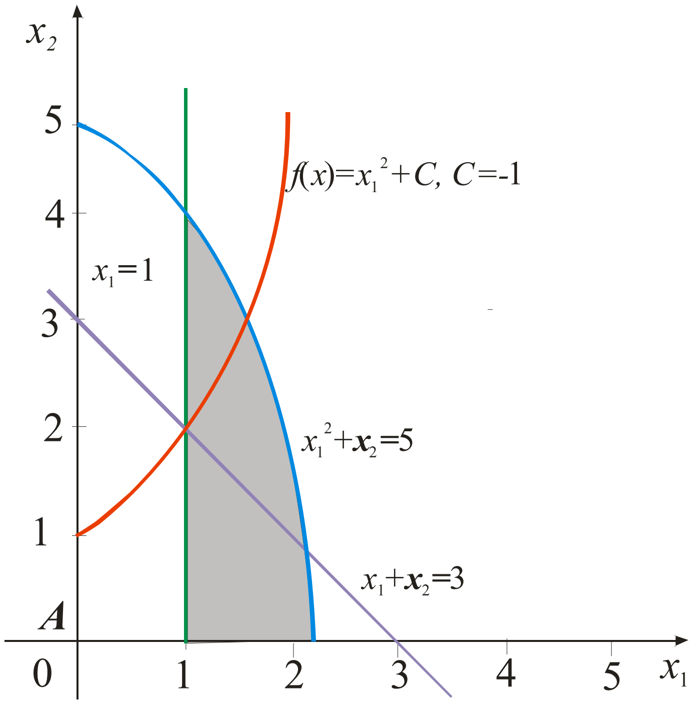


Рис. 2.4. Графическое решение задачи

Выписываем выражения для градиентов

$$\nabla f(x)=\begin{pmatrix}2x_1\\-1\end{pmatrix},\ \nabla h(x)=\begin{pmatrix}1\\1\end{pmatrix},\ \nabla g_1(x)=\begin{pmatrix}-1\\0\end{pmatrix},\ \nabla g_2(x)=\begin{pmatrix}2x_1\\1\end{pmatrix}.$$

Тогда система соотношений для нахождения решения по теореме Фрица — Джона будет иметь вид:

$$\lambda_0\begin{pmatrix}2x_1\\-1\end{pmatrix}+\lambda_1\begin{pmatrix}-1\\0\end{pmatrix}+\lambda_2\begin{pmatrix}2x_1\\1\end{pmatrix}+v\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

или

$$\begin{cases}\lambda_0 2x_1+\lambda_1\cdot(-1)+\lambda_2\cdot 2x_1+v\cdot 1=0\\ \lambda_0\cdot(-1)+\lambda_1\cdot 0+\lambda_2\cdot 1+v\cdot 1=0\end{cases}$$

После преобразования система равенств и неравенств выглядит следующим образом:

$$\begin{cases} \lambda_1\left(1 - x_1\right) = 0 \\ \lambda_2\left(x_1^2 + x_2 - 5\right) = 0 \\ x_1 + x_2 = 3 \\ 1 - x_1 \le 0 \\ x_1^2 + x_2 - 5 \le 0 \\ \lambda_0 \ge 0,\ \lambda_1 \ge 0,\ \lambda_2 \ge 0. \end{cases}$$

Решение проводится аналогично предыдущему примеру.

Ответ: $x^{*T} = (1, 2)$, $f^*(x) = -1$.

Лекция 2.1.3. Динамические модели в исследовании операций. Принцип оптимальности Беллмана для решения динамических задач. Сетевые модели и рекуррентные соотношения как основа методов нахождения решения динамических задач

Модели этой группы применяются при разработке правил управления запасами, устанавливающими момент пополнения запасов и размер пополняющего заказа; при разработке принципов календарного планирования производства и выравнивании занятости в условиях колеблющегося спроса на продукцию; при распределении дефицитных капитальных вложении между возможными направлениями их использования; при составлении календарных планов текущего и капитального ремонта оборудования и его замены.

Динамическое программирование представляет собой математический аппарат, разработанный с целью повышения эффективности вычислений при решении некоторого класса задач математического программирования путем их разложения (декомпозиции) на относительно небольшие и, следовательно, менее сложные подзадачи. Для динамического программирования характерен подход к решению задач по этапам, с каждым из которых ассоциирована одна управляемая переменная. Набор **рекуррентных вычислительных процедур**, связывающих различные этапы, обеспечивает получение допустимого, оптимального решения задачи *в целом* при достижении последнего этапа.

Происхождение названия **динамическое программирование** *(ДП)* связано с использованием методов ДП в задачах принятия решений через фиксированные промежутки времени (например, в задачах управления запасами). Однако методы решения ДП успешно применяются также для решения задач, в которых фактор времени не учитывается. По этой причине более удачным представляется термин **многоэтапное программирование**, отражающий пошаговый характер процесса решения задач.

Фундаментальным принципом, положенным в основу теории ДП, является **принцип оптимальности.** По существу, он определяет порядок поэтапного решения допускающей декомпозицию задачи (это более приемлемый путь, чем непосредственное решение задачи в исходной постановке) с помощью рекуррентных вычислительных процедур.

Приведем ниже одну из типовых ситуаций, принятие решения в которой связано с использованием математического аппарата ДП.

Задача распределения капиталовложений

Совет директоров фирмы рассматривает предложения по наращиванию производственных мощностей на трех принадлежащих фирме предприятиях. Для расширения всех трех предприятий выделены средства в объеме 5 млн долл. Каждое предприятие представляет проекты, которые характеризуются величинами (в миллионах долларов) суммарных затрат c и доходов R, связанных с реализацией каждого проекта. Соответствующие данные приведены в табл. 2.1, в которую включены также проекты с нулевыми затратами. Это позволяет учесть возможность отказа от расширения какого-либо предприятия. Цель фирмы состоит в получении максимального дохода от инвестиций в объеме 5 млн долл.

Таблица 2.1

Проект	Предприятие 1		Предприятие 2		Предприятие 3	
	c_1	R_1	c_2	R_2	c_3	R_3
1	0	0	0	0	0	0
2	1	5	2	8	1	3
3	2	6	3	9	—	—
4	—	—	4	12	—	—

Модель динамического программирования

Необходимо внести ясность в вопрос о структуре **рекуррентных** вычислений в ДП. Расчеты на некотором этапе осуществляются на базе сводной информации о максимальном суммарном доходе (самом длинном пути), полученной в результате *всех* предшествующих этапов. При этом все отдельные решения, найденные на предшествующих этапах, не представляют существенного интереса. Действительно, все последующие решения строятся некоторым *оптимальным* образом независимо от решений, полученных на предшествующих этапах. Это отражает сущность **принципа оптимальности**, составляющего основу вычислительной схемы ДП (принцип оптимальности Беллмана).

При формировании динамической модели необходимо определить **этапы** решения задачи. В задаче распределения капиталовложений каждому из предприятий поставим в соответствие некоторый этап, поскольку требуется выбрать оптимальный проект для каждого предприятия. Этапы связаны между собой посредством ограничения на суммарный объем капиталовложений. При построении модели необходимо учесть эту связь таким образом, чтобы получить возможность по отдельности решать подзадачи, соответствующие каждому этапу, не нарушая при этом условия допустимости. Введем следующие определения:

x_1 — объем капиталовложений, распределенных на этапе 1;

x_2 — объем капиталовложений, распределенных на этапах 1 и 2;

x_3 — объем капиталовложений, распределенных на этапах 1, 2 и 3.

Наиболее важным при построении динамической модели является способ определения величин x_j, который позволяет автоматически выводить из рассмотрения недоступные проекты (дуги). В ДП используется аналогичный прием, приводящий к понятию **состояния** системы. Этот термин не представляется бесспорно удачным, поскольку он описывает «состояние» системы лишь с позиций наличия ограниченного количества некоторого ресурса, что обеспечивает взаимную увязку всех этапов.

Рекуррентное соотношение ДП можно записать в следующем виде:

$$f_j(x_j) = \max_{\substack{\text{по допустимым} \\ \text{проектам } k}} \{R_j(k_j) + f_{j-1}(x_{j-1})\}, \quad j = 1,\ 2,\ 3.$$

Здесь $R_j(k_j)$ — доход, полученный на j-м этапе в результате реализации проекта k_j при вложении средств в размере x_j, $f_{j-1}(x_{j-1})$ — максимальный доход на (j-1)-м этапе после вложения средств в размере x_{j-1}.

При этом $f_0(x_0) = 0$ по определению. Приведенное равенство действительно является рекуррентным, так как величина $f_j(x_j)$ на этапе j вычисляется по известному значению $f_{j-1}(x_{j-1})$ на этапе j -1 при $f_0(x_0) = 0$.

Чтобы представить рекуррентное соотношение в корректной математической форме, необходимо сделать два замечания, которые к тому же помогут устранить кажущиеся различия между моделью ДП и сетевой моделью для рассматриваемой задачи.

Во-первых, заметим, что $f_j(x_j)$ — функция единственного аргумента. Отсюда следует, что правая часть рекуррентного соотношения должна быть выражена через x_j, а не через x_{j-1}. Разность между x_j и x_{j-1} равна величине затрат (c_j) на реализацию проекта k_j на этапе j, т. е. $c_j(k_j) = x_j - x_{j-1}$.

Теперь можно выразить x_{j-1} через x_j с помощью равенства

$$x_{j-1} = x_j - c_j(k_j).$$

Такая замена обеспечивает более корректную математическую запись рекуррентного соотношения.

Во-вторых, необходимо представить в математической форме условие, предписывающее рассмотрение только допустимых проектов. В этом случае можно также воспользоваться равенством $c_j(k_j) = x_j - x_{j-1}$. Однако это ограничение уже введено выше путем замены $x_{j-1} = x_j - c_j(k_j)$ в функции $f_{j-1}(x_{j-1})$. Чтобы обеспечить полную корректность, следует учесть не требующее пояснений неравенство $x_{j-1} \geq 0$, откуда $x_j - c_j(k_j) \geq 0$ или $x_j \geq c_j(k_j)$.

Таким образом, рекуррентное соотношение ДП имеет следующий вид:

$$\begin{aligned} &f_0(x_0) = 0 \\ &f_j(x_j) = \max_{c_j(k_j) \leq x_j} \left\{ R_j(k_j) + f_{j-1}\left(x_j - c_j(k_j)\right) \right\}, \; j = 1,\ 2,\ 3. \end{aligned}$$

Ниже приводятся результаты поэтапных расчетов на основе рекуррентного соотношения для рассматриваемой задачи. Они представлены в виде таблиц, так как в ДП обычно используется табличная форма записи числовых результатов.

Этап 1

$$f_1(x_1)=\max_{\substack{c_1(k_1)\le x_1\\ k_1=1,2,3}}\{R_1(k_1)\}.$$

Таблица 2.2

x_1	$R_1(k_1)$			Оптимальное решение	
	$k_1=1$	$k_1=2$	$k_1=3$	$f_1(x)_1$	k_1*
0	0	—	—	0	1
1	0	5	—	5	2
2	0	5	6	6	3
3	0	5	6	6	3
4	0	5	6	6	3
5	0	5	6	6	3

Этап 2

$$f_2(x_2)=\max_{\substack{c_2(k_2)\le x_2\\ k_2=1,2,3,4}}\{R_2(k_2)+f_1(x_2-c_2(k_2))\}$$

Таблица 2.3

x_2	$\{R_2(k_2)+f_1(x_2-c_2(k_2))\}$				Оптимальное решение	
	$k_2=1$	$k_2=2$	$k_2=3$	$k_2=4$	$f_2(x)_2$	k_2*
0	0+0 = 0	—	—	—	0	1
1	0+5 = 5	—	—	—	5	1
2	0+6 = 6	8+0 = 8	—	—	8	2
3	0+6 = 6	8+5 = 13	9+0 = 9	—	13	2
4	0+6 = 6	8+6 = 14	9+5 = 14	12+0 = 12	14	2 или 3
5	0+6 = 6	8+6 = 14	9+6 = 15	12+5 = 17	17	4

Этап 3

$$f_2(x_2) = \max_{\substack{c_3(k_3) \le x_3 \\ k_3 = 1,2.}} \{R_3(k_3) + f_2(x_3 - c_3(k_3))\}$$

Таблица 2.4

x_3	$\{R_3(k_3) + f_2(x_3 - c_3(k_3))\}$		Оптимальное решение	
	$k_3 = 1$	$k_3 = 2$	$f_3(x)_3$	k_3*
5	0 + 17 = 17	3 + 14 = 17	17	1 или 2

Оптимальное решение можно найти непосредственно из приведенных выше таблиц 2.2—2.4. Сначала рассматривается таблица, построенная на этапе 3. При x_3 = 5 оптимальный проект имеет либо $k_3^* = 1$, либо $k_3^* = 2$. Пусть сначала $k_3^* = 1$.Так как $c_3(1) = 0$, на этапах 2 и 1 $x_2 = x_3 - c_3(1) = 5$. Легко видеть, что на этапе 2 из x_2 = 5 следует $k_3^* = 4$. Далее $c_3(4) = 4$, откуда x_1= 5 – 4 = 1. На этапе 1 из x_1 = 1 вытекает $k_1^* = 2$. Таким образом, (2, 4, 1) есть оптимальный набор проектов для предприятий.

Рекуррентное соотношение для процедуры обратной прогонки

В предыдущем разделе вычисления проводили в соответствии с последовательностью $f_1 \to f_2 \to f_3$.

Такой метод вычислений известен как **алгоритм прямой прогонки,** поскольку расчеты осуществляются в естественном порядке следования этапов. Однако в специальной литературе, посвященной динамическому программированию, рассматривается рекуррентное соотношение, которое предписывает начинать вычисления с последнего этапа и затем «двигаться» назад до этапа 1. Такой метод вычислений известен как **метод обратной прогонки**.

Основное различие между процедурами прямой и обратной прогонки заключается в способе определения состояния системы. С целью пояснения опять обратимся к задаче распределения капиталовложений. Для процедуры обратной прогонки определим состояния y_j следующим образом:

y_1 — объем капиталовложений, распределенных на этапах 1, 2, 3.

y_2 — объем капиталовложений, распределенных на этапах 2, 3.

y_3 — объем капиталовложений, распределенных на этапе 3.

Положим

$f_3(y_3)$ — максимальный доход на этапе 3 при заданном y_3;

$f_2(y_2)$ — максимальный доход на этапах 2 и 3 при заданном y_2;

$f_1(y_1)$ — максимальный доход на этапах 1, 2 и 3 при заданном y_1.

Следуя аналогичным рассуждениям, проведенным в процедуре прямой прогонки $f_4(y_4)=0$.

Рекуррентное соотношение для процедуры обратной прогонки записывается в следующем виде:

$$f_j(y_j)=\max_{c_j(k_j)\le y_j}\{R_j(k_j)+f_{j+1}(y_j-c_j(k_j))\},\ j=1,\ 2,\ 3.$$

Порядок поэтапных вычислений определяется последовательностью $f_3 \to f_2 \to f_1$. Ниже приводятся результаты расчетов.

Этап 3

$$f_3(y_3)=\max_{\substack{c_3(k_3)\le y_3\\ k_3=1,2}}\{R_3(k_3)\}.$$

Таблица 2.5

y_3	$R_3(k_3)$		Оптимальные решения	
	$k_3=1$	$k_3=2$	$f_3(y_3)$	k_3^*
0	0	—	0	1
1	0	3	3	2
2	0	3	3	2
3	0	3	3	2
4	0	3	3	2
5	0	3	3	2

Этап 2

$$f_2(y_2) = \max_{\substack{c_2(k_2) \le y_2 \\ k_2 = 1,2,3,4}} \{R_2(k_2) + f_3(y_2 - c_2(k_2))\}$$

Таблица 2.6

y_2	$\{R_2(k_2) + f_3(y_2 - c_2(k_2))\}$				Оптимальное решение	
	$k_2 = 1$	$k_2 = 2$	$k_2 = 3$	$k_2 = 4$	$f_2(y_2)$	k_2^*
0	0+0 = 0	—	—	—	0	1
1	0+3 = 3	—	—	—	3	1
2	0+3 = 3	8+0 = 8	—	—	8	2
3	0+3 = 3	8+3 = 11	9+0 = 9	—	11	2
4	0+3 = 3	8+3 = 11	9+3 = 12	12+0 = 12	12	3 или 4
5	0+3 = 3	8+3 = 11	9+3 = 12	12+3 = 15	15	4

Этап 1

$$f_1(y_1) = \max_{\substack{c_1(k_1) \le y_1 \\ k_1 = 1,2,3}} \{R_1(k_1) + f_2(y_1 - c_1(k_1))\}$$

Таблица 2.7

y_1	$\{R_1(k_1) + f_2(y_1 - c_1(k_1))\}$			Оптимальное решение	
	$k_1 = 1$	$k_1 = 2$	$k_1 = 3$	$f_1(y_1)$	k_1^*
5	0+15 = 15	5+12 = 17	6+11 = 17	17	2 или 3

В качестве еще одной ситуации, для принятия решения в которой используется аппарат ДП, можно привести следующую задачу о загрузке.

Задача о загрузке

Самолет загружается предметами N различных типов. Каждый предмет типа j имеет вес w_j и стоимость v_j (j=1,2,...,N). Максимальная грузоподъемность самолета равна W. Требуется определить максимальную стоимость груза, вес которого не должен превышать максимальную грузоподъемность самолета. Предположим, что $W = 5$ и всего имеются три типа предметов, числовые сведения о которых приведены в таблице 2.8.

Таблица 2.8

j	w_j	v_j
1	2	65
2	3	80
3	1	30

Сначала рассмотрим задачу в общей постановке. Если обозначить количество предметов типа j через k_j, то задача принимает следующий вид: максимизировать $v_1 k_1+v_2 k_2+\ldots+ v_N k_N$ при ограничениях

$$w_1 k_1 + w_2 k_2 +\ldots+ w_N k_N \le W,$$

где k_j — неотрицательные числа.

Если отбросить требование целочисленности k_j, то решение задачи нетрудно найти с помощью симплекс-метода. В самом деле, так как остается лишь одно ограничение, базисной будет только одна переменная, и задача сводится к выбору типа j, для которого величина $v_j = \frac{W}{w_j}$ принимает максимальное значение. Исходная задача не является задачей линейного программирования, мы попытаемся использовать для её решения методы динамического программирования. Следует отметить, что рассматриваемая задача может быть также решена с помощью методов целочисленного программирования.

Каждый из трех основных элементов модели ДП определяется следующим образом.

1. *Этап* j ставится в соответствие типу j, j = 1, 2, . . . , N.

2. *Состояние* y_j на этапе j выражает суммарный вес предметов, решение о погрузке которых принято на этапах $j, j+1, \dots, N$; при этом $y_1=W$ и $y_j = 0, 1, \dots, W$ при $j = 2, 3, \dots, N$.
3. Варианты решения k_j на этапе j описываются количеством предметов типа j. Значение k_j заключено в пределах от нуля до $\left[\frac{W}{w_j}\right]$, где $\left[\frac{W}{w_j}\right]$ — целая часть числа (W/w_j).

Данная задача имеет несомненное сходство с задачей распределения капиталовложений и также относится к классу задач распределения ресурсов. Единственное различие состоит в том, что в задаче о загрузке структура вариантов решения несколько сложнее, чем в задаче распределения капиталовложений.

Пусть $f_j(y_j)$ — максимальная суммарная стоимость предметов, решения о погрузке которых приняты на этапах j, $j = 1, \dots, N$ при заданном состоянии y_j.

Рекуррентное соотношение (для процедуры обратной прогонки) имеет следующий вид:

$$f_N(y_N) = \max_{\substack{k_N = 0,1,\dots,\left[\frac{y_N}{w_N}\right] \\ y_N = 0,1,\dots W}} \{v_N k_n\}$$

$$f_j(y_j) = \max_{\substack{k_N = 0,1,\dots,\left[\frac{y_j}{w_j}\right] \\ y_j = 0,1,\dots W}} \{v_j k_j + f_{j+1}(y_i - w_j(k_j))\}, \quad j = 1,2,\dots N-1$$

Заметим, что максимальное *допустимое* значение k_j ограничено величиной $[y_j / w_j]$. Это позволяет автоматически исключать все не являющиеся допустимыми варианты при заданном значении переменной состояния y_j.

Для приведенного выше численного примера поэтапные расчеты осуществляются следующим образом.

Этап 3

$$f_3(y_3) = \max_{k_3}\{30k_3\}, \quad \max k_3 = [5/1] = 5$$

Таблица 2.9

y_3	$30\,k_3$						Оптимальные решения	
	$k_3 = 0$	1	2	3	4	5	$f_3(y_3)$	$k_3{}^*$
	$v_3\,k_3 = 0$	30	60	90	120	150		
0	0	—	—	—	—	—	0	0
1	0	30	—	—	—	—	30	1
2	0	30	60	—	—	—	60	2
3	0	30	60	90	—	—	90	3
4	0	30	60	90	120	—	120	4
5	0	30	60	90	120	150	150	5

Этап 2

$$f_2(y_2) = \max_{k_2}\{80k_2 + f_3(y_2 - 3k_2)\}, \quad \max k_2 = [5/3] = 1$$

Таблица 2.10

y_2	$\{80\,k_2 + f_3(y_2 - 3k_2)\}$		Оптимальные решения	
	$k_2 = 0$	$k_2 = 1$	$f_2(y_2)$	$k_2{}^*$
	$v_2 k_2 = 0$	$v_2 k_2 = 1$		
0	0+0 = 0	—	0	0
1	0+30 = 30	—	30	0
2	0+60 = 60	—	60	0
3	0+90 = 90	80+0 = 80	90	0
4	0+120 = 120	80+30 = 110	120	0
5	0+ 150 = 150	80+60 = 140	150	0

Этап 1

$$f_1(y_1) = \max_{k_1}\{65k_1 + f_2(y_1 - 2k_1)\}, \quad \max k_1 = [5/2] = 1$$

Таблица 2.11

y_1	$\{65k_1 + f_2(y_1 - 2k_1)\}$			Оптимальные решения	
	k_1=0	k_1=1	k_1=2	$f_1(y_1)$	k_1*
	$v_1\,k_1$=0	$v_1\,k_1$=65	$v_1\,k_1$=130		
0	0+0 = 0	—	—	0	0
1	0+30 = 30	—	—	30	0
2	0+60 = 60	65+0 = 65	—	65	1
3	0+90 = 90	65+30 = 95	—	95	1
4	0+120 = 120	65+60 = 125	130+0 = 130	130	2
5	0+150 = 150	65+90 = 155	130+30 = 160	160	2

При заданном $y_1 = W = 5$ оптимальным решением является

$$(k_1^*, k_2^*, k_3^*) = (2,0,1),$$

а суммарная стоимость груза равна 160.

Заметим, что на этапе 1 достаточно построить только одну строку таблицы, соответствующую значению y_1 = 5. Однако, располагая полной таблицей для значений y_1= 0, 1, 2, 3, 4 и 5, можно исследовать изменения в оптимальном решении, которые вызываются уменьшением максимальной грузоподъемности W = 5, т. е. провести анализ чувствительности решения. Вычислительная схема ДП автоматически обеспечивает возможность проведения такого анализа.

ТЕМА 2.2. СПЕЦИАЛЬНЫЕ МОДЕЛИ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ В ЭКОНОМИКЕ

Лекция 2.2.1. Операционные модели транспортного типа. Методы их решения: нахождение опорного решения транспортной задачи

Общая постановка транспортной задачи состоит в определении оптимального плана перевозок некоторого однородного груза из m пунктов отправления $A_1, A_2, \ldots, A_m$ в n пунктов назначения $B_1, B_2, \ldots, B_n$. При этом в качестве критерия оптимальности обычно берется либо минимальная стоимость перевозок всего груза, либо минимальное время его доставки. Рассмотрим транспортную задачу, в качестве критерия оптимальности которой взята минимальная стоимость перевозок всего груза. Обозначим через c_{ij} тарифы перевозки единицы груза из i-го пункта отправления в j-й пункт назначения, через a_i — запасы груза в i-м пункте отправления, через b_j — потребности в грузе в j-м пункте назначения, а через x_{ij} — количество единиц груза, перевозимого из i-го пункта отправления в j-й пункт назначения. Тогда математическая постановка задачи состоит в определении минимального значения функции

$$F = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \to \min \tag{2.16}$$

при условиях

$$\sum_{i=1}^{m} x_{ij} = b_j \quad (j = \overline{1, n}), \tag{2.17}$$

$$\sum_{j=1}^{n} x_{ij} = a_i \quad (i = \overline{1, m}), \tag{2.18}$$

$$x_{ij} \geq 0 \qquad (i = \overline{1, m};\ j = \overline{1, n}). \tag{2.19}$$

Поскольку переменные $x_{ij}\ (i = \overline{1, m};\quad j = \overline{1, n})$ удовлетворяют системам линейных уравнений (2.17) и (2.18) и условию неотрицательности (2.19), обеспечивается доставка необходимого количества груза в каждый из пунктов назначения, вывоз имеющегося груза из всех пунктов отправления, а также исключаются обратные перевозки.

Всякое неотрицательное решение систем линейных уравнений (2.17) *и* (2.18), *определяемое матрицей* $X=(x_{ij})$ $(i=\overline{1,m};\ \ j=\overline{1,n})$, *называется* ***планом транспортной задачи***.

Будем называть любой план перевозок ***допустимым****, если он удовлетворяет условиям* (2.17), (2.18)*, т. е. все заявки удовлетворены, все запасы исчерпаны).*

План $X^*=(x^*_{ij})$ $(i=\overline{1,m};\ \ j=\overline{1,n})$, *при котором функция* (2.16) *принимает свое минимальное значение, называется* ***оптимальным планом транспортной задачи.***

Обычно исходные данные транспортной задачи записывают в виде таблицы (табл. 2.12).

Таблица 2.12

Пункты отправления	Пункты назначения					Запасы
	B_1	…	B_j	…	B_n	
A_1	x_{11} c_{11}	…	x_{1j} c_{1j}	…	x_{1n} c_{1n}	a_1
…	…	…	…	…	…	…
A_i	x_{i1} c_{i1}	…	x_{ij} c_{ij}	…	x_{in} c_{in}	a_i
…	…	…	…	…	…	…
A_m	x_{m1} c_{m1}	…	x_{mj} c_{mj}	…	x_{mn} c_{mn}	a_m
Потребности	b_1	…	b_j	…	b_n	

Очевидно, общее наличие груза у поставщиков равно $\sum_{i=1}^{m} a_i$, а общая потребность в грузе в пунктах назначения равна $\sum_{j=1}^{n} b_j$ единиц.

Если общая потребность в грузе в пунктах назначения равна запасу груза в пунктах отправления, т. е.

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j, \tag{2.20}$$

то модель такой транспортной задачи называется ***закрытой.*** *Если же указанное условие не выполняется, то модель транспортной задачи называется* ***открытой.***

ТЕОРЕМА. *Для разрешимости транспортной задачи необходимо и достаточно, чтобы запасы груза в пунктах отправления были равны потребностям в грузе в пунктах назначения, т. е. чтобы выполнялось равенство* (2.20).

В случае превышения запаса над потребностью, т. е. $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$, вводится фиктивный ($n$+1)-й пункт назначения с потребностью $b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$ и соответствующие тарифы считаются равными нулю: $c_{in+1} = 0 \quad (i = \overline{1, m})$. Полученная задача — транспортная задача, для которой выполняется равенство (2.20).

Аналогично, при $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$ вводится фиктивный (m+1)-й пункт отправления с запасом груза $a_{m+1} = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i$ и тарифы полагаются равными нулю: $c_{m+1\,j} = 0 \quad (j = \overline{1,n})$. Этим задача сводится к обычной транспортной задаче, из оптимального плана которой получается оптимальный план исходной задачи. В дальнейшем будем рассматривать закрытую модель транспортной задачи. Если же модель конкретной задачи является открытой, то, исходя из сказанного выше, перепишем таблицу условий задачи так, чтобы выполнялось равенство (2.20).

Число переменных x_{ij} в транспортной задаче с m пунктами отправления и n пунктами назначения равно nm, а число уравнений в системах (2.17) и (2.18) равно $n+m$. Так как мы предполагаем, что выполняется условие (2.20), то число линейно независимых уравнений равно $n+m-1$.

Допустимый план будем называть ***опорным****, если в нем отличны от нуля не более* $m+n-1$ *базисных перевозок, а остальные перевозки равны нулю.*

Если в опорном плане число отличных от нуля компонент равно в точности $n+m$-1*, то план является* ***невырожденным****, а если меньше —* ***вырожденным****.*

Для определения опорного допустимого плана существует несколько методов. Ниже рассматриваются метод северо-западного угла и метод минимального элемента.

Как и для всякой задачи линейного программирования, оптимальный план транспортной задачи будет опорным и допустимым планом.

Ввиду исключительной практической важности транспортной задачи и специфики ее ограничений (каждая неизвестная входит лишь в два уравнения систем (2.17) и (2.18) и коэффициенты при неизвестных равны единице) для определения оптимального плана транспортной задачи разработаны специальные методы. Один из них — метод потенциалов — изложен ниже.

Определение опорного плана транспортной задачи

Как и при решении задачи линейного программирования симплексным методом, определение оптимального плана транспортной задачи начинают с нахождения какого-нибудь ее опорного плана. Этот план, как уже отмечалось, находят методом северо-западного угла, методом минимального элемента или методом аппроксимации Фогеля. Сущность этих методов состоит в том, что опорный план находят последовательно за $n+m-1$ шагов, на каждом из которых в таблице условий задачи заполняют одну клетку, которую называют **занятой.** Заполнение одной из клеток обеспечивает полностью либо удовлетворение потребности в грузе одного из пунктов назначения (того, в столбце которого находится заполненная клетка), либо вывоз груза из одного из пунктов отправления (из того, в строке которого находится заполняемая клетка).

В первом случае временно исключают из рассмотрения столбец, содержащий заполненную на данном шаге клетку, и рассматривают задачу, таблица условий которой содержит на один столбец меньше, чем было перед этим шагом, но то же количество строк и соответственно измененные запасы груза в одном из пунктов отправления (в том, за счет запаса которого была удовлетворена потребность в грузе пункта назначения на данном шаге). Во втором случае временно исключают из рассмотрения строку, содержащую заполненную клетку, и считают, что таблица условий имеет на одну строку меньше при неизменном количестве столбцов и при соответствующем изменении потребности в грузе в пункте назначения, в столбце которого находится заполняемая клетка.

После того как проделаны $m+n-2$ описанных выше шагов, получают задачу с одним пунктом отправления и одним пунктом назначения. При этом останется свободной только одна клетка, а запасы оставшегося пункта отправления будут равны потребностям оставшегося пункта назначения. Заполнив эту клетку, тем самым делают $(n+m-1)$-й шаг и получают искомый опорный план. Следует заметить, что на некотором шаге (но не на последнем) может оказаться, что потребности очередного пункта назначения равны запасам очередного пункта отправления. В этом случае также временно исключают из рассмотрения либо столбец, либо строку (т. е. что-нибудь одно).

Таким образом, либо запасы соответствующего пункта отправления, либо потребности данного пункта назначения считают равными нулю. Этот нуль записывают в очередную заполняемую клетку. Указанные выше условия гарантируют получение $n+m-1$ занятых клеток, в которых стоят компоненты опорного плана, что является исходным условием для проверки последнего на оптимальность и нахождения оптимального плана.

Метод северо-западного угла

При нахождении опорного плана транспортной задачи методом северо-западного угла на каждом шаге рассматривают первый из оставшихся пунктов отправления и первый из оставшихся пунктов назначения. Заполнение клеток таблицы условий начинается с левой верхней клетки для неизвестного x_{11} («северо-западный угол») и заканчивается клеткой для неизвестного x_{mn}, т. е. идет как бы по диагонали таблицы.

Пример. *На три базы A_1, A_2, A_3 поступил однородный груз в количествах, соответственно равных 140, 180 и 160 ед. Этот груз требуется перевезти в пять пунктов назначения* B_1, B_2, B_3, B_4, B_5 *соответственно в количествах 60, 70, 120, 130 и 100 ед. Тарифы перевозок единицы груза из каждого пункта отправления в соответствующие пункты назначения указаны в табл. 2.13.*

Требуется найти план перевозок данной транспортной задачи методом северо-западного угла.

Решение. Здесь число пунктов отправления $m = 3$, а число пунктов назначения $n=5$. Следовательно, опорный план задачи определяется числами, стоящими в 5+3–1=7 заполненных клетках.

Начнем заполнение транспортной таблицы с левого верхнего (северо-западного) угла, т.е. с клетки для неизвестного x_{11}, тем самым попытаемся удовлетворить потребности первого пункта назначения за счет запасов первого пункта отправления.

Таблица 2.13

Пункты отправления	Пункты назначения					Запасы
	B_1	B_2	B_3	B_4	B_5	
A_1	2	3	4	2	4	140
A_2	8	4	1	4	1	180
A_3	9	7	3	7	2	160
Потребности	60	70	120	130	100	480

Так как запасы пункта A_1 больше, чем потребности пункта B_1, то полагаем x_{11}= 60, записываем это значение в соответствующей клетке табл. 2.14 и временно исключаем из рассмотрения столбец B_1, считая при этом запасы пункта A_1 равными 140-60=80. Теперь в пункте A_1 осталось 80 единиц груза; этим количеством можно удовлетворить потребности пункта B_2. Положим x_{12} = 70, запишем это значение в соответствующей клетке табл. 2.14 и временно исключим из рассмотрения столбец B_2. После этого в A_1 остается еще 80 – 70=10 единиц груза; отдадим их пункту B_3. Потребности пункта B_3 больше оставшихся запасов пункта A_1. Положим x_{13}=10 и исключим из рассмотрения строку A_1. Значение x_{13}=10 запишем в соответствующую клетку табл. 2.14 и считаем потребности пункта B_3 равными 110 ед. Но заявка этого пункта еще не удовлетворена полностью; выделим недостающие 120–10=110 единиц из запасов других пунктов отправления.

Теперь перейдем к заполнению клетки для неизвестного x_{23} и т. д. Через шесть шагов остается один пункт отправления A_3 с запасом груза 100 ед. и один пункт назначения B_5 с потребностью 100 ед. Соответственно, имеется одна свободная клетка, которую и заполняем, полагая x_{35} =100 (табл. 2.15).

В результате получаем опорный план

$$X = \begin{pmatrix} 60 & 70 & 10 & 0 & 0 \\ 0 & 0 & 110 & 70 & 0 \\ 0 & 0 & 0 & 60 & 100 \end{pmatrix}.$$

Таблица 2.14

Пункты отправления	Пункты назначения					Запасы
	B_1	B_2	B_3	B_4	B_5	
A_1	60 [2]	70 [3]	10 [4]	[2]	[4]	140
A_2	[8]	[4]	[1]	[4]	[1]	180
A_3	[9]	[7]	[3]	[7]	[2]	160
Потребности	60	70	120	130	100	480

Таблица 2.15

Пункты отправления	Пункты назначения					Запасы
	B_1	B_2	B_3	B_4	B_5	
A_1	60 [2]	70 [3]	10 [4]	[2]	[4]	140
A_2	[8]	[4]	110 [1]	70 [4]	[1]	180
A_3	[9]	[7]	[3]	60 [7]	100 [2]	160
Потребности	60	70	120	130	100	480

Согласно данному плану перевозок, общая стоимость перевозок всего груза составляет

$$F=2\cdot 60 + 3\cdot 70 + 4\cdot 10+1\cdot 110 + 4\cdot 70 +7\cdot 60 + 2\cdot 100=1380.$$

Метод минимального элемента

В методе северо-западного угла на каждом шаге потребности первого из оставшихся пунктов назначения удовлетворялись за счет запасов первого из оставшихся пунктов отправления. Очевидно, выбор пунктов назначения и отправления целесообразно производить, ориентируясь на тарифы перевозок, а именно на каждом шаге выбираем какую-нибудь клетку, отвечающую минимальному тарифу (если таких клеток несколько, можно взять любую из них), и рассмотреть пункты назначения и отправления, соответствующие выбранной клетке. Сущность метода минимального элемента и состоит в выборе клетки с минимальным тарифом. Отметим, что этот метод, как правило, позволяет найти опорный план транспортной задачи, при котором общая стоимость перевозок груза меньше, чем общая стоимость перевозок при плане, найденном для данной задачи с помощью метода северо-западного угла. Поэтому наиболее целесообразно опорный план транспортной задачи находить методом минимального элемента.

Пример. *Найти опорный план транспортной задачи из предыдущего примера методом минимального элемента.*

Решение. Исходные данные задачи запишем в виде табл. 2.16. В каждой строке и каждом столбце отмечаем галочкой клетки с наименьшей стоимостью перевозки. После чего проставляем перевозки: вначале заполняя клетки с двумя галками, потом — с одной, а потом — оставшиеся, при этом не должно нарушаться условие допустимости плана.

Таблица 2.16

Пункты отправления	Пункты назначения					Запасы
	B_1	B_2	B_3	B_4	B_5	
A_1	✓✓ 2	✓ 3	4	✓✓ 2	4	140
A_2	8	4	✓✓ 1	4	✓✓ 1	180
A_3	9	7	3	7	✓ 2	160
Потребности	60	70	120	130	100	480

Минимальный тариф, равный 1, находится в клетке для переменной x_{23}. Положим x_{23}= 120, запишем это значение в соответствующую клетку табл. 4.6 и исключим временно из рассмотрения столбец B_3 . А для пункта A_2 запас будем считать равным 180–120=60 ед.

Кроме того, минимальный тариф, равный 1, находится в клетке для переменной x_{25}. Положим $x_{25} = 60$, запишем это значение в соответствующую клетку табл. 2.17 и исключим временно из рассмотрения столбец B_3 .

Таблица 2.17

Пункты отправления	Пункты назначения					Запасы
	B_1	B_2	B_3	B_4	B_5	
A_1	2	3	4	2	4	140
A_2	8	4	120 / 1	4	60 / 1	180
A_3	9	7	3	7	2	160
Потребности	60	70	120	130	100	480

В оставшейся части таблицы с двумя строками A_1 и A_3 и четырьмя столбцами B_1, B_2, B_4 и B_5 клетка с наименьшим значением тарифа c_{ij} =2 находится на пересечении строки A_3 и столбца B_5, строки A_1 и столбцов B_1 и B_4. Не нарушая допустимости плана, положим $x_{35} = 40$, $x_{11} = 60$, $x_{14} = 80$ и внесем эти значения в соответствующие клетки табл. 2.18.

Таблица 2.18

Пункты отправления	Пункты назначения					Запасы
	B_1	B_2	B_3	B_4	B_5	
A_1	60 / 2	3	4	80 / 2	4	140
A_2	8	4	120 / 1	4	60 / 1	180
A_3	9	7	3	7	40 / 2	160 (120)
Потребности	60	70	120	130 (50)	100	480

Теперь исключим из рассмотрения строки A_1 и A_2 и столбцы B_1, B_3, B_5, т. к. по ним выполнены условия допустимости. После этого аналогично заполняем оставшуюся часть таблицы (табл. 2.19).

Таблица 2.19

Пункты отправления	Пункты назначения					Запасы
	B_1	B_2	B_3	B_4	B_5	
A_1	60 [2]	[3]	[4]	80 [2]	[4]	140
A_2	[8]	[4]	120 [1]	[4]	60 [1]	180
A_3	[9]	70 [7]	[3]	50 [7]	40 [2]	160
Потребности	60	70	120	130	100	480

В результате получим опорный план

$$X = \begin{pmatrix} 60 & 0 & 0 & 80 & 0 \\ 0 & 0 & 120 & 0 & 60 \\ 0 & 70 & 0 & 50 & 40 \end{pmatrix}$$

При данном плане перевозок общая стоимость перевозок составляет

$$F = 2\cdot 60 + 2\cdot 80 + 1\cdot 120 + 1\cdot 60 + 7\cdot 70 + 7\cdot 50 + 2\cdot 40 = 1380.$$

Лекция 2.2.2. Операционные модели транспортного типа. Методы их решения: нахождение оптимального решения транспортной задачи

Для определения оптимального плана транспортной задачи разработано несколько методов. Однако наиболее часто используется метод потенциалов.

Метод потенциалов

Общий принцип определения оптимального плана транспортной задачи методом потенциалов аналогичен принципу решения задачи линейного программирования симплексным методом, а именно сначала находят опорный план транспортной задачи, а затем его последовательно улучшают до получения оптимального плана.

Для определения опорного плана транспортной задачи будем пользоваться одним из методов, изложенных ранее. Эти методы гарантируют получение занятых в исходном плане $n+m-1$ клеток, причем в некоторых из них могут стоять нули. Построенный план следует проверить на оптимальность.

<u>ТЕОРЕМА</u>. *Если для некоторого опорного плана* $X^* = (x_{ij}^*)$ $(i = \overline{1,m}\ \ j = \overline{1,n})$ *транспортной задачи существуют такие* $n+m$ *чисел* $\alpha_1, \alpha_2, ..., \alpha_m, \beta_1, \beta_2, ..., \beta_n$, *что*

$$\alpha_i + \beta_j = c_{ij} \text{ при } x_{ij}^* > 0 \tag{2.21}$$

и

$$\alpha_i + \beta_j \le c_{ij} \text{ при } x_{ij}^* = 0 \tag{2.22}$$

для всех $i = \overline{1,m}$ *и* $j = \overline{1,n}$, *то* $X^* = (x_{ij}^*)$ *— оптимальный план транспортной задачи.*

Числа α_i *и* β_j, $(i = \overline{1,m}\ \ j = \overline{1,n})$ *называются* ***потенциалами****, соответственно пунктов назначения и пунктов потребления.*

Сформулированная теорема позволяет построить алгоритм нахождения решения транспортной задачи. Он состоит в следующем. Пусть одним из рассмотренных выше методов найден опорный план транспортной задачи. Для каждого из пунктов отправления и назначения определяют потенциалы α_i и β_j $(i = \overline{1,m}\ \ j = \overline{1,n})$. Эти числа находят из системы уравнений

$$\alpha_i + \beta_j = c_{ij}, \tag{2.23}$$

где c_{ij} — тарифы, стоящие в заполненных клетках таблицы условий транспортной задачи.

Так как число заполненных клеток равно $n + m - 1$, то система (2.23) с $n+m$ неизвестными содержит $n + m - 1$ уравнений. Поскольку число неизвестных превышает на единицу число уравнений, одно из неизвестных можно положить равным произвольному числу, например $\alpha_1 = 0$, и найти последовательно из уравнений (2.23) значения остальных неизвестных. Для исследования плана на оптимальность для каждой свободной клетки проверяется условие $\alpha_i + \beta_j \le c_{ij}$. Если хотя бы одна свободная клетка не удовлетворяет данному условию, то опорный план не будет оптимальным, его можно улучшить за счет загрузки этой клетки. Если таких клеток несколько, то наиболее перспективна для загрузки клетка, для которой разность (платеж) между тарифом клетки и суммой потенциалов наименьшая, т. е.

$$\gamma_{ij} = c_{ij} - \left(\alpha_i + \beta_j\right) < 0 .$$

Экономически платеж показывает, на сколько денежных единиц уменьшатся транспортные издержки от загрузки данной клетки единицей груза.

Если для всех свободных клеток чисела $\gamma_{ij} \ge 0$, то опорный план перевозок является оптимальным.

Если для опорного плана перевозок указанное условие оптимальности не выполняется, то за счет загрузки свободной клетки с отрицательным платежом план перевозок улучшается. Для наиболее перспективной свободной клетки строится замкнутый цикл с вершинами в загруженных клетках. Для этого рассматривают все свободные клетки, для которых $\gamma_{ij} < 0$, и среди данных чисел выбирают наименьшее. Клетку, которой это число соответствует, следует заполнить, она далее становится вершиной цикла. Вершинам этого цикла условно приписываются знаки: свободной клетке — плюс, следующей по часовой или против часовой стрелки занятой клетке — минус, следующей — снова плюс и т. д. Из поставок в клетках цикла с «отрицательными» вершинами выбирается наименьшее количество груза, которое и перемещается по клеткам этого цикла: прибавляется к поставкам в положительных вершинах и вычитается из поставок в отрицательных, в результате чего баланс цикла не нарушится.

Циклом *в таблице условий транспортной задачи называется ломаная линия, вершины которой расположены в занятых клетках таблицы, а звенья — вдоль строк и столбцов, причем в каждой вершине цикла встречается ровно два звена, одно из которых находится в строке, а другое — в столбце.*

Если ломаная линия, образующая цикл, пересекается, то точки самопересечения не являются вершинами. Примеры некоторых циклов показаны на рис. 2.5.

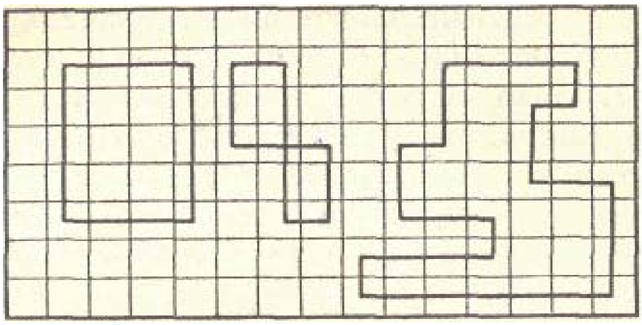


Рис. 2.5. Примеры построения циклов в транспортной задаче

При правильном построении опорного плана для любой свободной клетки можно построить лишь один цикл.

В результате указанных выше перемещений грузов в пределах клеток, связанных циклом с данной свободной клеткой, определяют новый опорный план транспортной задачи.

Описанный выше переход от одного опорного плана транспортной задачи к другому называется *сдвигом по циклу пересчета*.

Следует отметить, что при сдвиге по циклу пересчета число занятых клеток остается неизменным, а именно остается равным $n + m - 1$. При этом если в минусовых клетках имеется два (или более) одинаковых числа x_{ij}, то освобождают лишь одну из таких клеток, а остальные оставляют занятыми (с нулевыми поставками).

Новый опорный план транспортной задачи проверяют на оптимальность. Для этого определяют потенциалы пунктов отправления и назначения и находят числа γ_{ij} для всех свободных клеток. Если среди этих чисел не окажется отрицательных, это свидетельствует о получении оптимального плана. Если же отрицательные числа имеются, то следует перейти к новому опорному плану. В результате итерационного процесса после конечного числа шагов получают оптимальный план задачи.

Из изложенного следует, что процесс нахождения решения транспортной задачи методом потенциалов включает такие этапы:

1. Находят опорный план. При этом число заполненных клеток должно быть равным $n + m - 1$.
2. Находят потенциалы β_j и α_i, соответственно, пунктов назначения и отправления.
3. Для каждой свободной клетки определяют число γ_{ij}. Если среди чисел γ_{ij} нет отрицательных, то получен оптимальный план транспортной задачи; если же они имеются, то переходят к новому опорному плану.
4. Среди отрицательных чисел γ_{ij} выбирают наименьшее, строят для свободной клетки, которой оно соответствует, цикл пересчета и производят сдвиг по циклу пересчета.
5. Полученный опорный план проверяют на оптимальность, т. е. снова повторяют все действия, начиная с этапа 2.

В заключение отметим, что при определении опорного плана или в процессе решения задачи может быть получен вырожденный опорный план. Чтобы избежать в этом случае зацикливания, следует соответствующие нулевые элементы опорного плана заменить сколь угодно малым положительным числом ε и решать задачу как невырожденную. В оптимальном плане такой задачи необходимо считать ε равным нулю.

Пример. *В трех хранилищах горючего ежедневно хранится 50, 30 и 10 т бензина соответственно. Этот бензин ежедневно получают четыре заправочные станции в количествах, соответственно равных 30, 30, 10 и 20 т. Стоимость перевозки одной тонны бензина из хранилищ к заправочным станциям задаётся матрицей C:*

$$C = \begin{pmatrix} 1 & 2 & 4 & 1 \\ 2 & 3 & 1 & 5 \\ 3 & 2 & 4 & 4 \end{pmatrix}.$$

Составить такой план перевозок бензина, при котором общая стоимость перевозок будет наименьшей.

Решение. Построим таблицу транспортной задачи.

Таблица 2.20

Пункты отправления	Пункты назначения				Запасы
	B_1	B_2	B_3	B_4	
A_1	30 [1]	20 [2]	[4]	[1]	50
A_2	[2]	10 [3]	10 [1]	10 [5]	30
A_3	[3]	[2]	[4]	10 [4]	10
Потребности	30	30	10	20	90

Сначала, используя метод северо-западного угла, находим опорный план задачи. Этот план записан в табл. 2.20. Найденный опорный план проверяем на оптимальность. В связи с этим находим потенциалы пунктов отправления и назначения.

Для определения потенциалов получаем систему

$$\alpha_1 + \beta_1 = 1,\ \alpha_1 + \beta_2 = 2,\ \alpha_2 + \beta_2 = 3,\ \alpha_2 + \beta_3 = 1,\ \alpha_2 + \beta_4 = 5,$$
$$\alpha_3 + \beta_4 = 4,$$

содержащую шесть уравнений с семью неизвестными. Полагая $\alpha_1 = 0$, находим:

$$\beta_1 = 1,\ \beta_2 = 2,\ \alpha_2 = 1,\ \beta_3 = 0,\ \beta_4 = 4,\ \alpha_3 = 0.$$

Для каждой свободной клетки вычисляем число $\gamma_{ij} = c_{ij} - (\alpha_i + \beta_j)$: $\gamma_{13} = 4$, $\gamma_{14} = -3$, $\gamma_{21} = 0$, $\gamma_{31} = 2$, $\gamma_{32} = 0$, $\gamma_{33} = 4$. Так как среди чисел γ_{ij} имеется отрицательное ($\gamma_{14} = -3$), то построенный план перевозок не является оптимальным и надо перейти к новому опорному плану.

Таблица 2.21

Пункты отправления	Пункты назначения				Запасы	α_i
	B_1	B_2	B_3	B_4		
A_1	30 (1)	20 (2)	(4)	(1)	50	0
A_2	(2)	10 (3)	10 (1)	10 (5)	30	1
A_3	(3)	(2)	(4)	10 (4)	10	0
Потребности	30	30	10	20	90	
β_j	1	2	0	4		

Для свободной клетки с $\gamma_{14} = -3$ строим цикл пересчета (табл. 2.21) и производим сдвиг по этому циклу. Наименьшее из чисел в минусовых клетках равно 10 = min(20, 10). Клетка, в которой находится это число, становится свободной в новой табл. 2.22. Другие числа в табл. 2.22 получаются так: к числу 10, стоящему в плюсовой клетке табл. 2.21, добавим 10 и вычтем 10 из числа 20, находящегося в минусовой клетке табл. 2.21. Клетка на пересечении строки A_2 и столбца B_4 становится свободной. После этих преобразований получаем новый опорный план (табл. 2.22).

Этот план проверяем на оптимальность. Снова находим потенциалы пунктов отправления и назначения. Для этого составляем следующую систему уравнений:

$$\alpha_1+\beta_1=1,\ \alpha_1+\beta_2=2,\ \alpha_1+\beta_4=1,\ \alpha_2+\beta_2=3,\ \alpha_2+\beta_3=1,\ \alpha_3+\beta_4=4.$$

Полагаем $\alpha_1=0$, получаем

$$\beta_1=\beta_4=1,\ \beta_2=2,\ \beta_3=0,\ \alpha_3=3,\ \alpha_2=1.$$

Для каждой свободной клетки вычисляем число γ_{ij}; имеем:

$$\gamma_{13}=4,\ \gamma_{21}=0,\ \gamma_{24}=3,\ \gamma_{31}=-1,\ \gamma_{32}=-3,\ \gamma_{33}=1.$$

Таблица 2.22

Пункты отправления	Пункты назначения				Запасы	α_i
	B_1	B_2	B_3	B_4		
A_1	30 [1]	10 [2]	[4]	10 [1]	50	0
A_2	[2]	20 [3]	10 [1]	[5]	30	1
A_3	[3]	[2]	[4]	10 [4]	10	3
Потребности	30	30	10	20	90	
β_j	1	2	0	1		

Таким образом, видим, что данный план перевозок не является оптимальным. Поэтому переходим к новому опорному плану (табл. 2.23).

Сравнивая значения $\alpha_i+\beta_j$ новых потенциалов, отвечающих свободным клеткам табл. 2.23, с соответствующими числами c_{ij}, видим, что указанные значения сумм потенциалов для всех свободных клеток не превосходят соответствующих чисел c_{ij}.

Таблица 2.23

Пункты отправления	Пункты назначения				Запасы	α_i
	B_1	B_2	B_3	B_4		
A_1	30 [1]	[2]	[4]	20 [1]	50	0
A_2	[2]	20 [3]	10 [1]	ε [5]	30	0
A_3	[3]	10 [2]	[4]	[4]	10	–1
Потребности	30	30	10	20	90	
β_j	1	3	1	1		

Следовательно, полученный план

$$X^* = \begin{pmatrix} 30 & 0 & 0 & 20 \\ 0 & 20 & 10 & 0 \\ 0 & 10 & 0 & 0 \end{pmatrix}$$

является оптимальным.

При данном плане стоимость перевозок

$$F = 1 \cdot 30 + 1 \cdot 20 + 3 \cdot 20 + 1 \cdot 10 + 2 \cdot 10 = 140.$$

Лекция 2.2.3. Целочисленные задачи исследования операций. Метод Гомори для нахождения их решения. Задача о назначениях и венгерский метод

В ряде экономических задач требование целочисленности — свойство решения, оно естественно возникает в задачах размещения производства, о закупках парка машин для выполнения производственной программы, при определении количества сезонных рабочих и т. д.

Исторически первой задачей целочисленного типа является опубликованная 1932 г. венгерским математикам Е. Эгервари задача о назначениях персонала. Приведем ниже аналогичную указанной задачу о распределении механизмов по видам работ.

Задача о назначениях

Пусть требуется выполнить n различных работ и имеется n механизмов (машин) для их выполнения, причем каждый механизм может использоваться на любой работе. Производительность механизма на различных работах, вообще говоря, различна. Обозначим через C_{ij} производительность i-го механизма на j-й работе. Задача заключается в таком распределении механизмов по работам, при котором суммарная производительность максимальна.

Построим математическую модель этой задачи. Сопоставим каждому из возможных вариантов распределения машин по работам набор значений неизвестных x_{ij}, относительно которых условимся, что $x_{ij} = 1$, если в данном варианте i-й механизм назначается на j-ю работу, и $x_{ij} = 0$, если i-й механизм назначается не на j-ю работу. Для любого варианта среди чисел x_{ij} должно быть точно n единиц, причем должны выполняться условия:

$$\sum_{j=1}^{n} x_{ij} = 1, \quad j = 1,2,\ldots,n$$

(каждый механизм назначается на одну работу) и

$$\sum_{i=1}^{n} x_{ij} = 1, \quad i = 1,2,\ldots,n$$

(на каждую работу назначен один механизм). Суммарная производительность при данном варианте назначения машин на работы выразится суммой:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}\,.$$

Математическая модель задачи:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \to \max;$$

$$\sum_{j=1}^{n} x_{ij} = 1,\ \ j=\overline{1,n} \tag{4.9}$$

$$\sum_{i=1}^{n} x_{ij} = 1,\ \ i=\overline{1,n}$$

$$x_{ij} = \begin{cases} 1, \text{ если } i\text{-й исполнитель назначен на } j\text{-ю работу}; \\ 0, \text{ если } i\text{-й исполнитель не назначен на } j\text{-ю работу}. \end{cases}$$

$$i=\overline{1,n};\ \ j=\overline{1,n}\,.$$

Условия выводят задачу о назначениях из класса задач линейного программирования, так как они нелинейны. Задачи математического программирования, в которых на переменные наложены такие условия, называются задачами с булевыми переменными. Поскольку все остальные условия и целевая функция нашей задачи линейны, мы должны формально отнести сё к классу задач линейного программирования с булевыми переменными. Однако задачу можно рассмотреть как частный случай транспортной (и, следовательно, просто линейной) задачи. В самом деле, если отбросить последнее условие, заменив его условиями неотрицательности переменных, то задача превращается в обычную транспортную задачу, имеющую ту особенность, что в ней все a_i (i= 1,2, ..., n) и все b_j (j=1, 2,..., n) равны единице.

Если решить эту задачу методом потенциалов или любым другим методом, который при целых a_i и b_j приводит к целочисленному оптимальному решению, то полученное решение автоматически будет удовлетворять не учтенному нами условию булевости переменных: если сумма нескольких целочисленных неотрицательных переменных равна единице, то каждая из переменных может быть равна только нулю или единице. Таким образом, в задаче о назначениях ограничение на значения переменных можно заменить просто условием неотрицательности переменных, если только метод решения обеспечивает целочисленность оптимального решения, к которому он приводит. Хотя для транспортной задачи есть методы, которые проще методов решения общей задачи линейного программирования, особенности задачи позволяют решить её с помощью ещё более простых приемов. Проведем соответствующие рассуждения. Пусть

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \dots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} —$$

матрица эффективностей (например, производительностей) задачи о назначениях. В соответствии с постановкой этой задачи, решить её — значит выбрать в матрице C n элементов, по одному из каждой строки и каждого столбца (строки соответствуют назначенным объектам, например механизмам, а столбцы — объектам, на которые производится назначение, например работам) так, чтобы сумма выбранных элементов, которая равна общей эффективности, соответствующей данному выбору, была наибольшей по сравнению с её значениями при всех других таких выборах. Мы сейчас увидим, что эту задачу можно свести к выбору n нулевых элементов, по одному в каждом столбце и каждой строке («правильному выбору»), в некоторой матрице с неотрицательными элементами, у которой в каждой строке и каждом столбце есть нули.

Две матрицы $C = (c_{ij})$ и $D = (d_{ij})$ назовем здесь ***эквивалентными****, если одна из них получается из другой прибавлением к элементам каждой строки одного и того же числа (может быть, для разных строк эти числа различны) и к элементам каждого столбца одного и того же числа (может быть, для разных столбцов эти числа различны), например при $d_{ij}=c_{ij}+e_i+f_j$,*

$$i=1, 2,\dots, n;\ \ j = 1, 2,\dots, n.$$

Ясно, что тогда и вторая матрица получается из первой прибавлением к ее строкам и столбцам постоянных чисел (e_i и f_j соответственно), так что свойство эквивалентности взаимно.

Например, матрицы

$$C = \begin{pmatrix} 4 & 0 & -1 & 5 \\ 2 & 3 & 0 & -1 \\ 1 & 1 & 2 & 3 \\ -2 & -2 & -3 & 0 \end{pmatrix} \text{ и } D = \begin{pmatrix} 6 & 3 & 3 & 6 \\ 4 & 6 & 4 & 0 \\ 4 & 5 & 7 & 5 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

эквивалентны, так как D получается из C прибавлением к строкам C чисел 0, 0, 1, –1, а к столбцам — чисел 2, 3, 4, 1 соответственно.

ТЕОРЕМА. *Множества оптимальных назначений двух задач выбора с эквивалентными матрицами совпадают.*

Доказательство. Пусть матрица C эквивалентна матрице D и пусть $(1, j_1)$, $(2, j_2)$, ..., (n, j_n) — оптимальное назначение в задаче выбора с матрицей C (первый механизм назначается на j_1-ю работу, второй — на j_2-ю, и т. д.) Допустим от противного, что для задачи с матрицей D это назначение не является оптимальным, т.е. существует другой выбор $(1, j_1)$, $(2, j_2)$, ..., (n, j_n), что общая эффективность при этом выборе превышает общую эффективность при первом выборе:

$$d'_{1j_1} + d'_{2j_2} + \ldots + d'_{nj_n} \geq d_{1j_1} + d_{2j_2} + \ldots + d_{nj_n}.$$

Так как по условию $d_{ij}=c_{ij} + e_i + f_j$, то отсюда:

$$c'_{1j_1} + e_1 + f'_{j_1} + c'_{2j_2} + e_2 + f'_{j_2} + \ldots + c'_{nj_n} + e_n + f'_{j_n} >$$
$$> c_{1j_1} + e_1 + f_{j_1} + c_{2j_2} + e_2 + f_{j_2} + \ldots + c_{nj_n} + e_n + f_{j_n}.$$

Очевидно, что

$$f'_{j_1} + f'_{j_2} + \ldots + f'_{j_n} = f_{j_1} + f_{j_2} + \ldots + f_{j_n},$$

так как и в том и другом случае это сумма чисел, прибавляемых к различным столбцам матрицы C, только, может быть, идущих в разном порядке. Поэтому имеем:

$$c'_{1j_1} + c'_{2j_2} + \ldots + c'_{nj_n} > c_{1j_1} + c_{2j_2} + \ldots + c_{nj_n},$$

а это противоречит тому, что для задачи с матрицей C выбор

$(1, j_1)$, $(2, j_2)$, …, (n, j_n) — оптимальный.

Так как свойство эквивалентности матриц взаимное, то так же доказывается, что любой оптимальный выбор задачи с матрицей D является и оптимальным выбором задачи с матрицей C. Теорема доказана. ■

Приведенная теорема позволяет, если это требуется, переходить от данной задачи выбора с матрицей C к задаче выбора с любой другой матрицей D, при условии, что D эквивалентна C.

Нам будет удобно предварительно перейти от данной задачи выбора на максимум к задаче выбора с теми же условиями, но на минимум, т. е. от матрицы $C = (c_{ij})$ перейти к матрице $C = (-c_{ij})$ и искать выбор, дающий минимальную сумму элементов.

Теперь перейдем от задачи на минимум с матрицей $-C$ к задаче на минимум с эквивалентной ей матрицей, которая имела бы только неотрицательные элементы и в каждом столбце и каждой строке которой было бы хотя бы по одному нулевому элементу. Для этого сначала прибавим к каждому столбцу матрицы $-C$ наибольший из элементов соответствующего столбца матрицы C (или, что то же самое, вычтем элементы каждого столбца матрицы C из наибольшего элемента этого столбца). Получится неотрицательная матрица C_1, в каждом столбце которой есть хотя бы один нуль. Теперь вычтем из каждой строки матрицы C_1 минимальный элемент этой строки. Полученная матрица D и будет неотрицательной, в каждом столбце и каждой строке которой есть хотя бы один нуль.

Итак, проделываем последовательные преобразования матрицы *C:*

$$C=(c_{ij})\xrightarrow{(1)}C_1=(c'_{ij})=(\max c_{ij}-c_{ij})\xrightarrow{(2)}D=(d_{ij})=(c'_{ij}-\min c'_{ij}).$$

Назовем эти преобразования предварительными. Если в каждой строке матрицы $C_1=(c'_{ij})$, полученной после первого преобразования матрицы *C*, есть хотя бы один нуль, то второе преобразование, конечно, не изменит матрицу *C* и практически не производится.

Наименьшее возможное значение суммы *n* элементов неотрицательной матрицы равно, очевидно, нулю. Таким образом, наша задача сводится теперь к выбору в матрице *D* или в эквивалентной ей матрице с неотрицательными элементами *n* нулевых элементов, по одному в каждой строке и каждом столбце. Покажем, как это можно сделать. Неформальный смысл приводимого ниже алгоритма заключается в последовательных переходах от одного правильного неполного выбора нулей к другому, содержащему на один нуль больше, чем предыдущий, до тех пор, пока не получится полный правильный выбор. При этом на отдельных этапах может потребоваться переход к новой матрице, эквивалентной предыдущей.

Приведем алгоритм решения задач о назначениях. Пусть уже проделаны предварительные преобразования матрицы эффективностей *C* данной задачи и получена неотрицательная матрица *D*, содержащая хотя бы по одному нулевому элементу в каждой строке и в каждом столбце.

1. Отмечаем (например, звездочкой) какой-нибудь нуль в первом столбце матрицы *D*, (0*); отмечаем звездочкой какой-нибудь нуль во втором столбце, не лежащий в той строке, в которой находится 0* из первого столбца (если такой нуль во втором столбце найдется); отмечаем звездочкой один из нулей третьего столбца, лежащий в строке, где нет еще нулей со звездочкой (если такой нуль в третьем столбце найдется); и так далее, пока не пройдем все столбцы матрицы. Если число отмеченных звездочкой нулей равно *n*, то процесс окончен: места, занимаемые нулями со звездочкой, соответствуют n переменным x_{ij}, равным 1, в оптимальном решении исходной задачи. Если нулей со звездочкой меньше *n*, то поступаем следующим образом:

2. Помечаем (например, знаком «+» сверху) столбцы матрицы, в которых есть 0*, и считаем эти столбцы занятыми. В ходе процесса будут появляться и занятые строки. Элементы, стоящие на пересечении незанятого столбца и незанятой строки, будут считаться незанятыми, остальные элементы – занятыми. Если в матрице нет незанятых нулей, то переходим к п. 5. Если незанятые нули есть, то выбираем первый из них (просматривая поочередно строки матрицы слева направо). Отмечаем его каким-нибудь промежуточным значком (например, штрихом — 0′). Если в его строке нет нуля со звездочкой, то переходим к п. 4; если в его строке 0* есть, то делаем следующее:

3. Освобождаем (снимаем знак «+» и считаем снова незанятым) столбец, в котором находится 0*, лежащий в той же строке, что и отмеченный только что штрихом нуль. Помечаем (например, знаком «+» справа) строку, в которой находится 0, и считаем её занятой. Возвращаемся ко второй части п. 2.

4. Начиная с только что отмеченного 0′, строим цепочку из нулей: от этого 0′, по столбцу к 0*, от него по строке к 0′ и т. д., пока это возможно. Цепочка оборвется (может быть, на первом же 0′) на некотором 0′. Снимаем звездочки у нулей из цепочки и заменяем звездочками штрихи у нулей из цепочки. Новый набор нулей со звездочками содержит на один больше чем предыдущий и является также правильным. Снимаем все пометки, кроме звездочек, и возвращаемся ко второй части п. 1.

5. Отыскиваем минимальный элемент среди незанятых элементов матрицы (пусть он равен h) и вычитаем его из всех незанятых строк, а затем прибавляем ко всем занятым столбцам. Никакие пометки при этом не снимаются. Получается матрица, эквивалентная предыдущей и содержащая незанятые нули. Возвращаемся к третьей части п. 2. Рассмотрим пример использования алгоритма.

Пример. *Найти оптимальный вариант назначений, если матрица эффективностей такова:*

$$C=\begin{pmatrix} 2 & 3 & 3 & 5 & 4 \\ 4 & 2 & 4 & 6 & 2 \\ 2 & 2 & 2 & 4 & 3 \\ 4 & 3 & 4 & 3 & 5 \\ 0 & 1 & 0 & 2 & 0 \end{pmatrix}$$

Решение. Приведем цепочку матриц, получающихся в процессе решения задачи, с соответствующими проектами. Снятие значка отмечено заключением его в прямоугольник. Над стрелками переходов от матрицы указаны пункты алгоритма, которые использовались при соответствующих преобразованиях.

Процесс окончен, так как получилось n = 5 нулей со звездочкой. Оптимальный вариант назначений: $x_{15}=x_{24}=x_{31}=x_{43}=x_{52}=1$, остальные $x_{ij}=0$, т. е. первый механизм назначается на пятую работу, второй — на четвертую, третий — на первую, четвертый — на третью, пятый — на вторую.

Изложенный алгоритм в литературе называется венгерским алгоритмом (венгерский метод) решения задачи о назначениях.

$$\begin{pmatrix} 2 & 3 & 3 & 5 & 4 \\ 4 & 2 & 4 & 6 & 2 \\ 2 & 2 & 2 & 4 & 3 \\ 4 & 3 & 4 & 3 & 5 \\ 0 & 1 & 0 & 2 & 0 \end{pmatrix} \xrightarrow{(1)} \begin{pmatrix} 2 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 3 \\ 2 & 1 & 2 & 2 & 2 \\ 0 & 0 & 0 & 3 & 0 \\ 4 & 2 & 4 & 4 & 5 \end{pmatrix} \xrightarrow{(2)}$$

$$\begin{pmatrix} 2 & 0^* & 1 & 1 & 1 \\ 0^* & 1 & 0 & 0' & 3 \\ 1 & 0 & 1 & 1 & 1 \\ 0' & 0 & 0^* & 3 & 0 \\ 2 & 0 & 2 & 2 & 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0^* & 0' & 0 & 0 \\ 0^* & 2 & 0 & 0' & 3 \\ 0' & 0 & 0 & 0 & 0 \\ 0' & 1 & 0^* & 3 & 0 \\ 1 & 0 & 1 & 1 & 2 \end{pmatrix} \rightarrow$$

$$\to\begin{pmatrix} 1 & 0* & 0 & 0 & 0 \\ 0 & 2 & 0 & 0* & 3 \\ 0* & 0' & 0 & 0 & 0 \\ 0' & 1 & 0* & 3 & 0 \\ 1 & 0' & 1 & 1 & 2 \end{pmatrix} \xrightarrow{\textit{результат}} \begin{pmatrix} 1 & 0 & 0 & 0 & 0* \\ 0 & 2 & 0 & 0* & 3 \\ 0* & 0 & 0 & 0 & 0 \\ 0 & 1 & 0* & 3 & 0 \\ 1 & 0* & 1 & 1 & 2 \end{pmatrix}$$

Общая модель целочисленной задачи и метод Гомори для её решения

Задача о назначениях относится к частному случаю задачи целочисленного программирования.

Математическая модель задачи: в области, определенной условиями $\sum_{j=1}^{n} a_{ij}x_j = b_i,\ i=\overline{1,m}$, найти решение $x=(x_1,x_2,\ldots, x_n)$, при котором максимизируется значение функции

$$F(x)=\sum_{j=1}^{n} c_j x_j \text{ и при этом } x_j \text{ — целое } (j=\overline{1,n}).$$

Метод Гомори

Нахождение значения задачи целочисленного программирования методом Гомори начинают с определения симплексным методом оптимального плана задачи без учета целочисленности переменных. После того как этот план найден, просматривают его компоненты. Если среди компонент нет дробных чисел, то найденный план является оптимальным планом задачи целочисленного программирования. Если же в оптимальном плане задачи переменная x_j принимает дробное значение, то к системе уравнений добавляют неравенство следующего вида:

$$\sum_{j=1}^{n} f(a_{ij}^{*})\, x_j \ge f(b_i^{*})$$

и находят решение задачи, сформулированной выше. В данном неравенстве a_{ij}^{*} и b_i^{*} — преобразованные исходные величины a_{ij} и b_i, значения которых взяты из последней симплекс-таблицы, а $f(a_{ij}^*)$ и $f(b_i^*)$ — дробные части чисел (под дробной частью некоторого числа a понимается наименьшее неотрицательное число b такое, что разность между a и b есть целое). Если в оптимальном плане целочисленной задачи дробные значения принимают несколько переменных, то дополнительное неравенство определяется наибольшей дробной частью.

Если в оптимальном плане задачи с дополнительным ограничением по-прежнему присутствуют дробные значения, то снова добавляют одно дополнительное ограничение и процесс вычислений повторяют. Проводя конечное число итераций, либо получают оптимальный план задачи целочисленного программирования, либо устанавливают ее неразрешимость. Процесс определения оптимального плана задачи целочисленного программирования методом Гомори включает следующие основные этапы:

1. Используют симплексный метод, находят решение задачи без учета требования целочисленности переменных.

2. Составляют дополнительное ограничение для переменной, которая в оптимальном плане исходной задачи имеет максимальное дробное значение, а в оптимальном плане преобразованной задачи должна быть целочисленной.

3. Используя метод искусственного базиса, находят решение задачи, получающейся в результате присоединения дополнительного ограничения.

4. В случае необходимости составляют еще одно дополнительное ограничение и продолжают итерационный процесс до получения оптимального плана задачи или установления ее неразрешимости.

Условие неразрешимости линейной задачи в целых числах

Если в оптимальном плане задачи, решение которой находилось по методу Гомори, среди координат оптимального плана присутствуют дробные значения, а в строке коэффициентов матрицы A среди соответствующих этим дробным значениям координат нет дробных чисел, то ограничение Гомори построить невозможно и делается вывод о неразрешимости задачи в целых числах.

Лекция 2.2.4. Теория массового обслуживания: основные понятия, определения, теоремы

При исследовании операций часто приходится сталкиваться с системами, предназначенными для многоразового использования при решении однотипных задач. Возникающие при этом процессы получили название **процессов обслуживания,** а системы — **систем массового обслуживания** (СМО). Таким образом, теория массового обслуживания — это область прикладной математики, занимающаяся анализом процессов в системах производства, обслуживания, управления, в которых однородные события повторяются многократно. Примерами таких систем являются телефонные системы, ремонтные мастерские, вычислительные комплексы, билетные кассы, магазины, парикмахерские и т. п.

Как и любой раздел прикладной математики или исследования операций, теория массового обслуживания возникла в связи с решением практических задач. Если обратиться к самой ранней истории вопроса, то следует упомянуть опубликованную в 1907 г. статью Джоаннсена «Продолжительности ожидания и число попыток выйти на связь в телефонной сети». Это была одна из первых научных публикаций по проблеме очередей, однако первой математически корректной работой по теории очередей следует считать работу Эрланга «Теория вероятностей и телефонные разговоры». Таким образом, первые шаги в теории массового обслуживания были сделаны при исследовании систем связи. Системы связи как объекты приложений теории массового обслуживания и в настоящее время доминируют. После 1960 г. появилась другая важная сфера приложений теории массового обслуживания — проектирование и создание электронно-вычислительных комплексов. За последние годы существенно усовершенствованы модели массового обслуживания в транспортных системах. Следует отметить, что во многих случаях транспортные модели содержат как стохастические, так и детерминированные элементы. Системы технического обслуживания также относятся к числу типичных СМО.

Каждая СМО состоит из определенного числа обслуживающих единиц (приборов, устройств, пунктов, станций), которые будем называть ***каналами*** обслуживания. Каналами могут быть линии связи, рабочие точки, вычислительные машины, продавцы и др. По числу каналов СМО подразделяют на ***одноканальные* и *многоканальные*** (рис. 2.6).

Заявки поступают в СМО обычно не регулярно, а случайно, образуя так называемый ***случайный поток заявок (требований).*** Обслуживание заявок также продолжается какое-то случайное время. Случайный характер потока заявок и времени обслуживания приводит к тому, что СМО оказывается загруженной неравномерно: в какие-то периоды времени скапливается очень большое количество заявок, при этом они выстраиваются в очередь либо покидают СМО необслуженными, в другие же периоды СМО работает с недогрузкой или простаивает.

Предмет теории массового обслуживания — установление зависимостей между характером потока заявок, числом каналов обслуживания, производительностью отдельного канала и эффективным обслуживанием с целью нахождения наилучших путей управления этими процессами.

Задача теории массового обслуживания — установить зависимость результирующих показателей работы СМО (вероятности того, что заявка будет обслужена; математического ожидания числа обслуженных заявок и т. д.) от входных показателей (количества каналов в системе, параметров входящего потока заявок и т. д.). Задачи теории массового обслуживания носят оптимизационный характер и в конечном итоге включают экономический аспект по определению такого варианта системы, при котором будет обеспечен минимум суммарных затрат от ожидания обслуживания, потерь времени и ресурсов на обслуживание и простоев каналов обслуживания.

Результирующие показатели или интересующие характеристики СМО — показатели эффективности, которые описывают, способна ли данная система справляться с потоком заявок. В качестве них используются: среднее число заявок, обслуживаемых в единицу времени; среднее число заявок в очереди; среднее время ожидания обслуживания; вероятность отказа в обслуживании без ожидания; вероятность того, что число заявок в очереди превысит определенное значение, и т. п.

Рис. 2.6. Классификация систем массового обслуживания

СМО включает следующие элементы:

— источник требований,
— входящий поток требований,
— очередь, обслуживающее устройство
(обслуживающий аппарат, канал обслуживания),
— выходящий поток требований.

Входящий поток заявок на обслуживание

Если акты поступления требований на обслуживание и процедуры обслуживания реализуются строго по графику, то очереди можно избежать. Но на практике все происходит иначе, и в большинстве случаев поступление требований зависит от внешних обстоятельств. Следовательно, лучшее, что можно сделать, — это описать процесс поступления требований через случайные величины. К числу основных показателей, необходимых для описания входного потока, относятся:

— характеристики источника требований,
— тип требований,
— длина интервалов времени между поступлениями требований.

Механизм обслуживания

СМО различаются числом обслуживающих приборов, количеством одновременно обслуживаемых требований, продолжительностью и типом обслуживания. Например, в задачах, возникающих при рассмотрении потоков в сетях, и в задачах составления графиков работы ремонтных предприятий приходится иметь дело с несколькими обслуживающими приборами, функционирующими последовательно и (или) параллельно. Указанные выше характеристики процесса обслуживания также описываются с помощью случайных величин.

Дисциплина очереди

Все другие факторы, ассоциированные с правилами поведения очереди, можно включить в этот пункт. К их числу относятся, прежде всего, правила, в соответствии с которыми обслуживающий механизм принимает поступившую заявку к обслуживанию. Во многих ситуациях с целью придания описанию исследуемого процесса большего реализма приходится прибегать к понятию «приоритет». Ряд других факторов, которые также можно отнести к данному пункту, отражают особенности индивидуального поведения клиентов (уклонение от обслуживания, переход из одной очереди в другую, обманные действия и т. п.).

Очередь *— линейная цепочка выстроившихся один за другим объектов, нуждающихся в том или ином виде обслуживания.*

Классификация систем массового обслуживания

СМО классифицируют по разным признакам. К таким признакам относятся ***условия ожидания требованием начала обслуживания***. В соответствии с этим признаком системы подразделяются на следующие виды:

— СМО с отказами;

— СМО с ожиданием;

— СМО с ограниченной длиной очереди;

— СМО с ограниченным временем ожидания.

В СМО с отказами заявка, поступившая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем процессе обслуживания не участвует (например, заявка на телефонный разговор в момент, когда все каналы заняты, получает отказ и покидает СМО не обслуженной). В СМО с ожиданием заявка, пришедшая в момент, когда все каналы заняты, не уходит, а становится в очередь на обслуживание.

СМО с ожиданием подразделяются на разные виды в зависимости от того, как организована очередь: с ограниченной или неограниченной длиной очереди, с ограниченным временем ожидания и т. п.

По числу каналов или приборов СМО делятся на:

— одноканальные,

— многоканальные.

По месту нахождения источника требований СМО делятся:

— на разомкнутые, когда источник находится вне системы,

— на замкнутые, когда источник находится в самой системе.

К последнему виду относится, например, станочный участок, в котором станки являются источником неисправностей, а следовательно, и требований на их обслуживание.

Для классификации СМО значение имеет **дисциплина обслуживания,** определяющая порядок выбора заявок из числа поступивших и порядок распределения их между свободными каналами. По этому признаку обслуживание заявки может быть организовано по принципам:

— **«первая пришла — первая обслужена»,**

— **«последняя пришла — первая обслужена»**

(такой порядок может применяться, например, при извлечении для обслуживания изделий со склада, ибо последние из них оказываются часто более доступными)

или **обслуживание с приоритетом** (когда в первую очередь обслуживаются наиболее важные заявки). Приоритет может быть как **абсолютным,** когда более важная заявка вытесняет из-под обслуживания обычную (например, в случае аварийной ситуации плановые работы ремонтных бригад прерываются до ликвидации аварии), так и **относительным,** когда более важная заявка получает лишь «лучшее» место в очереди.

Потоки событий

Потоком событий *называется последовательность однородных событий, появляющихся одно за другим в случайные моменты времени.*

Поток событий может быть изображен точками с абсциссами $\Theta_1, \Theta_2, ..., \Theta_n, ...$ и с интервалами между ними

$$T_1 = \Theta_2 - \Theta_1,\ T_2 = \Theta_3 - \Theta_2,\ \ldots,\ T_n = \Theta_{n+1} - \Theta_n .$$

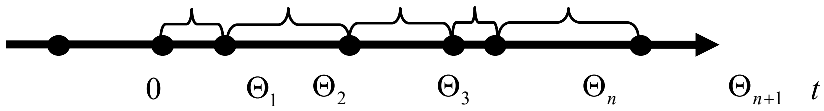


Рис. 2.7. Поток событий на временной оси

При вероятностном описании поток событий может быть представлен как последовательность случайных величин:

$$\Theta_1,\ \Theta_2 = \Theta_1 + T_1,\ \Theta_3 = \Theta_1 + T_1 + T_2, \ldots$$

Необходимо отметить, что на рис. 2.7 изображен не сам поток, а его конкретная реализация. Дадим классификацию потоков событий.

Поток событий называется ***регулярным****, если события следуют одно за другим через определенные равные промежутки времени.*

Примером регулярного потока может служить поток изделий на конвейере сборочного цеха (с постоянной скоростью движения).

Поток событий называется ***стационарным****, если его вероятностные характеристики не зависят от выбора начала отсчета или если вероятность попадания того или другого числа событий на любой интервал времени зависит только от длины τ этого интервала и не зависит от того, где именно на временной оси он расположен.*

Поток автомобилей на городском проспекте не является стационарным в течение суток, но этот поток можно считать стационарным в течение суток, скажем, в часы пик.

Поток событий называется ***ординарным****, если вероятность попадания на элементарный интервал времени Δt двух или более событий пренебрежимо мала по сравнению с вероятностью попадания одного события.*

В качестве примера ординарного потока можно взять поток поездов, подходящих к железнодорожной станции.

Поток событий называется ***без последействия****, если число событий, попадающих на любой интервал времени* τ*, не зависит от того, сколько событий попало на любой другой не пересекающийся с ним интервал.*

Поток пассажиров, входящих в метро, можно считать не имеющим последействия.

Поток событий называется ***простейшим****, если он стационарен, ординарен и не имеет последействия.*

Интервал времени *T* между двумя соседними событиями простейшего потока имеет показательное распределение, т. е.

$$f(t) = \lambda e^{-\lambda t} \quad при \quad t > 0, \tag{2.24}$$

где $\lambda = \dfrac{1}{M[T]}$ — величина, обратная среднему значению интервала *T*.

Интенсивностью λ *потока событий называется среднее число (математическое ожидание) числа событий, происходящее на единицу времени.*

Для стационарного потока $\lambda = const$, для нестационарного потока интенсивность в общем случае зависит от времени $\lambda = \lambda(t)$. Заметим, что регулярный поток не является простейшим, так как он обладает последействием: моменты появления событий в таком потоке жестко зафиксированы.

Понятие марковского случайного процесса

Процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем, т. е. состояние СМО меняется скачком в случайный момент появления каких-то событий (например, прихода новой заявки, окончания обслуживания и т. п.). При этом под случайным (вероятностным или стохастическим) процессом понимается процесс изменения во времени состояния какой-либо системы в соответствии с вероятностными закономерностями.

При анализе случайных процессов с дискретными состояниями удобно пользоваться геометрической схемой — ***графом состояний*** *(рис. 2.8).*

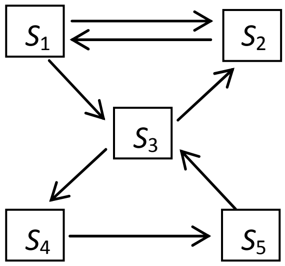


Рис. 2.8. Пример графа состояний системы

Обычно состояния системы изображаются прямоугольниками, а возможные переходы из состояния в состояние — стрелками (ориентированными дугами), соединяющими состояния.

Граф состояний системы с проставленными у стрелок интенсивностями называют ***размеченным***.

Процесс называется процессом с ***дискретными состояниями****, если его возможные состояния S_1, S_2, S_3,… можно заранее перечислить, а переходы из состояния в состояние происходят мгновенно (скачком). Процесс называется* ***процессом с непрерывным временем****, если моменты возможных переходов системы из состояния в состояние не фиксированы заранее, а происходят случайно.*

Анализ работы СМО существенно упрощается, если процесс этой работы — марковский.

Случайный процесс называется ***марковским*** *или* ***случайным процессом без последействия****, если для любого момента времени t_0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент времени t_0 и не зависят от того, когда и как система пришла в это состояние.*

Пример. *Система* ***S*** *— счетчик в такси — пример марковского процесса. Состояние системы в момент времени t характеризуется числом километров, пройденных автомобилем до данного момента. Пусть в момент t_0 счетчик показывал S_0. Вероятность того, что в момент $t > t_0$ счетчик покажет то или иное число километров S_1, зависит от S_0, но не зависит от того, в какие моменты времени изменялись показания счетчика.*

Многие процессы можно приближенно считать марковскими. Например, процесс игры в шахматы; система $\boldsymbol{S}$ — группа шахматных фигур. Состояние системы характеризуется числом фигур противника, сохранившихся на доске в момент t_0. Вероятность того, что в момент $t > t_0$ материальный перевес будет на стороне одного из противников, зависит в первую очередь от того, в каком состоянии находится система в данный момент t_0, а не от того, когда и в какой последовательности исчезли фигуры с доски до момента t_0.

Уравнения Колмогорова для вероятностей состояний

Рассмотрим математическое описание марковского процесса с дискретными состояниями и непрерывным временем. Системы, их описывающие, обычно исследуют с помощью уравнений Колмогорова для вероятностей состояний.

Плотностью вероятности перехода λ_{ij} *из состояния* S_i, *в состояние* S_j, *называется предел отношения вероятности этого перехода за время* Δt *к длине промежутка* Δt, *когда последний стремится к нулю:*

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t}, \tag{2.25}$$

где $P(\Delta t)$ *— вероятность того, что система, находившаяся в момент* t *в состоянии* S_i, *за время* Δt *перейдет в состояние* S_j.

Рассмотрим марковский процесс с дискретными состояниями и непрерывным временем на примере случайного процесса, граф которого изображен на рис. 2.9. Будем полагать, что все переходы системы из состояния S_i в S_j происходят под воздействием простейших потоков событий с интенсивностями λ_{ij}, $i, j = \overline{0,3}$. Рассматриваемая система $\boldsymbol{S}$ имеет четыре возможных состояния: S_0, S_1, S_2, S_3.

Вероятностью i состояния *называется вероятность* $p_i(t)$ *того, что в момент* t *система будет находиться в состоянии* S_i.

Очевидно, что для любого момента t сумма вероятностей всех состояний равна единице.

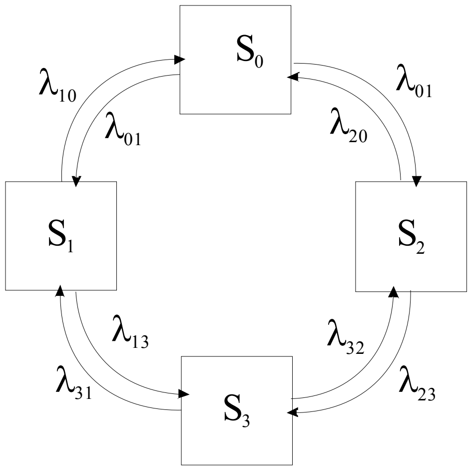


Рис. 2.9. Граф состояний системы ***S***

Рассмотрим систему в момент t и, задав малый промежуток Δt, найдем вероятность $p_0(t+\Delta t)$ того, что система в момент $t+\Delta t$ будет находиться в состоянии S_0.

Предположим, что система в момент t с вероятностью $p_0(t)$ находилась в состоянии S_0, и за время Δt не вышла из этого состояния. Тогда вывести систему из него можно с помощью суммарного простейшего потока с интенсивностью $(\lambda_{01}+\lambda_{02})$. Вероятность того, что система не выйдет из состояния S_0, равна $[1-(\lambda_{01}+\lambda_{02})\Delta t]$. Таким образом, вероятность того, что система будет находиться в состоянии S_0 и не выйдет из него за время Δt, равна

$$[1-(\lambda_{01}+\lambda_{02})\Delta t]\cdot p_0(t).$$

С другой стороны, система в момент t с вероятностями $p_1(t)$ и $p_2(t)$ находилась в состоянии S_1 и S_2 и за время Δt перешла в состояние S_0 соответственно.

Потоком с интенсивностью λ_{10} система будет переведена из состояния S_1 в состояние S_0 (или λ_{20} из S_2 в состояние S_0) с вероятностью, приближенно равной $\lambda_{10}\Delta t$ (или $\lambda_{20}\Delta t$). В этом случае вероятность того, что система будет находиться в состоянии S_0, равна $\lambda_{10}\Delta t\cdot p_1(t)$ (или $\lambda_{20}\Delta t\cdot p_2(t)$). Применяя теорему сложения вероятностей, получим

$$p_0(t+\Delta t)=\lambda_{10}\Delta t\cdot p_1(t)+\lambda_{20}\Delta t\cdot p_2(t)+[1-(\lambda_{01}+\lambda_{02})]\Delta t\cdot p_0(t),$$

откуда

$$\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t}=\lambda_{10}\cdot p_1(t)+\lambda_{20}\cdot p_2(t)-(\lambda_{01}+\lambda_{02})\cdot p_0(t). \qquad (2.26)$$

Переходя к пределу, при $\Delta t \to 0$ получим в левой части уравнения производную $\dot{p}_0(t)$.

Рассуждая аналогично для других состояний системы ***S***, можно получить систему дифференциальных уравнений Колмогорова для вероятностей состояний:

$$\begin{cases}\dot{p}_0(t)=\lambda_{10}\cdot p_1(t)+\lambda_{20}\cdot p_2(t)-(\lambda_{01}+\lambda_{02})\cdot p_0(t)\\ \dot{p}_1(t)=\lambda_{01}\cdot p_0(t)+\lambda_{31}\cdot p_3(t)-(\lambda_{10}+\lambda_{13})\cdot p_1(t)\\ \dot{p}_2(t)=\lambda_{02}\cdot p_0(t)+\lambda_{32}\cdot p_3(t)-(\lambda_{20}+\lambda_{23})\cdot p_2(t)\\ \dot{p}_3(t)=\lambda_{13}\cdot p_1(t)+\lambda_{23}\cdot p_2(t)-(\lambda_{31}+\lambda_{32})\cdot p_3(t)\end{cases} \qquad (2.27)$$

Таким образом, правило составления уравнений Колмогорова заключается в следующем:

1) в левой части каждого из уравнений стоит производная вероятности *i* состояния,
2) в правой части — сумма произведений вероятностей всех состояний (из которых идут стрелки в данное состояние) на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность данного (*i* состояния).

В системе (2.27) независимых уравнений на единицу меньше общего числа уравнений. Поэтому для решения системы необходимо добавить уравнение (2.28):

$$\sum_{i=0}^{3} p_i(t)=1. \qquad (2.28)$$

Особенностью решения дифференциальных уравнений является то, что требуется задать начальные условия. В данном случае вероятности состояний системы в начальный момент t = 0. Для данного случая систему уравнений

(2.27) естественно решать при условии, что в начальный момент система находилась в состоянии S_0, т. е. при начальных условиях

$$p_0(t=0)=1,\ \ p_1(t=0)=p_2(t=0)=p_3(t=0)=0\,.$$

В теории случайных процессов доказывается, что если число состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое состояние, то предельные вероятности существуют.

Предельная вероятность *состояния* ***S*** *показывает среднее относительное время пребывания системы в этом состоянии.*

Так как предельные вероятности постоянны, то, заменяя в уравнениях Колмогорова их производные нулевыми значениями, получим систему линейных алгебраических уравнений, описывающих стационарный режим. Для системы **S** с графом состояний, изображенном на рис. 2.27, такая система уравнений имеет вид

$$\begin{cases}(\lambda_{01}+\lambda_{02})\cdot p_0(t)=\lambda_{10}\cdot p_1(t)+\lambda_{20}\cdot p_2(t)\\(\lambda_{10}+\lambda_{13})\cdot p_1(t)=\lambda_{01}\cdot p_0(t)+\lambda_{31}\cdot p_3(t)\\(\lambda_{20}+\lambda_{23})\cdot p_2(t)=\lambda_{02}\cdot p_0(t)+\lambda_{32}\cdot p_3(t)\\(\lambda_{31}+\lambda_{32})\cdot p_3(t)=\lambda_{13}\cdot p_1(t)+\lambda_{23}\cdot p_2(t)\end{cases}\qquad(2.29)$$

Пример. *Техническая система* ***S*** *состоит из двух узлов — 1 и 2, каждый из них независимо друг от друга может отказать. Поток отказов первого узла — пуассоновский, с интенсивностью λ_1; второго — также пуассоновский, с интенсивностью λ_2. Каждый узел сразу после отказа начинает ремонтироваться (восстанавливается). Поток восстановлений для обоих узлов пуассоновский с интенсивностью λ. Составить граф состояний и написать уравнения Колмогорова для вероятностей состояний. Определить, при каких начальных условиях нужно решать эти уравнения, если в начальный момент времени (t = 0) система была исправна.*

Решение. Определим состояния системы ***S***:

S_{11} — оба узла исправны;

S_{21} — первый узел в ремонте, второй — исправен;

S_{12} — первый узел исправен, второй — в ремонте;
S_{22} — оба узла в ремонте.

Граф состояний в таком случае будет иметь вид, изображенный на рис. 2.10

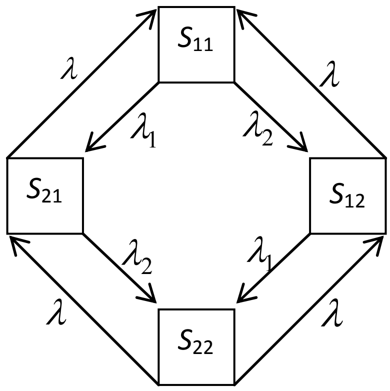


Рис. 2.10. Граф состояний системы ***S***

Система уравнений Колмогорова, соответствующая графу на рис. 2.10, имеет вид

$$
\begin{aligned}
\dot{p}_{11} &= -(\lambda_1 + \lambda_2)p_{11} + \lambda(p_{21} + p_{12}),\\
\dot{p}_{21} &= -(\lambda + \lambda_2)p_{21} + \lambda_1 p_{11} + \lambda p_{22},\\
\dot{p}_{12} &= -(\lambda + \lambda_1)p_{12} + \lambda_2 p_{11} + \lambda p_{22},\\
\dot{p}_{22} &= -2\lambda p_{22} + \lambda_1 p_{12} + \lambda_2 p_{21}.
\end{aligned}
$$

Начальные условия: $p_{11}(t{=}0) = 1$, $p_{12}(t{=}0){=}\, p_{21}(t{=}0){=}\, p_{22}(t{=}0){=}\, 0$.

Показатели эффективности систем массового обслуживания

Классификация целевых назначений теории массового обслуживания основана на выделении в ее структуре различных классов задач и, соответственно, областей применения получаемых результатов. Можно выделить следующие классы задач:

1) задачи анализа поведения системы;
2) статистические задачи;
3) операционные задачи.

Области приложений теории массового обслуживания весьма разнообразны. К их числу, например, относятся транспортные системы, электронно-вычислительные комплексы, система медицинского обслуживания и т. п. В каждой из них возникают задачи, относящиеся к указанным выше классам.

Задачи анализа поведения системы

Цель рассмотрения задач такого рода заключается в том, чтобы с помощью математических моделей, более или менее детально отражающих свойства реальных систем массового обслуживания, выявить операционные характеристики, определяющие поведение этих систем в процессе их функционирования. К числу основных операционных характеристик любой системы массового обслуживания относятся следующие:

— длина очереди в момент времени t, т. е. число требований, находящихся в системе в момент времени t;

— длина очереди на n стадии; при этом предполагается, что стадии реализуются в дискретном режиме и определяются теми или иными событиями (например, поступлением требования в систему или выбытием требования из нее);

— виртуальная продолжительность ожидания относительно момента времени t, т. е. время ожидания обслуживания для требования, которое поступит в систему в момент времени t;

— продолжительность периода, в течение которого n требование ожидает обслуживания;

— продолжительность периода занятости системы, начало которого соответствует t=0 i, т. е. длина периода занятости системы, начинающегося при наличии в системе i требований.

Наряду с указанными операционными характеристиками в исследованиях по теории массового обслуживания используются и их различные модификации, такие как полная продолжительность пребывания требования в системе, операционный цикл (сумма продолжительностей периода занятости и непосредственно следующего за ним периода простоя), суммарное полезное время (доля времени, в течение которого обслуживающий механизм находится в состоянии загруженности).

Многие системы массового обслуживания обладают тем свойством, что по истечении определенного времени их поведение в некотором смысле стабилизируется. Когда поведение системы полностью стабилизируется (т. е. перестает меняться с течением времени), говорят, что процесс ее функционирования будет стационарным, или равновесным. Условия, при которых системы переходят в стационарное состояние, представляют значительный интерес. На практике удобнее оперировать простыми и легко измеримыми показателями, нежели сложными формулами для распределений вероятностей и моментов. Выбор показателей такого рода, разумеется, зависит от характера исследуемой системы и целей анализа. В этой связи можно подчеркнуть, что наиболее часто используемым показателем служит степень загруженности обслуживающего прибора, определяемая коэффициентом нагрузки системы или приведенной интенсивностью:

$$\rho = \frac{\textit{интенсивность поступления требований}}{\textit{интенсивность обслуживания}} = \frac{\lambda}{\mu}. \qquad (2.30)$$

Статистические задачи

Статистическое исследование — неотъемлемая часть разработки математической модели реальной системы. В общем виде модель может существовать сама по себе, но приведение ее в количественное соответствие с конкретной системой массового обслуживания достигается путем статистического анализа эмпирических данных, оценивания фигурирующих в модели параметров и проверок исходных гипотез.

Параметрами системы, по существу, являются параметры, ассоциированные с процессом поступления требований и механизмом обслуживания (либо некоторые функции этих параметров).

Операционные задачи

Существование в теории массового обслуживания задач операционной направленности позволяет считать эту теорию одним из разделов исследования операций. Разумеется, некоторые из операционных задач по своей природе относятся к разряду статистических. Другие операционные задачи возникают при проектировании систем массового обслуживания, управлении реальными системами массового обслуживания и оценках их эффективности. Таким образом, при постановке операционных задач следует различать описательный и нормативный подходы. В первом случае описание системы через ее операционные характеристики (характеристики поведения) используется для принятия решений относительно режима функционирования данной системы. С другой стороны, путем математического моделирования протекающих в системе процессов устанавливаются нормативные требования по обеспечению эффективной работы системы.

Аналитическое исследование СМО наиболее простое, если все потоки событий, переводящие ее из состояния в состояние, простейшие (стационарные пуассоновские). Для СМО это допущение означает, что как поток заявок, так и поток обслуживания — простейшие.

*Под **потоком обслуживания** понимается поток заявок, обслуживаемых одна за другой одним непрерывно занятым каналом.*

Этот поток оказывается простейшим, только если время обслуживания заявки $T_{обсл}$ представляет собой случайную величину, имеющую показательное распределение. Параметр этого распределения μ есть величина, обратная среднему времени обслуживания: $\mu = \dfrac{1}{t_{обсл}}$, где $t_{обсл} = M[T_{обсл}]$. Если все потоки событий простейшие, то процесс, протекающий в СМО, представляет собой марковский случайный процесс с дискретными состояниями и непрерывным временем. При выполнении некоторых условий для этого процесса существует финальный стационарный режим, при котором как вероятности состояний, так и другие характеристики процесса не зависят от времени.

Показатели эффективности СМО делятся на характеризующие качество, условия работы обслуживающей системы и отражающие экономические особенности системы.

Показатели первой группы (качество и условия работы обслуживающей системы) обычно формируют на основе полученных из расчетов значений вероятностей состояний системы. Показатели второй группы (отражающие экономические особенности системы) рассчитывают на основе показателей первой группы.

Среди показателей первой группы можно выделить следующие:

1) среднее число заявок А, обслуживаемое СМО в единицу времени, или *абсолютная пропускная способность* СМО;
2) вероятность обслуживания поступившей заявки Q или *относительная пропускная способность* СМО, $Q = \frac{A}{\lambda}$;
3) *вероятность отказа* $P_{отк}$, т. е. вероятность того, что поступившая заявка не будет обслужена, получит отказ, $P_{отк} = 1 - Q$;
4) *среднее число заявок* в СМО (обслуживаемых или ожидающих в очереди) $\bar{z}$;
5) *среднее число заявок в очереди* $\bar{r}$ *;*
6) *среднее время пребывания заявки в СМО* (в очереди или под обслуживанием) $\bar{t}_{сист}$;
7) *среднее время пребывания заявки в очереди* $\bar{t}_{оч}$;
8) *среднее число занятых каналов* $\bar{k}$.

Рассмотрим далее основные виды СМО.

Лекция 2.2.5. Многоканальные СМО: многоканальные и одноканальные системы массового обслуживания

Многоканальная система массового обслуживания ***S*** состоит из ***n*** каналов. На вход системы поступают заявки, при этом интенсивность входящего потока равна λ. Заявка, попадающая в систему ***S,*** начинает обслуживаться, интенсивность потока обслуживания равна μ. Если все каналы заняты обслуживанием заявок, вновь пришедшая заявка отклоняется. В этом случае СМО имеет n+1 состояние:

S_0 — система свободна,

S_1 — один канал занят,

S_2 — два канала заняты,

…,

S_n — все каналы системы заняты обслуживанием заявок, поступивших в систему.

Графически данную СМО можно представить в виде графа состояний (рис. 2.11).

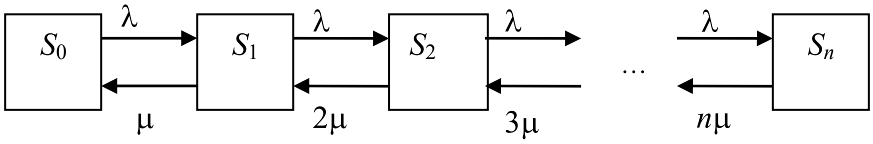


Рис. 2.11. Размеченный граф состояний многоканальной СМО с отказами

Приведенная интенсивность потока заявок

$$\rho = \frac{\lambda}{\mu}.$$

Финальные вероятности состояний

$$p_0 = \left[\sum_{i=0}^{n} \frac{\rho^i}{i!} \right]^{-1}, \; p_k = \frac{\rho^k}{k!} p_0, \, k = 1, 2, \ldots n.$$

Вероятность отказа

$$p_{отк} = p_n = \frac{\rho^n}{n!} p_0 .$$

Относительная пропускная способность

$$q = 1 - p_{отк} = 1 - p_n.$$

Абсолютная пропускная способность

$$A = \lambda\, q = \lambda\,(1 - p_n).$$

Среднее число занятых обслуживанием каналов

$$\tilde{k} = \frac{A}{\mu} = \frac{\lambda}{\mu}(1 - p_n) = \sum_{i=1}^{n} i \cdot p_i .$$

Среднее число заявок в системе

$$\tilde{z}_s = \tilde{r} + \tilde{k} = \tilde{k} .$$

Среднее время обслуживания

$$\tilde{t}_{обс} = \frac{\tilde{k}}{\lambda} .$$

Пример. *Имеется мини-АТС с тремя телефонами. Если все телефоны (каналы) заняты, то внешний звонок отклоняется. Среднее время обслуживания одной заявки каналом* μ *равно двум минутам. Поток заявок простейший с интенсивностью* $\lambda = 1{,}5\ \frac{\text{заяв.}}{\text{мин.}}$. *Составить граф состояний. Найти финальные вероятности состояний и основные характеристики эффективности СМО.*

Решение. Данная СМО будет иметь четыре состояния:

S_0 — все три канала связи свободны,

S_1 — два канала связи свободны, а один — занят,

S_2 — один канал связи свободен, а два — заняты,

S_3 — все три канала связи заняты.

Поскольку среднее время обслуживания одной заявки каналом равно 2 минутам, то

$$\mu = \frac{1}{\tilde{t}_{обсл.}} = \frac{1}{2} = 0{,}5\left(\frac{заяв.}{мин.}\right).$$

Граф состояний изображен на рис. 2.12.

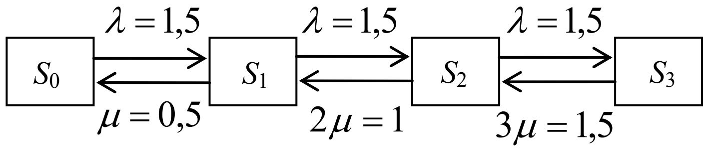


Рис. 2.12

Вычислим основные показатели СМО.

Финальные вероятности системы:

$$p_0 = \left\{1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \frac{\rho^3}{3!}\right\}^{-1}, \quad p_i = \frac{\rho^i}{i!}p_0, \quad i = 1,\ 2,\ 3,$$

$$\rho = \frac{\lambda}{\mu} = \lambda \cdot \tilde{t}_{обсл.} = 1{,}5 \cdot 2 = 3.$$

$$p_0 = \frac{1}{13}; \qquad p_1 = \frac{3}{13}; \qquad p_2 = \frac{9}{26}; \qquad p_3 = \frac{9}{26}.$$

Вероятность отказа $p_{отк.} = \frac{\rho^3}{3!}p_0 = \frac{9}{26}$.

Вероятность обслуживания $q = 1 - p_{отк} = \frac{17}{26}$.

Абсолютная пропускная способность $A = \lambda \cdot q = 1{,}5 \cdot \frac{17}{26} \approx 0{,}981$.

Среднее число занятых каналов $\tilde{k} = \frac{\text{A}}{\mu} = 1{,}96$.

Среднее время пребывания заявки в системе $\widetilde{t}_{сис} = \frac{\widetilde{k}}{\lambda} = \frac{1{,}96}{1{,}5} = 1{,}3$ (мин).

Многоканальные системы массового обслуживания с ограниченной длиной очереди

Многоканальная система массового обслуживания ***S*** имеет ***n*** каналов и ***m*** мест в очереди. На вход системы поступают заявки, при этом интенсивность входящего потока равна λ. Заявка, попадающая в систему ***S,*** начинает обслуживаться, интенсивность потока обслуживания равна μ. Если все каналы заняты обслуживанием заявок, вновь пришедшая заявка ставится в очередь ожидания обслуживания заявок. Если все места в очереди заняты, то вновь пришедшая заявка отклоняется. В этом случае СМО имеет $n+m+1$ состояние:

S_0 — система свободна;
S_1 — один канал занят, очереди нет;
S_2 — два канала заняты, очереди нет;
…;
S_n — все каналы системы заняты обслуживанием заявок, поступивших в систему, очереди нет;
S_{n+1} — все каналы системы заняты и занято одно место в очереди;
S_{n+2} — все каналы системы заняты и заняты два места в очереди
…;
S_{n+m} — все каналы системы и места в очереди заняты.

Графически СМО можно представить следующим образом (рис. 2.13).

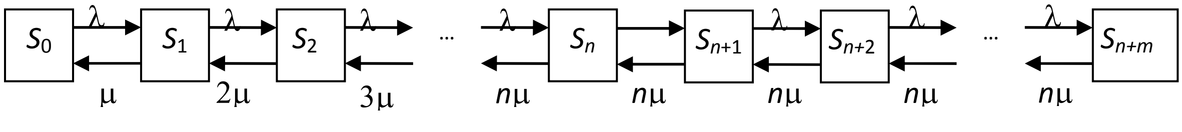


Рис. 2.13. Размеченный граф состояний многоканальной СМО с ограниченною очередью

Приведенная интенсивность потока заявок

$$\rho = \frac{\lambda}{\mu}.$$

Финальные вероятности состояний

$$p_0 = \left[\sum_{i=0}^{n} \frac{\rho^i}{i!} + \sum_{i=n+1}^{n+m} \frac{\rho^i}{n! \cdot n^{i-n}} \right]^{-1},$$

Для случая $\omega = \frac{\rho}{n} < 1$ формула финальной вероятности упрощается:

$$p_0 = \left[\sum_{i=0}^{n} \frac{\rho^i}{i!} + \frac{\rho^{n+1}}{n! \cdot n} \cdot \frac{1-\omega^m}{1-\omega} \right]^{-1},$$

$$p_k = \frac{\rho^k}{k!} p_0, \; k = \overline{1,n}, \qquad p_{n+k} = \frac{\rho^{n+k}}{n! \cdot n^k} p_0, \; 1 \le k \le m.$$

Вероятность отказа

$$p_{отк} = p_{n+m} = \frac{\rho^{m+n}}{n^m \cdot n!} p_0.$$

Относительная пропускная способность

$$q = 1 - p_{отк.}$$

Абсолютная пропускная способность

$$A = \lambda\, q = \lambda\, (1 - p_{n+m}).$$

Среднее число занятых обслуживанием каналов

$$\tilde{k} = \frac{A}{\mu} = \rho\left(1 - p_{n+m}\right).$$

Среднее число заявок в очереди ($\omega = \frac{\rho}{n} < 1$)

$$\tilde{r} = \frac{p_0 \rho^{n+1}}{n \cdot n!} \frac{1 - (m+1) \cdot \omega^m + m\omega^{m+1}}{(1-\omega)^2}.$$

Среднее число заявок в системе

$$\tilde{z}_s = \tilde{r} + \tilde{k} \,.$$

Среднее время обслуживания

$$\tilde{t}_{обс} = \frac{\tilde{k}}{\lambda}.$$

Среднее время в очереди

$$\tilde{t}_{оч} = \frac{\tilde{r}}{\lambda}.$$

Среднее время в системе

$$\tilde{t}_{сис} = \frac{\tilde{k}}{\lambda} + \frac{\tilde{r}}{\lambda}.$$

Пример. *На автозаправочной станции установлены три колонки. Около станции находится площадка на три машины для ожидания в очереди. На станцию прибывает в среднем две машины в минуту. Среднее время заправки одной машины минута. Требуется определить вероятность отказа и среднюю длину очереди.*

Решение. Так как

$$n = 3,\ m = 3,\ \lambda = 2,\ \tilde{t}_{обс} = 1, \text{ то } \mu = 1,\ \rho = \frac{2}{1} = 2,\ \omega = \frac{2}{3} < 1,$$

а финальная вероятность равна

$$p_0 = \left[1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{3 \cdot 3!} \cdot \frac{1 - \left(\frac{2}{3}\right)^3}{1 - \frac{2}{3}} \right]^{-1} \approx 0{,}122 \,,$$

Вероятность отказа равна вероятности пребывания в последнем состоянии S_{n+m}=S_6:

$$p_{отк} = p_{m+n} = \frac{\rho^{m+n}}{n^m \cdot n!} p_0 = \left(\frac{2}{3}\right)^3 \frac{2^3}{3!} 0.122 \approx 0{,}048.$$

Средняя длина очереди

$$\tilde{r} = \frac{p_0 \rho^n}{n!} \sum_{i=1}^{m} i\, \omega^i = \frac{0.122 \cdot 2^3}{3!} \cdot \left[\frac{2}{3} + 2\left(\frac{2}{3}\right)^2 + 3\left(\frac{2}{3}\right)^3\right] = 0{,}35.$$

Многоканальные системы массового обслуживания с ожиданием

Многоканальная система массового обслуживания ***S*** имеет ***n*** каналов и бесконечную очередь. На вход системы поступают заявки, при этом интенсивность входящего потока равна λ. Заявка, попадающая в систему ***S,*** начинает обслуживаться, интенсивность потока обслуживания равна μ. Если все каналы заняты обслуживанием заявок, вновь пришедшая заявка ставится в очередь. В этом случае СМО имеет бесконечное число состояний:

S_0 — система свободна;
S_1 — один канал занят, очереди нет;
S_2 — два канала заняты, очереди нет;
…;
S_n — все каналы системы заняты обслуживанием заявок, поступивших в систему, очереди нет;
S_{n+1} — все каналы системы заняты и занято одно место в очереди;
S_{n+2} — все каналы системы заняты и заняты два места в очереди;
...

В этом случае СМО можно представить в виде графа состояний (рис. 2.14).

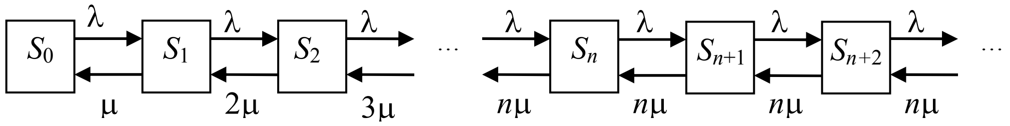


Рис. 2.14. Размеченный граф состояний многоканальной СМО с бесконечной очередью

Финальные вероятности состояний

$$p_0 = \left[\sum_{i=0}^{n} \frac{\rho^i}{i!} + \frac{\rho^{n+1}}{n \cdot n!} \cdot \frac{1}{1-\omega} \right]^{-1} \text{ при } \omega = \frac{\rho}{n} < 1,$$

$$p_k = \frac{\rho^k}{k!} p_0, \; k = \overline{1,n}, \qquad p_{n+k} = \frac{\rho^{n+k}}{n! \cdot n^k} p_0, \quad 1 \le k \le m.$$

Вероятность отказа

$$p_{отк} = 0.$$

Относительная пропускная способность

$$q = 1.$$

Абсолютная пропускная способность

$$A = \lambda.$$

Среднее число занятых обслуживанием каналов

$$\tilde{k} = \rho.$$

Среднее число заявок в очереди (ω<1)

$$\tilde{r} = \frac{\omega \cdot p_0}{(1-\omega)^2} = \frac{\rho^{n+1} \cdot p_0}{n \cdot n! \cdot (1-\omega)^2}.$$

Среднее число заявок в системе

$$\tilde{z}_s = \tilde{r} + \tilde{k}.$$

Среднее время обслуживания

$$\tilde{t}_{обс} = \frac{1}{\mu}.$$

Среднее время в очереди

$$\tilde{t}_{оч} = \frac{\tilde{r}}{\lambda}.$$

Среднее время в системе

$$\widetilde{t}_{сис} = \frac{\widetilde{k}}{\lambda} + \frac{\widetilde{r}}{\lambda}.$$

Одноканальные системы массового обслуживания с отказами

Одноканальная система массового обслуживания S состоит из одного канала. На вход системы поступают заявки, при этом интенсивность входящего потока равна λ. Заявка, попадающая в систему S**,** начинает обслуживаться, интенсивность потока обслуживания равна μ. Если канал занят обслуживанием заявок, вновь пришедшая заявка отклоняется. В этом случае СМО имеет два состояния:

S_0 — система свободна,

S_1 — система занята.

Графически СМО можно представить следующим образом (рис. 2.15).

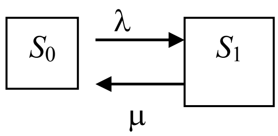


Рис. 2.15. Размеченный граф состояний одноканальной СМО с отказами

Финальные вероятности состояний

$$p_0 = \frac{1}{1+\rho}, \quad p_1 = \frac{\rho}{1+\rho}.$$

Вероятность отказа

$$p_{отк} = p_1 = \frac{\rho}{1+\rho}.$$

Относительная пропускная способность

$$q = \frac{1}{1+\rho}.$$

Абсолютная пропускная способность

$$A = \frac{\lambda}{1+\rho}.$$

Среднее число занятых обслуживанием каналов

$$\tilde{k} = \frac{\rho}{1+\rho}.$$

Среднее число заявок в очереди

$$\tilde{r} = 0.$$

Среднее число заявок в системе

$$\tilde{z}_s = \tilde{k}.$$

Среднее время обслуживания системой

$$\tilde{t}_s = \frac{\tilde{k}}{\lambda}.$$

Среднее время ожидания в очереди

$$\tilde{t}_{оч} = 0.$$

Пример. *Известно, что заявки на телефонные переговоры в телевизионном ателье поступают с интенсивностью $\lambda = 90$ заявок в час, а средняя продолжительность переговоров по телефону составляет две минуты. Определить показатели эффективности работы СМО при наличие одного телефонного аппарата.*

Решение. По условию задачи имеем $\lambda = 90\left(\frac{1}{ч}\right)$, $\tilde{t}_{обс} = 2(мин)$. Тогда интенсивность потока обслуживания $\mu = \frac{1}{\tilde{t}_{обс}} = 0{,}5\left(\frac{1}{мин}\right) = 30\left(\frac{1}{ч}\right)$.

Относительная пропускная способность СМО равна

$$q = \frac{30}{90 + 30} = 0{,}25,$$

т. е. в среднем только 25 % поступающих заявок осуществляют переговоры по телефону. Вероятность отказа в этом случае равна 0,75.

Абсолютная пропускная способность СМО равна

$$A = 90 \cdot 0{,}25 = 22{,}5,$$

т. е. в среднем в час будут обслужены 22,5 заявки на переговоры. Следовательно, при одном телефоне ателье плохо справляется с потоком заявок.

Одноканальные системы массового обслуживания с ограниченной длиной очереди

Одноканальная система массового обслуживания S имеет 1 канал и m мест в очереди. На вход системы поступают заявки, при этом интенсивность входящего потока равна λ. Заявка, попадающая в систему S, начинает обслуживаться, интенсивность потока обслуживания равна μ. Если канал занят, то вновь пришедшая заявка ставится в очередь ожидания обслуживания заявок. Если все места в очереди заняты, то вновь пришедшая заявка отклоняется. В этом случае СМО имеет m+2 состояние:

S_0 — система свободна;

S_1 — канал занят, очереди нет;

S_2 — канал занят, занято одно место в очереди;

…;

S_{m+2} — канал и все места в очереди заняты.

Графически данную СМО можно изобразить так (рис. 2.16).

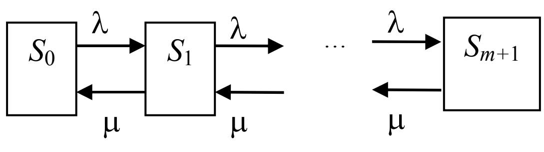


Рис. 2.16. Размеченный граф состояний одноканальной СМО с ограниченною очередью

Финальные вероятности состояний

$$p_0 = \frac{1-\rho}{1-\rho^{m+2}}, \qquad p_k = \rho^k p_0,\ k=1, 2, \ldots m+1.$$

Вероятность отказа

$$p_{отк} = \rho^{m+1} p_0 .$$

Относительная пропускная способность

$$q = 1-\rho^{m+1} p_0 .$$

Абсолютная пропускная способность

$$A = \lambda \cdot \left(1-\rho^{m+1} \cdot p_0\right).$$

Среднее число занятых обслуживанием каналов

$$\tilde{k} = \frac{\rho}{1+\rho}.$$

Среднее число заявок в очереди

$$\tilde{r} = \rho^2 \frac{1-\rho^m \left(m+1-m\rho\right)}{\left(1-\rho^{m+2}\right)\left(1-\rho\right)}.$$

Среднее число заявок в системе

$$\tilde{z}_s = \tilde{k} + \tilde{r} .$$

Среднее время обслуживания системой

$$\tilde{t}_s = \frac{\tilde{z}}{\lambda}.$$

Среднее время ожидания в очереди

$$\tilde{t}_{оч} = \frac{\tilde{r}}{\lambda}.$$

Пример. *В частном стоматологическом кабинете работает один врач. В приемной этого врача имеется три кресла для ожидания. Подсчитать характеристики эффективности данной простейшей одноканальной СМО с тремя местами в очереди при условии, что интенсивность потока заявок равна четырем заявкам в час, а время обслуживания одной заявки — 30 минут. Выяснить, как эти характеристики изменятся, если увеличить число мест в очереди до четырех.*

Решение. По условию задачи имеем

$$\lambda = 4 \quad \left(\frac{заявки}{час}\right), \; \tilde{t}_{обс} = 30 \quad (мин).$$

Тогда интенсивность потока обслуживания

$$\mu = \frac{1}{\tilde{t}_{обс}} = 2 \quad \left(\frac{заявки}{час}\right), \; \rho = \frac{\lambda}{\mu} = 2\,.$$

При $m = 3$ финальные вероятности будут равны

$$p_0 = \frac{1}{31}, \cdots, p_4 = \frac{16}{31}.$$

Зная финальные вероятности, найдем характеристики эффективности системы массового обслуживания:

$$\begin{aligned}
&q \approx 0.484, \\
&A \approx 1.93\left(\frac{заявки}{час}\right), \\
&\tilde{k} \approx 0.968(каналов), \\
&\tilde{r} \approx 2.19(заявки), \\
&\tilde{z} \approx 3.16(заявки), \\
&\tilde{t}_{оч} \approx 0.55(час), \\
&\tilde{t}_{сис} \approx 0.79(час).
\end{aligned}$$

При $m = 4$ финальные вероятности будут равны

$$p_0 = \frac{1}{63}, \cdots, p_5 = \frac{32}{63}.$$

В этом случае характеристики эффективности СМО будут равны:

$$q \approx 0.493,$$

$$A \approx 1.96\left(\frac{заявки}{час}\right),$$

$$\tilde{r} \approx 3.11(заявки),$$

$$\tilde{z} \approx 4.09(заявки),$$

$$\tilde{t}_{оч} \approx 0.78(час),$$

$$\tilde{t}_{сис} \approx 1,02(час).$$

Из полученных данных следует, что увеличение числа мест в очереди с трех до четырех приводит к незначительному увеличению абсолютной и относительной пропускной способности, но при этом происходит некоторое увеличение среднего числа заявок в очереди и в системе в целом, а также соответствующих средних времен.

Одноканальные системы массового обслуживания с ожиданием

Одноканальная система массового обслуживания ***S*** имеет бесконечную очередь. На вход системы поступают заявки, при этом интенсивность входящего потока равна λ. Заявка, попадающая в систему ***S***, начинает обслуживаться, интенсивность потока обслуживания равна μ. Если канал занят обслуживанием заявок, вновь пришедшая заявка ставится в очередь. В этом случае СМО имеет бесконечное число состояний:

S_0 — система свободна;
S_1 — один канал занят, очереди нет;
S_2 — канал занят, занято одно место в очереди;
…

Графически данную СМО можно представить таким образом (рис. 2.17).

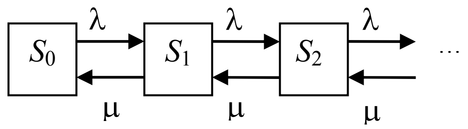


Рис. 2.17. Размеченный граф состояний одноканальной СМО с бесконечной очередью

Финальные вероятности состояний

$$p_0 = 1 - \rho, \quad p_k = \rho^k p_0, k=1, 2, \ldots$$

Вероятность отказа

$$p_{отк} = 0.$$

Относительная пропускная способность

$$q = 1.$$

Абсолютная пропускная способность

$$A = \lambda.$$

Среднее число занятых обслуживанием каналов

$$\tilde{k} = \rho.$$

Среднее число заявок в очереди

$$\tilde{r} = \frac{\rho^2}{1-\rho}.$$

Среднее число заявок в системе

$$\tilde{z}_s = \sum_{k=1}^{\infty} k p_k = (1-\rho) \sum_{k=1}^{\infty} k \rho^k, \text{ для } \rho < 1 \quad \tilde{z}_s = \frac{\rho}{1-\rho}.$$

Среднее время обслуживания системой

$$\tilde{t}_s = \frac{\rho}{\lambda \cdot (1-\rho)}.$$

Среднее время ожидания в очереди

$$\tilde{t}_{оч} = \frac{\rho^2}{\lambda \cdot (1-\rho)}.$$

Пример. *Железнодорожная сортировочная горка, на которую подается простейший поток составов с интенсивностью λ = 2 состава в час, представляет собой одноканальную СМО с неограниченной очередью. Время обслуживания (роспуска) состава на горке имеет показательное распределение со средним значением $\tilde{t}_{обс}$ = 20 мин. Найти финальные вероятности состояний СМО, среднее число составов, связанных с горкой, среднее число составов в очереди, среднее время пребывания состава в СМО, среднее время пребывания состава в очереди.*

Решение. По условию задачи имеем

$$\lambda = 2\left(\frac{состава}{ч}\right),\ \tilde{t}_{обс} = 20\ (мин).$$

Тогда интенсивность потока обслуживания

$$\mu = \frac{1}{\tilde{t}_{обс}} = \frac{1}{3}\left(\frac{состава}{ч}\right),\ \rho = \frac{\lambda}{\mu} = \frac{2}{3}.$$

Финальные вероятности тогда будут равны

$$p_0 = 1 - \frac{2}{3} = \frac{1}{3},$$

$$p_1 = \left(\frac{2}{3}\right)\cdot\left(\frac{1}{3}\right) = \frac{2}{9},$$

$$p_2 = \left(\frac{2}{3}\right)^2\cdot\left(\frac{1}{3}\right) = \frac{4}{27},$$

$$\ldots$$

$$p_k = \left(\frac{2}{3}\right)^k\cdot\left(\frac{1}{3}\right) = \frac{2^k}{9^{k+1}},$$

Зная финальные вероятности, найдем значения среднего числа обслуживаемых составов, среднего числа составов в очереди и соответствующее время пребывания в очереди и в системе в целом:

$$\tilde{z}_s = \tilde{k} = \frac{\rho}{1-\rho} = 2(состава),\quad \tilde{r} = \frac{4}{3}(состава),$$

$$\tilde{t}_s = 1(час),\quad \tilde{t}_{оч} = \frac{2}{3}(час).$$

Лекция 2.2.6. Модели принятия решений в условиях неопределенности и риска

Теория игр и её использование в задачах принятия решений

В экономике очень распространенной является ситуация, когда необходимо принять решение в условиях неопределенности, т. е. если две (или более) сторон преследуют различные цели, а результаты любого действия каждой из сторон зависят от мероприятий партнера. В качестве примера таких конфликтных ситуаций можно привести взаимоотношения между поставщиком и потребителем, покупателем и продавцом, банком и клиентом. Во всех этих примерах конфликтная ситуация порождается различием интересов партнеров и стремлением каждого из них принять оптимальное решение. При этом каждому из них приходится считаться не только со своими целями, но и с целями партнера и учитывать неизвестные заранее решения, которые эти партнеры будут принимать.

Если имеется несколько конфликтующих сторон (лиц), каждое из которых принимает некоторое решение, определяемое заданным набором правил, и каждому из лиц известно возможное конечное состояние конфликтной ситуации с заранее определенными для каждой из сторон платежами, то говорят, что имеет место ***игра.*** Задача теории игр состоит в выборе такой линии поведения данного игрока, отклонение от которой может лишь уменьшить его выигрыш.

*Ситуация называется **конфликтной**, если в ней участвуют стороны, интересы которых полностью или частично противоположны.*

Игра *— это действительный или формальный конфликт, в котором имеются по крайней мере два участника (игрока), каждый из которых стремится к достижению собственных целей.*

*Допустимые действия каждого из игроков, направленные на достижение некоторой цели, называются **правилами игры.***

*Количественная оценка результатов игры называется **платежом.***

*Игра называется **парной**, если в ней участвуют только две стороны.*

*Парная игра называется **игрой с нулевой суммой,** если сумма платежей равна нулю, т. е. если проигрыш одного игрока равен выигрышу второго.*

*Однозначное описание выбора игрока в каждой из возможных ситуаций, при которой он должен сделать личный ход, называется **стратегией** игрока.*

*Стратегия игрока называется **оптимальной**, если при многократном повторении игры она обеспечивает игроку максимально возможный средний выигрыш (или, что то же самое, минимально возможный проигрыш).*

Пусть имеются два игрока, один из которых может выбрать i-ю стратегию из m своих возможных стратегий (i=1,...,m), а второй, не зная выбора первого, выбирает j-ю стратегию из n своих возможных стратегий (j=1,...,n). В результате первый игрок выигрывает величину a_{ij}, а второй проигрывает эту величину.

Из чисел a_{ij} составим матрицу:

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

*Матрица A называется **платежной** (или матрицей игры).*

Игру, определяемую матрицей A, которая имеет m строк и n столбцов, называют игрой размерности $m \times n$. Строки матрицы A соответствуют чистым стратегиям первого игрока, а столбцы — чистым стратегиям второго.

*Число $\alpha = \max_i(\min_j a_{ij})$ называется **нижней ценой игры** или максимином, а соответствующая ему стратегия (строка) максиминной.*

*Число $\beta = \min_j(\max_i a_{ij})$ называется **верхней ценой игры** или минимаксом, а соответствующая ему стратегия (строка) минимаксной.*

ТЕОРЕМА 1. *Нижняя цена игры никогда не превосходит верхней цены игры.*

Если $\alpha = \beta = w$*, то* w *называется* ***ценой игры***.

Игра, для которой $\alpha = \beta$*, называется игрой с* ***седловой точкой***.

Пример. *Найти седловую точку для платежной матрицы A.*

Решение. Седловая точка — пара ($i_o = 3$; $j_o = 1$), при которой

$$w = \alpha = \beta = 2.$$

Заметим, что хотя выигрыш в ситуации (3;3) также равен

$$2 = w = \alpha = \beta,$$

она не будет седловой точкой, т. к. этот выигрыш не является максимальным среди выигрышей третьего столбца.

$$\begin{array}{rcl}
 & & \min\limits_j a_{ij} \\
 & & \| \\
A = & \begin{pmatrix} 1 & -3 & -2 \\ 0 & 5 & 4 \\ 2 & \cancel{3} & 2 \end{pmatrix} & \left.\begin{matrix} -3 \\ 0 \\ 2 \end{matrix}\right\} \max\limits_i \min\limits_j a_{ij} = 2 \\
\max\limits_i a_{ij} = & \underbrace{\begin{matrix} 2 & 5 & 4 \end{matrix}}_{\min\limits_{\mathrm{j}} \max\limits_i a_{ij} = 2} &
\end{array}$$

Пример. *Найти седловую точку для платежной матрицы B.*

<u>Решение.</u> Из анализа матрицы выигрышей видно, что $\alpha < \beta$, т. е. данная матрица не имеет седловой точки. Если игрок 1 выбирает свою чистую максиминную стратегию i = 2, то игрок 2, выбрав свою минимаксную j = 2, проиграет только 20. В этом случае игроку 1 выгодно выбрать стратегию i = 1, т. е. отклониться от своей чистой максиминной стратегии и выиграть

30. Тогда игроку 2 будет выгодно выбрать стратегию j = 1, т. е. отклониться от своей чистой минимаксной стратегии и проиграть 10. В свою очередь игрок 1 должен выбрать свою 2-ю стратегию, чтобы выиграть 40, а игрок 2 ответит выбором 2-й стратегии и т. д.

$$\begin{array}{llcl}
 & & & \min\limits_{j} a_{ij} \\
B = & \begin{pmatrix} 10 & 30 \\ 40 & 20 \end{pmatrix} & \begin{matrix} \to \\ \to \end{matrix} & \left.\begin{matrix} 10 \\ 20 \end{matrix}\right\} \max\limits_{i} \min\limits_{j} a_{ij} = 20 \\
\max\limits_{i} a_{ij} & \begin{matrix} \downarrow & \downarrow \end{matrix} & & \\
 & \underbrace{\begin{matrix} 40 & 30 \end{matrix}} & & \\
 & \min\limits_{j} \max\limits_{i} a_{ij} = 30 & &
\end{array}$$

Если игра, заданная матрицей, не имеет седловой точки, то для нахождения ее решения используются смешанные стратегии.

Вектор, каждая из компонентов которого показывает относительную частоту (вероятность) использования игроком соответствующей чистой стратегии, называется ***смешанной стратегией*** *данного игрока.*

Из этого определения непосредственно следует, что сумма компонент указанного вектора равна единице, а сами компоненты не отрицательны. Чистая стратегия есть частный случай смешанной стратегии. Действительно, если в смешанной стратегии какая-либо i-я чистая стратегия применяется с вероятностью 1, то все остальные чистые стратегии не применяются. И эта i-я чистая стратегия является частным случаем смешанной стратегии. Для соблюдения секретности каждый игрок применяет свои стратегии независимо от выбора другого игрока.

Обозначим смешанную стратегию первого игрока как вектор $X = (x_1, x_2 ..., x_m)$, а второго игрока — как вектор $Y = (y_1, y_2 ..., y_n)$, где $x_i \geq 0 \ (i = \overline{1, m})$, $y_j \geq 0 \ (j = \overline{1, n})$, причем $\sum\limits_{i=1}^{m} x_i = 1$, $\sum\limits_{j=1}^{n} y_j = 1$.

***Средний выигрыш игрока** 1 в матричной игре с матрицей А выражается в виде математического ожидания его выигрышей*

$$\mathrm{E}\,(\mathrm{A}, \mathrm{x}, \mathrm{y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\, x_i\, y_j = \mathrm{x}\ \mathrm{A}\ \mathrm{y}^{\mathrm{T}}.$$

Первый игрок имеет целью за счёт изменения своих смешанных стратегий X максимально увеличить свой средний выигрыш E $(A,\ x,\ y)$, а второй — за счёт своих смешанных стратегий Y стремится сделать E $(A,\ x,\ y)$ минимальным, т.е. для решения игры необходимо найти такие x и y, при которых достигается верхняя цена игры

$$\beta = \min_{y} \max_{x}\ E\,(A,\ x,\ y).$$

Аналогичной должна быть ситуация и для игрока 2, т. е. нижняя цена игры должна быть

$$\alpha = \max_{x} \min_{y}\ E\,(A,\ x,\ y).$$

Подобно играм, имеющим седловые точки в чистых стратегиях, вводится следующее определение: **оптимальными смешанными стратегиями** игроков 1 и 2 называются такие наборы $x^o,\ y^o$ соответственно, которые удовлетворяют равенству

$$\min_{y} \max_{x}\ E\,(A,\ x,\ y) = \max_{x} \min_{y}\ E\,(A,\ x,\ y) = E\,(A,\ x^o,\ y^o) \text{ или}$$

$$w = \sum_{i=1}^{m}\ \sum_{j=1}^{n} a_{ij} x_i^{\,*} y_j^{\,*}\ .$$

*Величина $E\ (A,\ x^o, y^o)$ называется при этом **ценой игры** и обозначается через w.*

*$x^o,\ y^o$ называются **оптимальными смешанными стратегиями**, соответственно, игроков 1 и 2, если они образуют седловую точку:*

$$E\,(A,\ x,\ y^o) \le E\,(A,\ x^o,\ y^o) \le E\,(A,\ x^o,\ y).$$

Оптимальные смешанные стратегии и цена игры называются ***решением матричной игры.***

ТЕОРЕМА 2. *Всякая матричная игра с нулевой суммой имеет решение в смешанных стратегиях.*

ТЕОРЕМА 3. *Для того чтобы число* w *было ценой игры, а вектора* X^* *и* Y^* *— векторами оптимальных стратегий, необходимо и достаточно выполнение следующих неравенств:*

$$\sum_{i=1}^{m} a_{ij} x_i^* \geq w \quad (j = \overline{1,\ n}) \quad \text{и} \quad \sum_{j=1}^{n} a_{ij} y_i^* \geq w \quad (i = \overline{1,\ m}).$$

Если теорема 2 дает ответ на вопрос о существовании решения игры, то теорема 4 позволяет найти решение для игры 2×2, $2 \times n$, $n \times 2$.

ТЕОРЕМА 4. *Если один из игроков применяет оптимальную смешанную стратегию, то его выигрыш равен цене игры w вне зависимости от того, с какими частотами будет применять второй игрок стратегии, вошедшие в оптимальную (в том числе и чистые стратегии).*

Определение оптимальных стратегий и цены игры и составляет процесс нахождения решения игры.

Свойства решений матричных игр

Обозначим через $G\ (X,Y,A)$ игру двух лиц с нулевой суммой, в которой игрок 1 выбирает стратегию $x \in X$, игрок 2 — $y \in Y$, после чего игрок 1 получает выигрыш $A = A\ (x, y)$ за счёт игрока 2.

Стратегия x^1 игрока 1 ***доминирует (строго доминирует)*** *над стратегией x^2, если*

$$A\,(x^1, y) \geq A\,(x^2, y) \quad (A\,(x^1, y) > A\,(x^2, y)),\ y \in Y.$$

Стратегия y^1 игрока 2 ***доминирует (строго доминирует)*** *над стратегией y^2, если*

$$A\,(x, y^1) \leq A\,(x, y^2) \quad (A\,(x, y^1) < A\,(x, y^2)),\ x \in X.$$

При этом стратегии x^2 и y^2 называются ***доминируемыми (строго доминируемыми).***

Спектром смешанной стратегии *игрока в конечной антагонистической игре называется множество всех его чистых стратегий, вероятность которых согласно этой стратегии положительна.*

СВОЙСТВО 1. *Если чистая стратегия одного из игроков содержится в спектре некоторой его оптимальной стратегии, то выигрыш этого игрока в ситуации, образованной данной чистой стратегией и любой оптимальной стратегией другого игрока, равен значению конечной антагонистической игры.*

СВОЙСТВО 2. *Ни одна строго доминируемая чистая стратегия игрока не содержится в спектре его оптимальной стратегии.*

Игра $G' = (X', Y', A')$ называется подыгрой игры $G\,(X,Y,A)$, если $X' \subset X$, $Y' \subset Y$, а матрица A' является подматрицей матрицы A. Матрица A' при этом строится следующим образом. В матрице A остаются строки и столбцы, соответствующие стратегиям X' и Y', а остальные «вычеркиваются». Всё то что «останется» после этого в матрице A и будет матрицей A'.

СВОЙСТВО 3. Пусть G = (X,Y,A) — конечная антагонистическая игра, G′ = (X\x′,Y,A) — подыгра игры G, а x′ — чистая стратегия игрока 1 в игре G, доминируемая некоторой стратегией $\bar{x}$, спектр которой не содержит x′. Тогда всякое решение (x^o, y^o, w) игры G′ является решением игры G.

СВОЙСТВО 4. *Пусть G = (X, Y, A) — конечная антагонистическая игра, G′ = (X,Y\y′,A) — подыгра игры G, а y′ — чистая стратегия игрока 2 в игре G, доминируемая некоторой стратегией $\bar{y}$, спектр которой не содержит y′. Тогда всякое решение игры G′ является решением G.*

СВОЙСТВО 5. *Если для чистой стратегии x′ игрока 1 выполнены условия свойства 3, а для чистой стратегии y′ игрока 2 выполнены условия свойства 4, то всякое решение игры G′ = (X\x′,Y\y′,A) является решением игры G = (X, Y, A).*

СВОЙСТВО 6. *Тройка (x^o, y^o, w) является решением игры G = (X, Y, A) тогда и только тогда, когда (x^o, y^o, кw + a) — решение игры G(X, Y, кA + a), где a — любое вещественное число, а к > 0.*

СВОЙСТВО 7. *Для того чтобы $x^o = (x_1^o, \dots, x_i^o, \dots, x_m^o)$ была* ***оптимальной смешанной стратегией*** *матричной игры с матрицей A и ценой игры w, необходимо и достаточно выполнение следующих неравенств:*

$$\sum_{i=1}^{m} a_{ij}\, x_i^o \ge w, \qquad (j = \overline{1,n}). \tag{2.31}$$

Аналогично для игрока 2: чтобы $y^o = (y_1^o,\ \dots, y_j^o,\ \dots, y_n^o)$ была оптимальной смешанной стратегией игрока 2, необходимо и достаточно выполнение следующих неравенств:

$$\sum_{j=1}^{n} a_{ij}\, y_j^o \le w, \qquad (\mathrm{i} = \overline{1,m}). \tag{2.32}$$

Из последнего свойства вытекает: чтобы установить, являются ли предполагаемые (*x, y*) и *w* решением матричной игры, достаточно проверить, удовлетворяют ли они неравенствам (2.31) и (2.32). С другой стороны, найдя неотрицательные решения неравенств (2.31) и (2.32), совместно со следующими уравнениями

$$\sum_{i=1}^{m} x_i = 1, \qquad \sum_{j=1}^{n} y_j = 1. \tag{2.33}$$

получим решение матричной игры.

Таким образом, решение матричной игры сводится к нахождению неотрицательных параметров решений линейных неравенств (2.31) (2.32) и линейных уравнений (2.33). Однако это требует большого объёма вычислений, которое растёт с увеличением числа чистых стратегий игроков. Например, для матрицы 3×3 имеем систему из 6 неравенств и 2 уравнений. Поэтому в первую очередь следует, по возможности используя свойства 2 и 3, уменьшить число чистых стратегий игроков. Затем необходимо во всех случаях проверить выполнение неравенства

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij}.$$

Если оно выполняется, то игроки имеют чистые оптимальные стратегии (игрок 1 — чистую максиминная, а игрок 2 — чистую минимаксная). В противном случае хотя бы у одного игрока оптимальные стратегии будут смешанными. Для матричных игр небольшого размера эти решения можно найти, применяя свойства 1—5.

Замечание. *Отметим, что исключение доминируемых (не строго) стратегий может привести к потере некоторых решений. Если же исключаются только строго доминируемые стратегии, то множество решений игры не изменится.*

Пример. *Пусть G = (X, Y, A), где X = {1, 2, 3, 4}; Y = {1, 2, 3, 4}, а функция выигрыша A задана следующим образом:*

$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 2C & C & 2C & 3C \\ 3C & 3C/2 & C & 2C \\ 2C & 2C & C & C \\ C & C & C & C/2 \end{pmatrix} \end{array},$$

где C > 0. Требуется найти решение игры.

Решение. Прежде всего заметим, что по свойству 6 достаточно решить игру $G^1 = (X, Y, A)$, где $A^1 = \dfrac{1}{C}A$. В матричной форме игра G^1 определяется матрицей выигрышей

$$A^1 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 2 & 1 & 2 & 3 \\ 3 & 3/2 & 1 & 2 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1/2 \end{pmatrix} \end{array}$$

Элементы четвёртой строки этой матрицы «≤» соответствующих элементов третьей строки, и поэтому третья стратегия игрока 1 доминирует над четвёртой. Кроме того, элементы первого столбца матрицы A^1 «≥» соответствующих элементов второго столбца, следовательно, вторая стратегия игрока 2 доминирует над его первой стратегией.

Далее, из свойства 5 следует, что всякое решение игры

$$G^2 = (X \backslash \{4\},\ Y \backslash \{1\},\ A^1)$$

является решением игры G^1. В матричной форме игру G^2 можно представить матрицей

$$A^2 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 2 & 3 \\ 3/2 & 1 & 2 \\ 2 & 1 & 1 \end{pmatrix} \end{array}.$$

Очевидно, что элементы второй строки «≥» полусуммы соответствующих элементов первой и третьей строк. Кроме того, элементы третьего столбца матрицы A^2 «≥» соответствующих элементов второго столбца. Применяя свойство 5, получим, что всякое решение игры

$$G^3 = (X \setminus \{4,2\},\ Y \setminus \{1,4\},\ A^2)$$

есть решение игры G^2, а следовательно и игры G^1. Игра G^3 определяется матрицей

$$A^3 = \begin{array}{c} \\ 1 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 2 & 3 \end{array} \\ \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \end{array}.$$

Матрица A^3 не имеет седловой точки, так как не выполнено равенство

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij},$$

а игра G^3 не имеет решения в чистых стратегиях, т.е. оптимальные стратегии игроков смешанные. Эти стратегии (в данном случае) легко найти из анализа структуры матрицы A^3. Поскольку матрица A^3 симметрична, можно предположить, что игроки в оптимальной стратегии используют свои чистые стратегии с равными вероятностями.

Действительно, если игрок 1 выбирает с равными вероятностями стратегии 1 и 3, то при применении любой из двух чистых стратегий игроком 2 математическое ожидание выигрыша игрока 1 будет равным либо

$$\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 2 = \frac{3}{2},$$

либо

$$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 1 = \frac{3}{2}.$$

Аналогично, если игрок 2 использует свои чистые стратегии 2 и 3 с равными вероятностями, то математическое ожидание его проигрыша будет равно $\frac{3}{2}$. Следовательно, указанные стратегии являются оптимальными в игре G^3, а величина $\frac{3}{2}$ — значением игры G^3. Из предыдущего следует, что эти стратегии оптимальны и в G^1.

Таким образом, стратегия $X = (\frac{1}{2}, 0, \frac{1}{2}, 0)$ — оптимальная стратегия игрока 1, стратегия $Y = (0, \frac{1}{2}, \frac{1}{2}, 0)$ — оптимальной стратегией игрока 2 в игре G^1, а значение игры G^1 равно $\frac{3}{2}$. В силу свойства 4 решением игры G будет тройка $(X, Y, \frac{3C}{2})$.

Игры порядка 2×2

В общем случае игра 2×2 определяется матрицей

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Прежде всего необходимо проверить, есть ли у данной игры седловая точка. Если да, то игра имеет решение в чистых стратегиях, причем оптимальными стратегиями игроков 1 и 2 будут, соответственно, чистая максиминная и чистая минимаксная стратегии. Если же игра с матрицей выигрышей A не имеет чистых стратегий, то оба игрока имеют только такие оптимальные стратегии, которые используют все свои чистые стратегии с положительными вероятностями. В противном случае один из игроков (например, 1) имеет чистую оптимальную стратегию, а другой — только смешанные. Не ограничивая общности, можно считать, что оптимальной стратегией игрока 1 является выбор с вероятностью 1 первой строки. Далее, по свойству 1 следует, что $a_{11} = a_{12} = w$ и матрица имеет вид

$$\begin{pmatrix} w & w \\ a_{21} & a_{22} \end{pmatrix}.$$

Для матриц такого вида одна из стратегий игрока 2 является доминируемой. Следовательно, по свойству 4 этот игрок имеет чистую стратегию, что противоречит предположению.

Пусть $X = (\xi,\ 1-\xi)$ — оптимальная стратегия игрока 1. Так как игрок 2 имеет смешанную оптимальную стратегию, из свойства 1 получим, что (свойство 7)

$$\begin{cases} a_{11}\xi + a_{21}(1-\xi) = w, \\ a_{12}\xi + a_{22}(1-\xi) = w. \end{cases}$$

Отсюда следует, что при $w \neq 0$ столбцы матрицы A не могут быть пропорциональны с коэффициентом пропорциональности, отличным от единицы. Если же коэффициент пропорциональности равен единице, то матрица A принимает вид $\begin{pmatrix} a_{11} & a_{11} \\ a_{12} & a_{12} \end{pmatrix}$ и игрок 1 имеет чистую оптимальную стратегию (он выбирает с вероятностью 1 ту из строк, элементы которой не меньше соответствующих элементов другой), что противоречит предположению. Следовательно, если $w \neq 0$ и игроки имеют только смешанные оптимальные стратегии, то определитель матрицы A отличен от нуля. Из этого следует, что последняя система уравнений имеет единственное решение.

Решая её, находим

$$\xi = \frac{a_{22} - a_{21}}{a_{11} - a_{12} - a_{21} + a_{22}}; \tag{2.34}$$

$$w = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} - a_{12} - a_{21} + a_{22}}. \tag{2.35}$$

Аналогичные рассуждения приводят нас к тому, что оптимальная стратегия игрока 2 $Y = (\eta,\ 1-\eta)$ удовлетворяет системе уравнений

$$\begin{cases} a_{11}\eta + a_{12}(1-\eta) = w \\ a_{21}\eta + a_{22}(1-\eta) = w \end{cases}. \tag{2.36}$$

Откуда

$$\eta = \frac{a_{22} - a_{12}}{a_{11} - a_{12} - a_{21} + a_{22}}. \qquad (2.37)$$

Графический метод решения игр 2×n и m×2

Рассмотрим графический метод решения матричных игр $2\times n$ и $m\times 2$ метод на примерах.

Пример. *Рассмотрим игру, заданную платёжной матрицей:*

$$\begin{array}{cc} & 2 \\ & \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ 1 \begin{array}{c} A_1 \\ A_2 \end{array} & \begin{pmatrix} 2 & 3 & 11 \\ 7 & 5 & 2 \end{pmatrix} \end{array}$$

Требуется найти ее решение.

Решение. На плоскости xOy введём систему координат и на оси Ox отложим отрезок единичной длины A_1, A_2, каждой точке которого поставим в соответствие некоторую смешанную стратегию игрока 1 (x, 1 – x). В частности, точке A_1 (0;0) отвечает стратегия A_1, точке A_2 (1; 0) — стратегия A_2 и т. д. (рис. 2.18).

В точках A_1 и A_2 восстановим перпендикуляр и на полученных прямых будем откладывать выигрыш игроков. На первом перпендикуляре (в данном случае он совпадает с осью O*y*) отложим выигрыш игрока 1 при стратегии A_1, а на втором — при стратегии A_2. Если игрок 1 применит стратегию A_1, то выиграет при стратегии B_1 игрока 2 — 2, при стратегии B_2 — 3, а при стратегии B_3 — 11. Числам 2, 3, 11 на оси O*x* соответствуют точки B_1, B_2 и B_3.

Если же игрок 1 применит стратегию A_2, то его выигрыш при стратегии B_1 равен 7, при B_2 — 5, а при B_3 — 2. Эти числа определяют точки B'_1, B_2', B_3' на перпендикуляре, восстановленном в точке A_2. Соединяя между собой точки B_1 и B'_1, B_2 и B'_2, B_3 и B'_3 получим три прямые, расстояние до которых от оси Ox определяет средний выигрыш при любом сочетании соответствующих стратегий. Например, расстояние от любой точки отрезка $B_1B'_1$ до оси Ox определяет средний выигрыш w_1 при любом сочетании стратегий A_1 A_2 (с частотами x и $1-x$) и стратегией B_1 игрока 2. Это расстояние равно

$$2x_1 + 6(1 - x_2) = w_1.$$

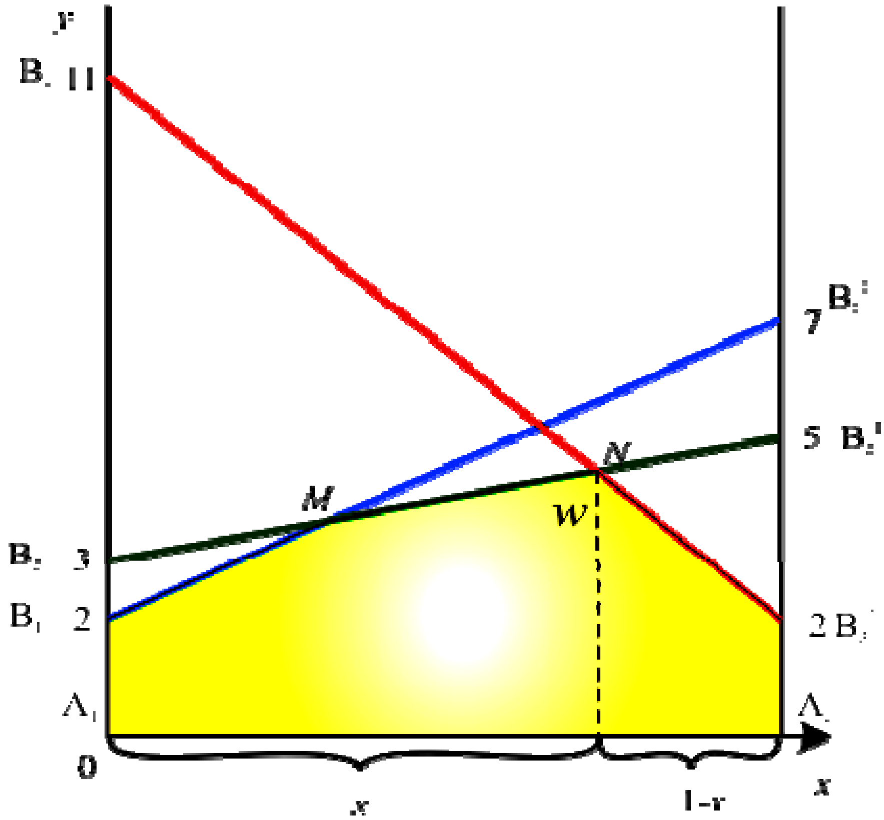


Рис. 2.18. Графическое решение матричной игры

Таким образом, ординаты точек, принадлежащих ломаной $B_1MNB'_3$, определяют минимальный выигрыш игрока 1 при применении им любых смешанных стратегий. Эта минимальная величина является максимальной в точке N; следовательно, этой точке соответствует оптимальная стратегия $X^* = (x,\ 1-x)$, а её ордината равна цене игры w. Координаты точки N находим как точку пересечения прямых $B_2\,B'_2$ и $B_3\,B'_3$.

Соответствующие два уравнения имеют вид

$$\begin{cases} 3x + 5(1-x) = w \\ 11x + 2(1-x) = w \end{cases} \Rightarrow x = \frac{3}{11},\ w = \frac{49}{11}.$$

Следовательно, $X = \left(\dfrac{3}{11}; \dfrac{9}{11}\right)$ при цене игры $w = \dfrac{49}{11}$. Оптимальную стратегию мы можем найти при помощи матрицы $\begin{pmatrix} 3 & 11 \\ 5 & 2 \end{pmatrix}$.

Оптимальные стратегии для игрока 2 можно найти из системы

$$\begin{cases} 3y + 5(1-y) = w \\ 5y + 2(1-y) = w \end{cases} \Rightarrow y = \frac{9}{11}$$

и, следовательно, $Y = \left(0; \dfrac{9}{11}; \dfrac{2}{11}\right)$. Из рис. 2.18 видно, что стратегия B_1 не войдет в оптимальную стратегию.

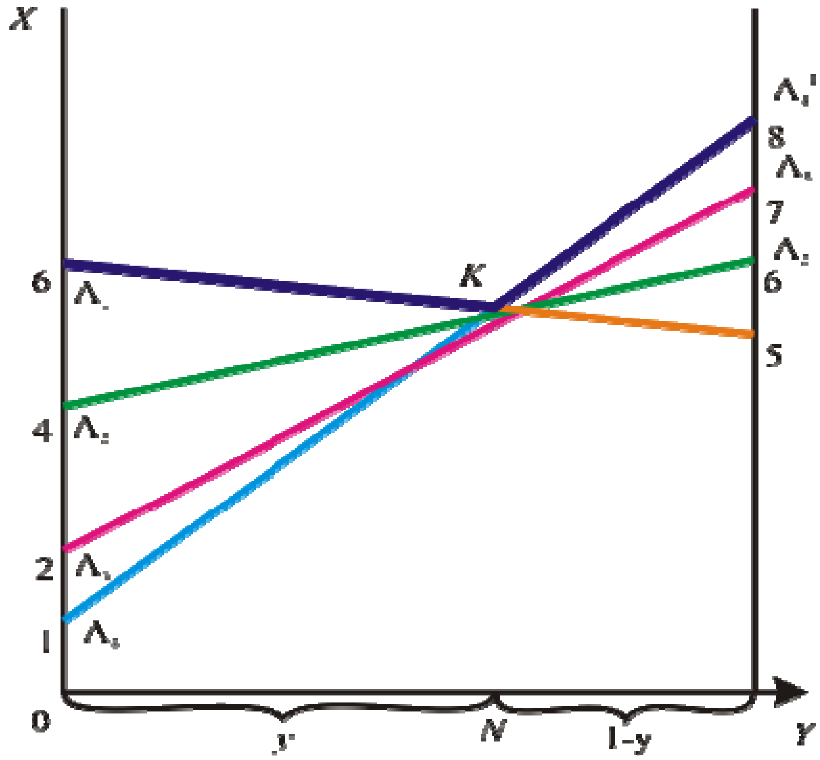


Рис. 2.19. Графическое решение матричной игры

Пример. *Найти решение игры, заданной матрицей*

$$\begin{array}{cc} & 2 \\ & \begin{array}{cc} B_1 & B_2 \end{array} \\ 1 & \begin{array}{c} A_1 \\ A_2 \\ A_3 \\ A_4 \end{array}\begin{pmatrix} 6 & 5 \\ 4 & 6 \\ 2 & 7 \\ 1 & 8 \end{pmatrix} \end{array}$$

Решение. Матрица имеет размерность 2×4. Строим прямые, соответствующие стратегиям игрока 1 (рис. 2.19).

Ломаная $A_1KA'_4$ соответствует верхней границе выигрыша игрока 1, а отрезок NK — цене игры. Решение игры таково:

$$Y = \left(\frac{3}{8}; \frac{5}{8}\right), \qquad X = \left(\frac{7}{8}; 0; 0; \frac{1}{8}\right), \qquad w = \frac{43}{8}.$$

Сведение задач теории игр к задачам линейного программирования

Рассмотрим игру $m \times n$, определяемую матрицей

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

Для оптимальной стратегии (теорема 3) первого игрока $X^* = (x_1^*, x_2^*, x_m^*)$ и цены игры w выполняется неравенство:

$$\sum_{i=1}^{m} a_{ij} x_i^* \geq w \quad (j = \overline{1,\ n}).$$

Предположим для определенности, что $w > 0$.

Это всегда может быть достигнуто благодаря тому, что прибавление ко всем элементам матрицы A одного и того же постоянного числа C не приведет к изменению оптимальных стратегий, а лишь увеличит цену игры на С. Разделив теперь обе части последнего неравенства на w, получим:

$$\frac{\sum_{i=1}^{m} a_{ij} x_i^*}{w} \geq 1 \quad (j = \overline{1, n}).$$

Положим $\frac{x_i^*}{w} = u_i^*$, тогда

$$\sum_{i=1}^{m} a_{ij} u_i^* \geq 1 \quad (j = \overline{1,\ n}),\ u_i^* \geq 0 \quad (i = \overline{1, m}).$$

Используя введенное обозначение, перепишем условие

$$\sum_{i=1}^{m} x_i^* \geq 1 \quad \text{в виде} \quad \sum_{i=1}^{m} u_i^* \frac{1}{w}.$$

Так как первый игрок стремится получить максимальный выигрыш, то он должен обеспечить минимум выражению $\frac{1}{w}$. С учетом этого определение оптимальной стратегии второго игрока сводится к нахождению минимального значения функции

$$F^* = \sum_{i=1}^{m} u_i \text{ при условиях } \sum_{i=1}^{m} a_{ij} u_i \geq 1 \quad (j = \overline{1,\ n}),\ u_i \geq 0 \quad (i = \overline{1, m}). \quad (2.38)$$

Аналогичные рассуждения показывают, что определение оптимальной стратегии второго игрока сводится к нахождению максимального значения функции

$$F^* = \sum_{j=1}^{n} v_j \text{ при условиях } \sum_{j=1}^{n} a_{ij} v_j \leq 1\ (i = \overline{1, m}),\ v_j \geq 0\ \ (j = \overline{1, n}),\ v_j = \frac{y_j}{w}. \quad (2.39)$$

Таким образом, чтобы найти решение данной игры, определяемой матрицей А, нужно составить следующую пару двойственных задач и найти их решение.

Прямая задача: найти минимальное значение функции

$$F^* = \sum_{i=1}^{m} u_i \text{ при условиях } \sum_{i=1}^{m} a_{ij} u_i \geq 1 \quad (j = \overline{1,\ n}),\ u_i \geq 0 \quad (i = \overline{1,m}).$$

Двойственная задача: найти максимальное значение функции

$$F^* = \sum_{j=1}^{n} v_j \text{ при условиях } \sum_{j=1}^{n} a_{ij} v_j \leq 1 \ (i = \overline{1,\ m}),\ v_j \geq 0 \quad (j = \overline{1,n}).$$

Используя решение пары двойственных задач, получаем формулы для определения координат оптимальных смешанных стратегий и цены игры:

$$x_j^* = \frac{u_i^*}{\sum_{i=1}^{m} u_i^*} = w u_i^*, \qquad y_j^* = \frac{v_j^*}{\sum_{j=1}^{n} v_j^*} = w v_j^*, \tag{2.40}$$

$$w = \frac{1}{\sum_{j=1}^{n} v_j^*} = \frac{1}{\sum_{i=1}^{m} u_i^*}; \qquad (i = \overline{1,m}; \quad j = \overline{1,n}). \tag{2.41}$$

Таким образом, алгоритм нахождения решения игры с использованием методов линейного программирования включает следующие этапы:

1. Составить пару двойственных задач линейного программирования, эквивалентных данной матричной игре.
2. Определить планы пары двойственных задач.
3. Используя соотношения между планами пары двойственных задач и оптимальными стратегиями и ценой игры, найти решение игры.

Пример. *Найти решение игры, определяемой матрицей*

$$A = \begin{pmatrix} 0 & 1 & -1 \\ 0 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

Решение. При решении этой игры к каждому элементу матрицы A прибавим 1 и получим следующую матрицу:

$$A_1 = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}$$

Составим теперь пару взаимно-двойственных задач:

$$u_1 + u_2 + u_3 \to \min \qquad\qquad v_1 + v_2 + v_3 \to \max$$

$$\begin{cases} u_1 + u_2 + 2u_3 \ge 1, \\ 2u_1 + \quad\;\; u_3 \ge 1, \\ \qquad\qquad u_3 \ge 1, \\ u_1, u_2, u_3 \ge 0, \end{cases} \qquad\qquad \begin{cases} v_1 + 2v_2 \qquad \le 1, \\ v_1 \qquad + v_3 \le 1, \\ 2v_1 + v_2 \qquad \le 1, \\ v_1, v_2, v_3 \ge 0. \end{cases}$$

Решим двойственную задачу с помощью симплекс-метода (табл. 2.24).

Таблица 2.24

B	***C***	P_0	P_1	P_2	P_3	P_4	P_5	P_6	λ
P_4	0	1	1	2	0	1	0	0	1/0
P_5	0	1	1	0	1	0	1	0	1/1
P_6	0	1	2	1	0	0	0	1	1/0
	Δj	0	–1	–1	–1	0	0	0	
P_4	0	1	1	2	0	1	0	0	1/2
P_3	1	1	1	0	1	0	1	0	1/0
P_6	0	1	2	1	0	0	0	1	1
	Δj	1	0	–1	0	0	1	0	
P_2	1	1/2	1/2	1	0	1/2	0	0	
P_3	1	1	1	0	1	0	1	0	
P_6	0	0	3/2	0	0	–1/2	0	1	
	Δj	3/2	1/2	0	0	1/2	1	0	

Из оптимальной симплекс-таблицы следует, что $(v_1, v_2, v_3) = (0; \frac{1}{2}; 1)$, а из соотношений двойственности — $(u_1, u_2, u_3) = (\frac{1}{2}; 1; 0)$. Значит, цена игры с платёжной матрицей A_1 равна

$$w_1 = \frac{1}{v_1 + v_2 + v_3} = \frac{1}{\frac{1}{2} + 1} = \frac{2}{3} = \frac{1}{u_1 + u_2 + u_3},$$

а игры с платежной матрицей A —

$$w = w_1 - 1 = \frac{2}{3} - 1 = -\frac{1}{3}.$$

При этом оптимальные стратегии игроков имеют вид

$$X = (x_1, x_2, x_3) = (wv_1; wv_2; wv_3) = \left(\frac{2}{3}\cdot\frac{1}{2};\ \frac{2}{3}\cdot 1;\ \frac{2}{3}\cdot 0\right) = \left(\frac{1}{3};\ \frac{2}{3};\ 0\right),$$

$$Y = (y_1, y_2, y_3) = (wu_1; wu_2; wu_3) = \left(\frac{2}{3}\cdot 0;\ \frac{2}{3}\cdot\frac{1}{2};\ \frac{2}{3}\cdot 1\right) = \left(0;\ \frac{1}{3};\ \frac{2}{3}\right).$$

Принятие решений в условиях неопределенности и риска

Принятие решений, касающихся выбора стратегии поведения субъектов экономики, зачастую сталкивается с ситуациями, когда отсутствует достоверная информация о будущих состояниях среды. Однако и тогда необходимо подходить к принятию решений таким образом, чтобы снизить риск возможных негативных последствий в случае осуществления наихудшего из возможных состояний среды.

Методы принятия решений в условиях отсутствия достоверной информации о возможных последствиях изучаются теорией риска. Одна из наиболее важных сфер применения этой теории — оценка инвестиционных проектов.

С точки зрения принятия решений различаются основные типы моделей:

1. Принятие решений в условиях определенности (когда точно известны последствия и исходы любой альтернативы или выбора решения).
2. Принятие решений в условиях риска (когда известны вероятности наступления исходов или последствий для каждого решения).
3. Принятие решений в условиях неопределенности (когда не известны вероятности наступления исходов для каждого решения).

Если имеет место полная неопределенность в отношении возможности реализации состояний среды (т. е. нельзя даже приблизительно указать вероятность наступления каждого возможного исхода), то обстоятельства, с которыми мы имеем дело при выборе решения, можно представить как вид стратегической игры, в которой один игрок — человек, принимающий решения, а другой — некая объективная действительность («природа»), мотивы проявления которой нам неизвестны. Такие ситуации отличаются от задач классической теории игр, поскольку вторая сторона («природа») не имеет каких-либо интересов, что меняет подход к выбору оптимальной стратегии.

Условия игры с «природой» представляются платежной матрицей, в которой строки соответствуют стратегиям человека (обозначим их $A_1, A_2, \ldots, A_m$), а столбцы — стратегиям природы ($B_1, B_2, \ldots, B_n$). Элемент платежной матрицы a_{ij}, находящийся на пересечении i-й стратегии человека и j-й стратегии природы, соответствует выигрышу человека при реализации пары стратегий (A_i, B_j). Указанная платежная матрица имеет следующий вид (табл. 2.25):

Таблица 2.25

Стратегии природы Стратегии человека	B_1	B_2	…	B_n
A_1	a_{11}	a_{12}	…	a_{1n}
A_2	a_{21}	a_{22}	…	a_{2n}
…	…	…	…	…
A_m	a_{m1}	a_{m2}	…	a_{mn}

Элементы платежной матрицы могут принимать как положительные, так и отрицательные значения. В случае, если элементы a_{ij} интерпретируются как доход, то элементы матрицы, как правило, положительны (за исключением случаев, когда при сочетании стратегий наносится убыток). Отрицательные значения элементов a_{ij} характеризуют, в основном, затраты.

При выборе решения в условиях неопределенности обычно используются критерии, рассматриваемые ниже.

Критерий Лапласа

Великий математик Лаплас рассуждал следующим образом: «Я не знаю ничего о будущих состояниях природы; поэтому я могу считать их равновероятными». Исходя из таких рассуждений, задачей человека является выбор такой стратегии, которая позволит ему максимизировать свой выигрыш.

В математическом виде критерий Лапласа записывается следующим образом:

$$\max_i \sum_{j=1}^{n} (1/n) a_{ij} . \qquad (2.42)$$

Критерий Вальда

Согласно статистику и экономисту Вальду, «если мне неизвестны состояния природы, я буду поступать самым осторожным образом, из каждой строки таблицы я извлеку самый маленький результат a_{ij} и выберу строку, в которой находится наибольшее из этих чисел». В случае антагонистической игры этот критерий переходит в критерий фон Неймана; вторым игроком здесь является природа. Из этого следует, что нужно выбирать строку, для которой мы имеем

$$\max_i \min_j a_{ij} . \qquad (2.43)$$

Критерий Гурвица

Согласно Леониду Гурвицу, неразумно, приняв во внимание самый маленький выигрыш, не учитывать самый большой; следует субъективным образом ввести некоторый коэффициент оптимизма α. Пусть в данной строке α — самое маленькое число и А — самое большое. Вычислим теперь для каждой строки $H = \alpha A + (1 - \alpha)A$ и выберем строку, для которой H принимает наибольшее значение. Математическое выражение критерия Гурвица имеет следующий вид:

$$H = \max_i \left\{ \alpha \min_j a_{ij} + (1-\alpha) \max_j a_{ij} \right\}. \qquad (2.44)$$

Заметим, что для $\alpha = 0$ снова получается критерий Вальда (критерий законченного пессимиста), в то время как для $\alpha = 1$ критерий превращается в критерий абсолютного оптимизма.

Критерий Сэвиджа

Л. Дж. Сэвидж рекомендует для каждого из состояний природы найти наиболее благоприятный случай и вычесть соответствующее значение из всех элементов столбца, отвечающего этому состоянию. Такой подсчет дает «сожаление между действительным выбором и наиболее благоприятным, если бы были известны намерения природы». Величина сожаления определяется как

$$r_{ij} = \max_i a_{ij} - a_{ij}. \qquad (2.45)$$

В этом случае выбирается стратегия, при которой величина риска r_{ij} в наихудших условиях минимизируется, т.е. выбирается стратегия, для которой реализуется

$$\max_i \min_j r_{ij}. \qquad (2.46)$$

Пример. *Предприниматель планирует построить небольшую гостиницу в парковой зоне, которая будет использоваться для отдыха и лечения. Площадка под гостиницу уже куплена за 250 000 долл., однако предприниматель не знает, сколько комнат нужно оборудовать в этой гостинице: 20, 30, 40 или 50. Обозначим количество построенных комнат как S, а возможные состояния среды (количество занятых комнат) — как R.*

Решение. Определим необходимые затраты на строительство и эксплуатацию гостиницы.

1. Ежегодные затраты, не зависящие от числа построенных комнат

Благоустройство территории — 100 000 долл. Допускается, что постройка и благоустройство будут длиться в течение 10 лет, указанные затраты будут также погашаться десять лет. Отсюда годовая часть затрат на первичное благоустройство составляет 10 000 долл.

Затраты па ремонт и содержание. Допускается, что затраты составляют фиксированную величину, не зависящую от числа комнат и пропорциональную важности здания. Эта фиксированная часть затрат в год равна 1500 долл.

Один ночной дежурный — 15 долл. в день. Пусть вместе с различными премиями это составляет 6 000 долл. в год.

Один служащий для уборки — 20 долл. в день. Пусть вместе с дополнительной оплатой это 8 000 долл. в год. Стоимость покупки площадки не учитывается, так как стоимость этой недвижимости полагается примерно равной капиталу, который она собой представляет, вложенному в банк с обычными процентами. Итого, общие фиксированные ежегодные затраты 25 500 долл.

2. Ежегодные затраты, пропорциональные числу построенных комнат, в долл. (табл. 2.26)

Таблица 2.26

Количество построенных комнат (S)	20	30	40	50
Постройка, благоустройство, меблировка комнат. Одна комната стоит 4000 долл., практически амортизация длится 10 лет, что дает при различных предложениях	80 000	120 000	160 000	200 000
На 10 комнат полагается одна горничная; ежегодные затраты, включая дополнительные расходы, составляют 6000 долл. на одну горничную	12 000	18 000	24 000	30 000
Содержание и ремонт (пропорциональная часть) — 150 долл. в год на одну комнату	3 000	4 500	6 000	7 500
Страхование на случай пожара — 25 долл. за комнату в год	500	750	1 000	1 250
ИТОГО	95 500	143 250	191 000	238 750

3. Ежегодные затраты, пропорциональные среднему числу занятых комнат R (табл. 2.27)

Таблица 2.27.

Количество занятых комнат (R)	0	10	20	30	40	50
Стирка, уборка — 5 долл. в день на комнату	0	18 000	36 000	54 000	72 000	90 000
Электричество, газ и вода — 5 долл. в день на комнату	0	18 000	36 000	54 000	72 000	90 000
Итого	0	36 000	72 000	108 000	144 000	180 000

Рассмотрим теперь, как будут формироваться доходы предпринимателя от эксплуатации гостиницы. Средняя цена гостиничного номера повышенной комфортности (какие и предполагает оборудовать предприниматель) составляет 60 долл. в день. В зависимости от значений R доходы предпринимателя выражаются в следующих суммах в долл. (табл. 2.28):

Таблица 2.28

R	0	10	20	30	40	50
Доходы	0	219000	438000	657000	876000	1095000

На основании представленных выше значений доходов и затрат можно получить таблицу ежегодного дохода для различных значений R и S. Доход в тыс. долл. (табл. 2.29)

Таблица 2.29

	R=0	R=10	R=20	R=30	R=40	R=50
S=20	–121	62	245	245	245	245
S=30	–168,475	14,25	197,25	380,25	380,25	380,25
S=40	–216,5	–33,5	149,5	332,5	515,5	515,5
S=50	–264,25	–81,25	101,75	284,75	467,75	650,75

Описанная ситуация характеризуется тем, что неизвестно, сколько комнат может быть реально занято. Если принять решение о строительстве 20 комнат, то на этом предприниматель заработает 245 тыс. долл., но может потерпеть убыток в 121 000; для 30 комнат максимально возможный доход, конечно, выше, но убыток также становится большим и т.д. Как принять решение в условиях неопределенности?

Воспользуемся критериями выбора решения в условиях неопределенности, приведенными выше.

Критерий Лапласа. Согласно этому критерию предприниматель считает вероятность наступления различных значений R (0, 10, 20, 30, 40, 50) равной 1/6. При этих условиях для соответствующих значений S вероятностный доход составит:

S = 20 153,5

S = 30 197,25

S= 40 210,5 ←

S= 50 193, 5

Согласно этому критерию предприниматель должен выбрать S = 40. Однако гипотеза о равновероятности не совсем его удовлетворила, поскольку он не считает, что природа к нему враждебна, но все же подозревает, что вероятность среднего числа занятых комнат должна удовлетворять совсем другим законам, а не закону равных вероятностей.

При использовании **критерия Вальда** в нашем случае доход предпринимателя для разных значений S составит:

S = 20 –121; ←

S = 30 –168,75;

S= 40 –216,5;

S= 50 –264,25.

Выбирая S = 20, предприниматель гарантирован от убытка, превышающего 121 000 долл. Однако критерий крайнего пессимизма скорее приведет к выбору решения ничего не строить, поскольку любое помещение капитала рискованно.

Используя **критерий Гурвица**, предприниматель вычислил значение H для различных значений коэффициента α (табл. 2.30):

Таблица 2.30

	α =0,1	α =0,2	α =0,5	α =0,8	α =0,9
S=20	–84,40 ←	–47,80←	62	171,80	206,4
S=30	–113,25	–58,95	105,75	270,45	325,35
S=40	–143,30	–70,10	149,50	369,10	442,30
S=50	–172,75	–81,25	193,25←	467,75←	559,25←

В задаче предпринимателя преобладающее отношение пессимиста приводит к выбору S=20; наоборот, оптимистическое отношение приводит к выбору S = 50.

Для построения матрицы сожалений при использовании **критерия Сэвиджа** в нашем примере нужно вычесть: 121 из столбца R = 0; 62 — из столбца R=10 и т. д. В результате получим матрицу сожалений (табл. 2.31):

Таблица 2.31

	R=0	R=10	R=20	R=30	R=40	R=50
S=20	0	0	0	–135,25	–270,50	–405,75
S=30	–47,75	–47,75	–47,75	0	–135,25	–270,50
S=40	–95,50	–95,50	–95,50	–47,75	0	–135,25
S=50	–143,25	–143,25	–143,25	–95,50	–47,75	0

Минимальные значения сожаления для различных значений, соответственно, равны:

S = 20 – 405, 75;

S = 30 – 270,50;

S = 40 – 135,25;

S = 50 – 143,25.

Оказывается, что, выбирая S=40, предприниматель будет «иметь сожаление», которое не сможет превысить 135,25 (т. е. быть меньшим чем

–135,25).

В результате применения разных критериев предприниматель должен сделать выбор среди следующих различных решений:

а) согласно критерию Лапласа построить 40 комнат;

б) по критерию Вальда построить только 20 комнат;

в) следуя критерию Гурвица, принять число 20, если он пессимист, и 50, если он оптимист;

г) наконец, если применяется критерий Сэвиджа, построить 40 комнат.

Как предпринимателю выбрать решение? Выбор критерия образует как раз высшую форму свободы, которая существует у принимающих экономические решения лиц (разумеется, при условии, что они располагают достаточными средствами, чтобы поставить перед собой подобную задачу). Всякий критерий должен согласовываться с намерениями лица, принимающего решения, и соответствовать его характеру. Как видно из рассмотренного примера, каждый выбор критерия влечет за собой принятие решения, которое может быть совершенно отлично от решения, принятого в соответствии с другим критерием.

Предположим, что предприниматель постарался собрать дополнительную информацию относительно шансов на успех в проектируемом предприятии и получил подтверждение того факта, что существующие аналогичные гостиницы зарегистрировали следующий средний спрос в течение последних лет (табл. 2.32):

Таблица 2.32

Спрос	0	10	20	30	40	50
Вероятность	0,01	0,09	0,20	0,30	0,30	0,1

В условиях частичной определенности о вероятности наступления каждого из состояний среды гораздо легче делать свой выбор. Достаточно вычислить математическое ожидание дохода при каждой гипотезе:

$E(S = 20) = –121 \cdot 0{,}01+62 \cdot 0{,}09+245 \cdot [0{,}2+0{,}3+0{,}3+0{,}1] = 224{,}87;$

$E(S = 30) = –68{,}75 \cdot 0{,}01 + 14{,}25 \cdot 0{,}09 + 197{,}25 \cdot 0{,}2 + 380{,}25 \cdot$
$\cdot [0{,}3+0{,}3+0{,}1]= 305{,}22;$

$E(S = 40) = –216{,}5 \cdot 0{,}01 - 33{,}5 \cdot 0{,}09 + 149{,}5 \cdot 0{,}2 + 332{,}5 \cdot 0{,}3 + 515{,}5 \cdot$
$\cdot [0{,}3 + 0{,}1] = 330{,}675;$

$E(S = 50) = –264{,}25 \cdot 0{,}01 – 81{,}25 \cdot 0{,}09 + 101{,}75 \cdot 0{,}2 + 284{,}75 \cdot 0{,}3 –$

$– 467{,}75 \cdot 0{,}3 + 50{,}75 \cdot 0{,}1 = 301{,}12$

и сравнить эти ожидания. Тогда наиболее благоприятным в среднем решением остается, несомненно, постройка гостиницы из 40 комнат.

Выбор критерия. Обозначим через α субъективную вероятность получить плохие результаты, через γ — вероятность полного успеха; промежуточные ситуации будут оцениваться вероятностью β такой, чтобы $\alpha + \gamma + \beta = 1$.

При данных условиях, если $\sum P$ представляет собой сумму несчастливых исходов, $\sum I$ — сумму промежуточных исходов и $\sum S$ — сумму удовлетворительных результатов, то взвешенное среднее

$$\frac{\alpha \sum P}{m} + \frac{\beta \sum I}{n} + \frac{\gamma \sum S}{p}$$

представляет собой при каждой гипотезе субъективную оценку математического ожидания, где *m, n* и *p* — количества результатов, отнесенных к каждой категории. Основная задача будет состоять в определении того, к какой категории следует отнести каждый результат. Вернемся к таблице значений дохода. Первое решение могло бы состоять в том, чтобы считать несчастливыми результаты, соответствующие убытку, т. е. результаты, которые в таблице отрицательны; но на это можно возразить, что, когда мы берем взвешенное среднее, сумма $\frac{\sum P}{m}$ в случае, например, *S*=40 равна (–216,5 + + 33,5)/2 = –125, тогда как при *S* = 30 она составляет $\frac{\sum P}{m} = 168{,}75$.

Интуитивно ясно, что риск возрастет с увеличением числа построенных комнат. Поэтому принято единственными неудачными исходами считать исходы, находящиеся в первом столбце таблицы (R = 0); небольшие же убытки или небольшие выигрыши будут считаться эквивалентными (например, при R=10 считаем, что 14,25 $\approx$ –33,5).

Успешными исходами в этом решении считаются те, которые соответствуют максимальному спросу R=50. Таким образом, мы приходим к следующему разбиению (табл. 2.33).

Таблица 2.33

R \ S	0	10	20	30	40	50
20						
30						
40						
50						

Неудачные результаты — Промежуточные результаты — Благоприятные результаты

Рассмотрим конкретный пример предпринимателя, предполагая, что он оценивает вероятность разорения в 10 %, но что вероятность успеха он не осмеливается оценить больше чем в 20 %. Каждое решение (построить 20, 30, 40, 50 комнат) обозначим номером I, II, III, IV.

Таблица 2.34

γ												
1	IV											
0,9	IV	IV										
0,8	IV	IV	IV									
0,7	IV	IV	IV	IV								
0,6	IV	IV	IV	IV	IV							
0,5	IV	IV	IV	IV	IV	IV						
0,4	IV	IV	IV	IV	IV	IV	IV					
0,3	IV	IV	IV	IV	IV	IV	IV	IV				
0,2	III	III	III	III	III	III	II	II	I			
0,1	III	III	III	II	II	II	I	I	I	I		
0	II	II	II	II	II	I	I	I	I	I	I	
	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	α

Таким образом, если $\alpha = 0{,}1$; $\gamma = 0{,}2$; то $\beta = 0{,}7$, и можно вычислить:

$\xi(20) = -121 / 1 \cdot 0{,}1 + (62 + 3 \cdot 245) / 4 \cdot 0{,}7 + 245 / 1 \cdot 0{,}2 = 176{,}4$;

$\xi(30) = -168{,}75 / 1 \cdot 0{,}1 + (14{,}25 + 197{,}5 + 2 \cdot 380{,}25) / 4 \cdot 0{,}7 + 380{,}25 / 1 \cdot 0{,}2 = 229{,}3$:

$\xi(40) = -216{,}5 / 1 \cdot 0{,}1 + (-33{,}5 + 149{,}5 + 332{,}5 + 515{,}5) / 4 \cdot 0{,}7 + 515{,}5 / 1 \cdot 0{,}2 = 250{,}2$;

$\xi(50) = -264{,}5 / 1 \cdot 0{,}1 + (-81{,}25 + 101{,}75 + 284{,}75 + 467{,}75) / 4 \cdot 0{,}7 + 650{,}75 / 1 \cdot 0{,}2 = 238{,}8$.

Чтобы усовершенствовать этот метод на плоскости оценок α и γ, можно составить перечень значений этих коэффициентов, изменяющихся от 0 до 1 с интервалом 0,1, и построить карту наиболее благоприятных решений.

Если неудачными событиями считать только те, когда вообще нет клиентов, а в качестве благоприятных событий рассматривать те, при которых все комнаты заняты, то разбиение результатов представляется следующей схемой.

Таблица 2.35

S \ R	0	10	20	30	40	50
20						
30						
40						
50						

Неудачные результаты — Промежуточные результаты — Благоприятные результаты

СПИСОК ЛИТЕРАТУРЫ

1. Абрамов, Л. М. **Математическое программирование.** / *Л. М. Абрамов, В. Ф. Капустин.* — Л.: Изд-во Ленингр. ун-та, 1976. — 184 с.

2. Акулич, И. Л. **Математическое программирование в примерах и задачах:** Учеб. пособие. — 2-е изд., испр. и доп. / *И. Л. Акулич.* — М.: Высш. шк., 1993. — 336 с.

3. Ален, Р. **Математическая экономия.** / *Р. Ален.* — М.: Иностранная лит., 1963. — 667 с.

4. Ашманов, С. А. **Линейное программирование.** / *С. А. Ашманов.* — М.: Наука, 1981. — 340 с.

5. Баканов, М. И. **Экономический анализ: ситуации, тесты, примеры, задачи, выбор оптимальных решений, финансовое прогнозирование:** Учеб. пособие. / *М. И. Баканов, А. Д. Шеремет.* — М.: Финансы и статистика, 1999. — 656 с.

6. Баканов, М. И. **Теория экономического анализа:** Учебник. — 4-е изд., доп. и перераб. / *М. И. Баканов, А. Д. Шеремет.* — М.: Финансы и статистика, 2000. — 416 с.

7. Банди, Б. **Основы линейного программирования:** Пер. с англ. / *Б. Банди.* — М.: Радио и связь, 1989. — 176 с.

8. Вентцель, Е. С. **Инженерные приложения теории вероятностей.** / *Е. С. Вентцель.* — М.: Наука, 1980. — 477 с.

9. Вентцель, Е. С. **Исследование операций: задачи, принципы, методология.** / *Е. С. Вентцель.* — М.: Наука, 2000. — 208 с.

10. Габасов, Р. **Методы линейного программирования.** Ч. 1. Общие задачи. / *Р. Габасов, Ф. М. Кириллова.* — Минск: Изд-во БГУ им. В. И. Ленина, 1977. — 176 с.

11. Габасов, Р. **Методы линейного программирования.** Ч. 2. Транспортные задачи. / *Р. Габасов, Ф. М. Кириллова.* — Минск: Изд-во БГУ им. В. И. Ленина, 1977. — 240 с.

12. Гасс, С. М. **Линейное программирование.** / *С. М. Гасс.* — М.: Физматгиз, 1961. — 304 с.

13. Гермейер, Ю. Б. **Игры с непротивоположными интересами.** / *Ю. Б. Гермейер.* — М.: Наука, 1976. — 327 с.

14. Глухов, В. В. **Математические методы и модели для менеджмента.** / *В. В. Глухов, М. Д. Медников, С. Б. Коробко.* — СПб.: Лань, 2000. — 480 с.

15. Гольштейн, Е. Г. **Линейное программирование, теория, методы и приложения.** / *Е. Г. Гольштейн, Д. Б. Юдин.* — М.: Наука, 1969. — 383 с.

16. Давыдов, Э. Г. **Исследование операций.** / *Э. Г. Давыдов.* — М.: Высш. школа, 1990. — 382 с.

17. Дюбин, Г. Н. **Введение в прикладную теорию игр**. / *Г. Н. Дюбин, В. Г. Суздаль.* — М.: Наука, 1981. — 336 с.

18. Заварыкин, В. М. **Численные методы:** Учеб. пособие для студентов физ.-мат. спец. пед. ин-тов. / *В. М. Заварыкин, В. Г. Житомирский, М. П. Лапчик.* — М.: Просвещение, 1990. — 176 с.

19. Замков, О. О. **Математические методы в экономике.** / *О. О. Замков, А. В. Толстопятенко, Ю. Н. Черемных.* — М.: ДИС, 1997. — 365 с.

20. Зандер, Е. В. **Практикум по исследованию операций: нелинейные, динамические и специальные модели.** Ч. 2. / *Е. В. Зандер, В. П. Злодеев.* — Красноярск: РИЦ Краснояр. гос. ун-та, 1998. — 54 с.

21. **Исследование операций в экономике**. / *Под ред. Н. Ш. Кремера.* — М.: Банки и биржи, ЮНИТИ, 1997. — 407 с.

22. **Исследование операций.** — В 2 т. Пер. с англ. / *Под ред. Дж. Моудера, С. Элмаграби.* — М.: Мир, 1981. — Т. 1. — 712 с.

23. **Исследование операций.** — В 2 т. Пер. с англ. / *Под ред. Дж. Моудера, С. Элмаграби.* — М.: Мир, 1981. — Т. 2. — 677 с.

24. Карасев, А. И. **Математические методы и модели в планировании.** / *А. И. Карасев, Н. Ш. Кремер, Т. И. Савельева.* — М.: Экономика, 1987. — 239 с.

25. Карлин, С. **Математические методы в теории игр, программировании, экономике.** / С. Карлин. — М.: Мир, 1964. — 838 с.

26. Косоруков, О. А. **Исследование операций:** Учебник. / *О. А. Косоруков, А. В. Мищенко.* — М.: Экзамен, 2003. — 446 с.

27. Кофман А. **Займемся исследованием операций**. / *А. Кофман, Р. Фор.* — М., Мир, 1966. — 279 с.

28. Кузнецов, А. В. **Высшая математика. Математическое программирование.** / *Под общ. ред. проф. А. В. Кузнецов, В. А. Сакович, Н. И. Холод.* — Минск: Высш. школа, 1994. — 288 с.

29. Кузнецов, Ю. Н. **Математическое программирование:** Учеб. пособие. — 2-е изд., перераб и доп. / *Ю. Н.Кузнецов, В. И. Кузубов, А. Б. Волощенко.* — М.: Высш. школа, 1980. — 300 с.

30. Льюис, Р. Д. **Игры и решения**. / *Р. Д. Льюис, Х. Райфа.* — М.: Изд-во иностр. лит, 1961. — 642 с.

31. Ляшенко, И. Н. **Линейное и нелинейное программирование.** / *И. Н. Ляшенко, Е. А. Карагодова, Н. В. Черникова, Н. З. Шор.* — Издательское объединение «Вища школа», 1975. — 372 с.

32. Мак Кинси, Дж. **Введение в теорию игр**. / *Дж. Мак Кинси.* — М.: Физматгиз, 1960. — 420 с.

33. Мастяева, И. Н. **Прикладная математика и математическое моделирование в бизнесе.** / *И. Н. Мастяева, Г. Я. Горбоцов, В. Б. Турундаевский.* — М.: МЭСИ, 1997. — 131 с.

34. **Математическая экономика на персональном компьютере**: Пер. с яп. / *М. Кубонива, М. Табата, С. Табата, Ю. Хасэбэ; под ред. М. Кубонива.* — М.: Высш. школа, 1980. — 303 с.

35. Морозов, В. В. **Исследование операций в задачах и упражнениях**. / *В. В. Морозов, А. Г. Сухарев, В. В. Федоров.* — М., 1986. — 285 с.

36. Мошкович, Л. И. **Ситуационный анализ в экономике**. / *Л. И. Мошкович, Е. В. Зандер, В. П. Злодеев.* — Красноярск: Краснояр. гос. ун-т, 1996. — 34 с.

37. Мулен, Э. **Теория игр**. / *Э. Мулен.* — М., 1985. — 199 с.

38. Нейман, Дж. **Теория игр и экономическое поведение**. / *Джон фон Нейман.* — М.: Наука, 1997. — 708 с.

39. Оуэн, Г. **Теория игр**. / *Г. Оуэн.* — М.: Мир, 1971. — 230 с.

40. Солодовников, А. С. **Введение в линейную алгебру и линейное программирование.** / *А. С. Солодовников.* — М.: Изд. Просвещение, 1966. — 184 с.

41. Схрейвер, А. **Теория линейного и целочисленного программирования:** В 2 т. Пер с англ. / *А. Схрейвер.* — М.: Мир, 1991. — Т. 1. — 360 с.

42. Таха, Х. **Введение в исследование операций**. / *Х. Таха*. — М.: Мир, 1985. — 479 с.

43. Тынкевич, М. А. **Экономико-математические методы (исследование операций).** Изд. 2-е, испр. и доп. / *М. А. Тынкевич.* — Кемерово, 2000. — 177 с.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Учебное издание

Исследование операций в экономике

Евгения Викторовна ЗАНДЕР

Валерий Павлович ЗЛОДЕЕВ

Леонид Иосифович МОШКОВИЧ

Анна Робертовна СЕМЁНОВА

Редактор — *О. Ф. Александрова*
Корректор — *Т. Е. Бастрыгина*
Лицензия

Компьютерная верстка — *А. И. Пыжев*

Печать офсетная. Подписано в печать 00.00.0000. Формат 64 × 80 / 16.
Бумага типографская. Гарнитура Таймс.
Усл.-печ. л. 0,0. Усл.-изд. л. 0,0. Тираж 0000 экз.
Заказ № 0000.
Цена договорная.

Издательский центр Института естественных и гуманитарных наук
Сибирского федерального университета.
660041 Красноярск, пр. Свободный, 79.

ТЕМА 1. РАЗОВЫЕ ПЛАТЕЖИ: ПРОСТЫЕ И СЛОЖНЫЕ ПРОЦЕНТЫ

Лекция 1. Предмет финансовой математики

Вопросы, рассматриваемые в лекции

1. *Финансовая математика – основа количественного анализа финансовых операций.*
2. *Основные задачи финансовой математики.*
3. *Область приложения методов количественного анализа финансовых операций.*
4. *Понятие процентов, процентной ставки. Виды процентных ставок.*
5. *Период начисления и способы начисления процентов.*
6. *Простые и сложные процентные ставки.*
7. *Ставки наращения, дисконтные (учётные) ставки. Декурсивные и антисипативные проценты. Фиксированные и плавающие процентные ставки.*

Финансовая математика – основа количественного анализа финансовых операций

В последние годы в нашей стране значительные изменения произошли в сфере приложений математики. Если раньше развитие прикладной математики в решающей степени стимулировало задачи естественных наук и связанных с ними отраслей промышленности (что в значительной степени явно или неявно определялось военно-промышленным комплексом), то сегодня трудности в этих областях заставили математиков активно искать новые сферы приложения своих знаний.

Социально-экономические причины перенесли интересы, во всяком случае специалистов по прикладной математике, на новые области, которые практически не были известны в нашей стране до начала 90-х годов. Активное развитие банковской, страховой, инвестиционной деятельности поставило необходимость привлечения в эти области специалистов совершенно нового для нашей страны типа. Одной из новых для нашей страны областей оказалась финансовая математика.

Первая компания по страхованию жизни, действующая на научных принципах, была организована в Лондоне в 1762 году (Справедливое Общество Страхования Жизни). Секретарю этой компании, который регистрировал собрания руководства, а также выписывал полисы страхователям, было дано название *актуарий*. В 1775 году на этот пост был назначен математик. Он был ответствен за вычисление приемлемых ставок страховых взносов и обеспечивал надежность финансовых операций компании. С тех пор название *актуарий* стало все шире употребляться для тех. кто выполнял эту финансовую и математическую работу.

Область математики, которая занимается математическими проблемами финансов, называется *актуарная математика.*

В настоящее время финансовая математика испытывает период интенсивного развития, особенно в той ее части, которая связана с современным стохастическим анализом. Именно методы общей теории случайных процессов оказались наиболее подходящими для адекватного описания эволюции ценных бумаг.

Весь комплекс количественных методов анализа финансовых операций условно делится на две основных части. В первой части, называемой «Финансовая математика» рассматриваются: анализ операций по начислению простых и сложных процентов, а также вопросы их практического применения; количественный анализ наиболее часто используемых потоков платежей. Иногда эту часть еще называют финансовой арифметикой. Во второй части комплекса количественных методов анализа финансовых операций, называемой «Финансовая математика – II», рассматриваются характеристики методов количественного анализа, применяемых при решении конкретных практических задач из различных областей финансово-кредитной деятельности (измерение доходности финансовых инструментов, измерение эффективности производственных инвестиций, расчет лизинговых операций, оценка страховых аннуитетов и т.д.).

Количественный финансовый анализ – одно из самых динамичных направлений экономической науки – сформировался на стыке науки о финансах и математики. Он нацелен на решение широкого круга задач, основными из которых являются:

- ✓ измерение конечных результатов финансовой операции;
- ✓ выявление зависимости конечных результатов от основных параметров операции, измерение взаимосвязи этих параметров, определение их допустимых граничных значений;
- ✓ разработка планов выполнения финансовых операций;
- ✓ нахождение параметров эквивалентного изменения условий операции.

Любая финансово-кредитная операция, инвестиционный проект или коммерческое соглашение предполагают наличие ряда условий их выполнения, с которыми согласны участвующие стороны. К таким условиям относятся следующие количественные данные: денежные суммы, временные параметры, процентные ставки и некоторые другие дополнительные величины. Каждая из перечисленных характеристик может быть представлена самым различным образом.

Говоря о классификации платежей необходимо отметить, что существуют постоянные и переменные во времени платежи, единовременные (разовые) или в рассрочку (см. рис.1). Примером постоянных платежей может являться выплата в течение семестра стипендии студентам или пенсия (здесь не берется в расчет индексация). С переменными платежами мы сталкиваемся в розничной торговле, когда каждая покупка определяется своим размером платежа. Последовательность таких платежей переменна во времени. Кроме того, каждый такой платеж в отдельности будет еще и разовым. Примером платежей в рассрочку может быть платежи, которые осуществляются в потребительском кредите.

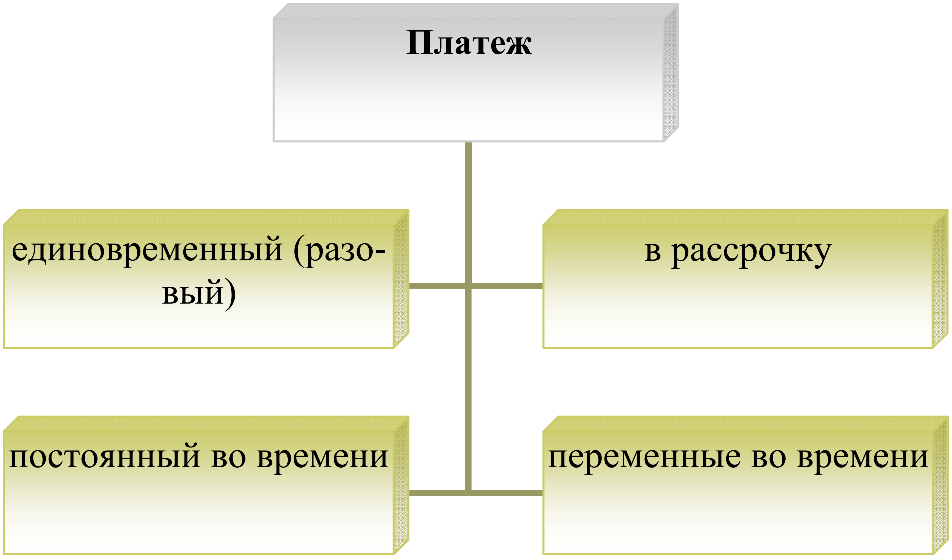


Рис. 1. Классификация платежей

Процентами называют абсолютную величину дохода от предоставления денег в долг в любой его форме: выдача ссуды, продажа товара в кредит, помещение денег на депозитный счет, учет векселя, покупка сберегательного сертификата или облигации и т.д.

При заключении финансового или кредитного соглашения стороны (кредитор и заемщик) договариваются о размере процентной ставки. Процентная ставка измеряет уровень доходности отнесением абсолютного эффекта (полученного дохода в виде суммы процентных денег, начисленных за весь срок) к исходной сумме долгового обязательства. Временной интервал, к которому приурочена процентная ставка, называют периодом начисления. В качестве периода начисления принимают год (в большинстве случаев), полугодие, квартал, месяц или даже день.

Существует более десятка видов процентных ставок и методов начисления процентов (см. рис. 2). При выборе принципа расчета процентных денег существует два подхода: от настоящего к будущему и, наоборот, от будущего к настоящему. Соответственно применяют ставки наращения и дисконтные, или учетные, ставки.

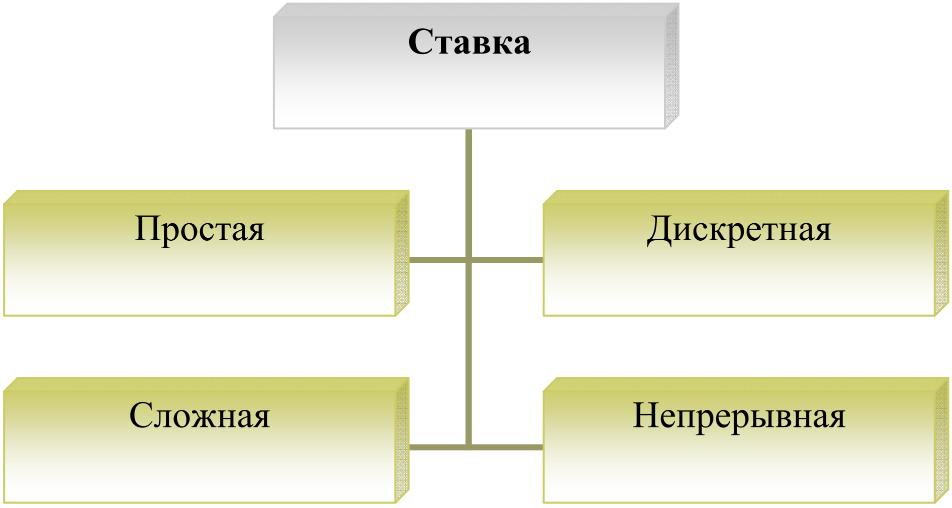


Рис. 2. Классификация процентных ставок.

В финансовой литературе проценты, полученные по ставке наращения, принято называть декурсивными, по учетной ставке — антисипативными.

Процентные ставки могут быть фиксированными (в контракте указываются их размеры) или плавающими. В последнем случае указывается не сама ставка, а изменяющаяся во времени база (базовая ставка) и размер надбавки к ней — маржи. Классическим примером базовой ставки может служить лондонская межбанковская ставка ЛИБОР. В России применяются базовые ставки по рублевым кредитам МИБОР. Размер маржи определяется рядом условий, в частности финансовым положением заемщика, сроком кредита и т.д. Он может быть постоянным на протяжении срока ссудной операции или переменным.

Определившись с принципом расчета процентных денег, в зависимости от длительности финансовой операции, используют простые и сложные процентные ставки. Простые процентные ставки используются в краткосрочных денежно-кредитных операциях (до года), в случае если срок операции свыше года – используют сложную процентную ставку.

Простые проценты начисляются по ставке i на одну и ту же постоянную базу – исходную сумму долга P, что за счет многократного прибавления постоянной величины процентного дохода за один период Pi приводит к росту $S=P(1+iN)$ за полный срок N по закону арифметической прогрессии.

Сложные проценты характеризуются тем, что база начисления растет в результате регулярного присоединения к ней процентных денег, причитающихся кредитору за предыдущие расчетные периоды. Получается геометрическая прогрессия с постоянным знаменателем, равным множителю наращения $(1 + i)$ за один период по ставке процентов i.

В практических расчетах применяют дискретные проценты, которые начисляются за фиксированные интервалы времени (год, полугодие и т.д.). Иначе говоря, время рассматривается как дискретная переменная. В некоторых случаях – в доказательствах и аналитических финансовых расчетах, связанных с процессами, которые можно рассматривать как непрерывные, в общих теоретических разработках и значительно реже на практике (в случае электронных торгов) – возникает необходимость в применении непрерывных процентов, когда наращение или дисконтирование производится непрерывно, за бесконечно малые промежутки времени. В подобных ситуациях применяют специальные непрерывные процентные ставки.

Время устанавливается в виде фиксированных сроков платежей, интервалов поступлений доходов, моментов погашения задолженности и т.д. (см. рис. 3).

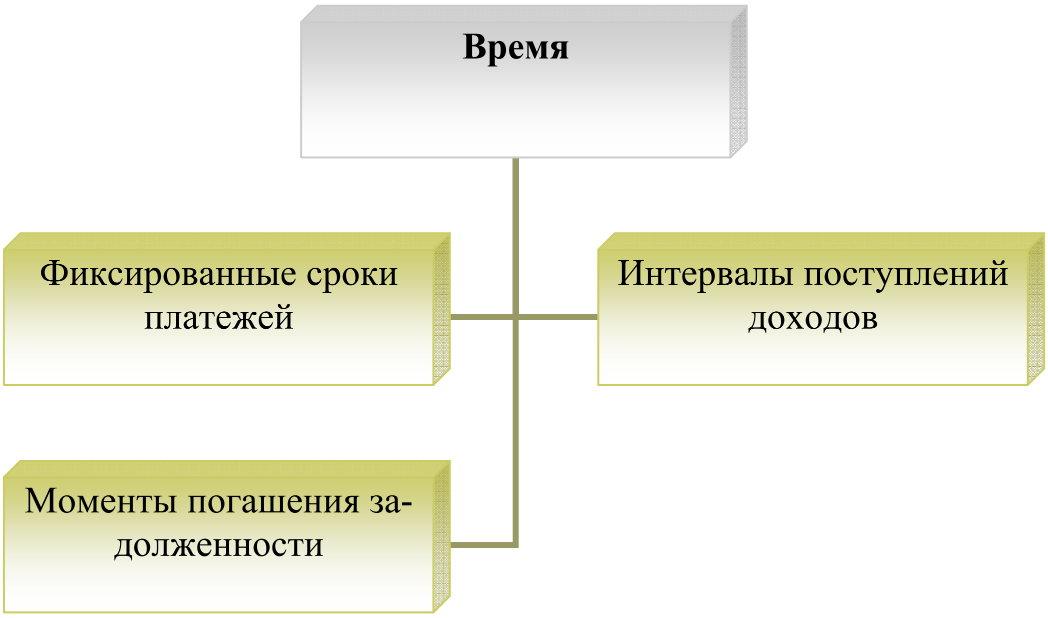


Рис. 3. Фактор времени в денежно-кредитных операциях.

В рамках одной финансовой операции перечисленные показатели образуют некоторую взаимоувязанную систему, подчиненную соответствующей логике. Изменение значения одной из этих величин в большей или меньшей мере скажется на результате соответствующей операции. Отсюда следует, что такие системы могут и должны являться объектом приложения количественного финансового анализа. Методы данного анализа и составляют предмет финансовой математики.

Количественный финансовый анализ предназначен для решения разнообразных задач (см. рис. 4). Эти задачи можно разделить на две большие группы: традиционные или "классические", и новые, нетрадиционные, постановка и интенсивная разработка которых наблюдается в последние два—три десятилетия.

Рис. 4. Классификация задач финансовой математики

Количественный финансовый анализ применяется как в условиях определенности, так и неопределенности. В первом случае предполагается, что данные для анализа заранее известны и фиксированы. Например, при выпуске обычных облигаций однозначно оговариваются все параметры – срок, купонная доходность, порядок выкупа. Анализ заметно усложняется, когда приходится учитывать неопределенность – динамику денежного рынка (уровень процентной ставки, колебания валютного курса и т.д.), поведение контрагента.

Рамки финансовой математики достаточно широки – от элементарных начислений процентов до относительно сложных расчетов, например оценки влияния различных факторов на эффективность выпуска облигаций или методов сокращения риска путем диверсификации портфеля финансовых инвестиций и т.д.

К основным задачам финансовой математики относятся (см. рис. 5):

- ✓ измерение конечных финансовых результатов операции (сделки, контракта) для каждой из участвующих сторон;
- ✓ разработка планов выполнения финансовых операций, например, планов погашения задолженности по кредиту;
- ✓ измерение зависимости конечных результатов операции от основных ее параметров;

✓ определение допустимых критических значений параметров операции и расчет параметров эквивалентного (безубыточного) изменения первоначальных условий операции.

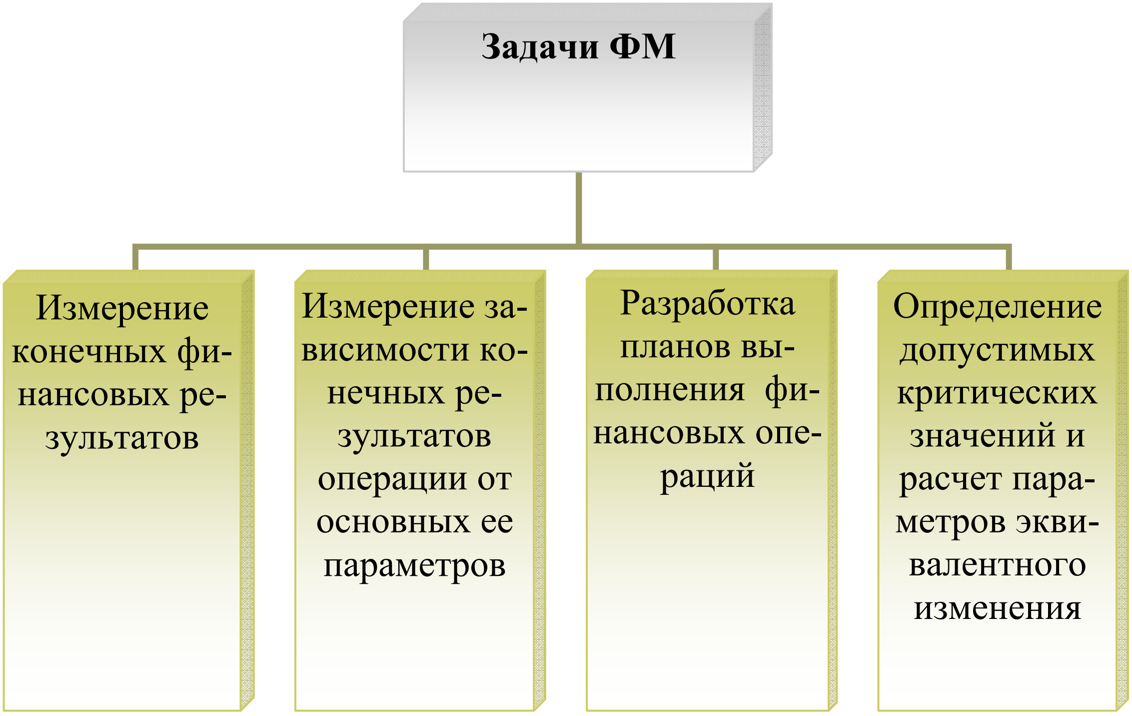


Рис. 5. Классификация задач финансовой математики

Знание методов, применяемых в финансовой математике, необходимо при непосредственной работе в любой сфере финансов и кредита, в том числе и на этапе разработки условий контрактов. Нельзя обойтись без них при финансовом проектировании, а также при сравнении и выборе долгосрочных инвестиционных проектов. Финансовые вычисления являются необходимой составляющей расчетов в долгосрочном личном страховании, например проектировании и анализе состояния пенсионных фондов (расчет тарифов, оценка способности фондов выполнить свои обязательства перед пенсионерами и т.д.), долгосрочном медицинском страховании.

Время как фактор в финансовых расчетах

В практических финансовых операциях суммы денег обязательно связывают с конкретными моментами или периодами времени. Для этого в контрактах фиксируют соответствующие даты и периодичность выплат. Фактор времени, особенно в долгосрочных операциях, играет не меньшую, а иногда даже и большую роль, чем размеры денежных сумм. Необходимость учета временного фактора вытекает из сущности финансирования, кредитования и инвестирования и выражается в принципе неравноценности денег, относящихся к разным моментам времени.

Влияние фактора времени усиливается в период инфляции. Этот фактор часто лежит в основе явного или скрытого мошенничества и недобросовестности. Очевидным следствием принципа изменения ценности денег во времени является неправомерность суммирования денежных величин, относящихся к разным моментам времени, особенно при принятии решений финансового порядка. Однако такое суммирование вполне допустимо там, где фактор времени не имеет принципиального значения. Например, в бухгалтерском учете для получения итогов по периодам и в финансовом контроле, но, повторяем, не при принятии финансовых решений долгосрочного характера. Неправомерно также и непосредственное сравнение разновременных денежных величин. Их сравнение допустимо только при "приведении" таких сумм к одному моменту времени. Способы приведения обсуждаются ниже для разных вариантов производства платежей.

Другим важным принципом является финансовая эквивалентность – равенство (эквивалентность) финансовых обязательств сторон, участвующих в операции. Например, покупатель облигации оплачивает ее рыночную цену, а эмитент обязуется периодически выплачивать ему купонный доход и вернуть в конце срока сумму, равную номиналу облигации, или страхователь оплачивает стоимость страхования, а страховщик обязуется выплатить ему страховую сумму, но только при наступлении страхового события. В первом случае платежи обеих сторон безусловны, во втором – платеж страховщика имеет вероятностный характер.

Лекция 2. Наращивание и дисконтирование по простым процентным ставкам

Вопросы, рассматриваемые в лекции

1. *Сущность процентных платежей. Вычисление наращенных сумм на основе простых процентных ставок.*
2. *Три метода процентных расчётов: «Германская практика». «Французская практика». «Английская практика».*
3. *Переменные ставки. Начисление процентов при изменении сумм депозита во времени. Реинвестирование по простым процентным ставкам.*
4. *Погашение задолженности частями. Актуарный метод и правило торговца.*
5. *Наращение процентов в потребительском кредите.*
6. *Дисконтирование и его сущность.*

7. *Математическое дисконтирование. Определение уровня процентной ставки и продолжительности ссуды.*
8. *Банковское дисконтирование. Определение сроков ссуды, величин простых процентных и учётных ставок.*
9. *Определение срока ссуды и величины процентной ставки.*

Формула наращения по простым процентам

Под наращенной суммой ссуды S (долга, депозита, других видов выданных в долг или инвестированных денег) понимают первоначальную ее сумму P с начисленными процентами к концу срока начисления. Наращенная сумма определяется умножением первоначальной суммы долга P на множитель наращения, который показывает, во сколько раз наращенная сумма больше первоначальной.

К наращению по простым процентам обычно прибегают при выдаче краткосрочных ссуд (на срок до 1 года) или в случаях, когда проценты не присоединяются к сумме долга, а периодически выплачиваются. Для записи формулы наращения простых процентов используют стандартные обозначения:

I – проценты за весь срок ссуды;

P – первоначальная сумма долга;

S – наращенная сумма, т. е. сумма в конце срока;

i – ставка наращения процентов;

n – срок ссуды.

Пусть срок ссуды измеряется в годах, тогда i – годовая процентная ставка. Каждый год она приносит проценты в сумме $P \cdot i$. Тогда начисленные за весь срок проценты будут определены формулой (1):

$$I = \underbrace{Pi + Pi + \ldots + Pi}_{\text{за n лет – срок финансовой операции}} = Pin \tag{1}$$

Наращенная сумма за весь период финансовой операции определяется формулой (2):

$$S = P + I = P + Pin = P(1 + in) \tag{2}$$

Выражение (2) называют формулой наращения по простым процентам, а множитель (1 + *in*) – множителем наращения простых процентов. Графически рост по простым процентам представлен на рис. 6.

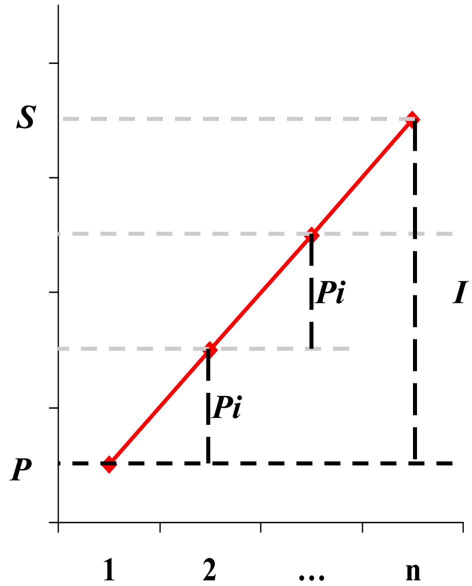


Рис. 6. График роста по простым процентам

Пример 1. Определить проценты и сумму накопленного долга, если ссуда равна 700 тыс.руб., срок 4 года, проценты простые по ставке 15% годовых.
Решение. Проценты за весь срок финансовой операции составят

$$\boldsymbol{I} = 700 \text{ x } 4 \text{ x } 0{,}15 = 420 \text{ (тыс. руб.).}$$

Наращенная сумма в этом случае будет равна

$$\boldsymbol{S} = 700 + 420 = 1120 \text{ (тыс. руб.).}$$

Если увеличить ставку в два раза, сумма процентов при этом удвоится. Однако наращенная сумма увеличится в

$$\frac{1 + 4 \cdot 2 \cdot 0{,}15}{1 + 4 \cdot 0{,}15} = 1{,}375 \text{ раза.}$$

◆

Процентная ставка, как правило, устанавливается в расчете за год, при сроке ссуды менее года необходимо определить, какая часть годового процента уплачивается кредитору. Аналогичная поступают в случаях, когда срок ссуды меньше периода начисления.

Рассмотрим типичный случай: процентная ставка с годовым периодом начисления. При этом срок ссуды необязательно равен целому числу лет. Выразим срок n в виде дроби (3)

$$n = \frac{t}{K} \tag{3}$$

где t – число дней ссуды,

K – число дней в году, или временная база начисления процентов.

При расчете процентов применяют две временные базы:

K = 360 дней (12 месяцев по 30 дней) или K = 365 (366) дней. Если K = 360, то получают обыкновенные или коммерческие проценты, а при использовании действительной продолжительности года рассчитывают точные проценты. Число дней ссуды также можно измерить приближенно и точно. В первом случае продолжительность ссуды определяется из условия, согласно которому любой месяц принимается равным 30 дням. В свою очередь точное число дней ссуды определяется путем подсчета числа дней между датой выдачи ссуды и датой ее погашения. День выдачи и день погашения считаются за один день. В банковской практике различных стран срок в днях и расчетное количество дней в году при начислении процентов определяется поразному. Различают три системы начисления: германскую, французскую и английскую:

- ✓ **Германская система**: K = 360 дней (12 месяцев по 30 дней в каждом месяце). Метод применяется тогда, когда не требуется большой точности, например при промежуточных расчетах. Он принят в практике коммерческих банков Германии, Швеции, Дании. Метод условно обозначается как 360/360.
- ✓ **Французская система**: K = 360 дней (12 месяцев с номинальным количеством дней в каждом месяце). Этот метод, иногда называемый банковским, распространен в межстрановых ссудных операциях коммерческих банков, во внутристрановых — во Франции, Бельгии, Швейцарии. Он обозначается, как 365/360 или АСТ/360. Этот вариант дает несколько больший результат, чем применение точных процентов. Заметим, что при числе дней ссуды, превышающем 360, данный способ приводит к тому, что сумма начисленных процентов будет больше, чем предусматривается годовой ставкой.

✓ **Английская система**: K = 365 (366) дней (12 месяцев с номинальным количеством дней в каждом месяце). Этот вариант, естественно, дает самые точные результаты. Данный способ применяется центральными банками многих стран и крупными коммерческими банками, например, в Великобритании, США. В коммерческих документах он обозначается как 365/365 или ACT/ACT.

Пример 2. Женщина 6 августа взяла в долг под проценты деньги у своей соседки. Размер долга составил 15000 руб., а проценты – 25% годовых. Рассчитайте сумму, которую должна отдать женщина, методом точных процентов с точным числом дней, если долг должен быть возвращен 15 сентября и год не високосный.

Решение. Сумма долга P = 15000 руб., ставка i = 0,25. Метод точных процентов с точным числом дней предусматривает отношение реального числа дней операции к реальному числу дней в году. Так как год не високосный, то временная база K = 356. Число дней операции будет состоять из 26 дней августа (с 6 по 31 августа) и 15 дней сентября (с 1 по 15 сентября), а так как день начала и окончания финансовой операции считается как один день, то получается, что t = 26 + 15 – 1 = 40 дней. Используя формулу наращения по простым процентам (2) и формулу (3) для расчета периода финансовой операции, получим

$$S = P\left(1+\frac{t}{K}i\right) = 15000\left(1+\frac{40}{365}0,25\right) \approx 15410,96\,(\text{руб.})$$

◆

В кредитных соглашениях иногда предусматриваются изменяющиеся во времени процентные ставки (см. рис. 7).

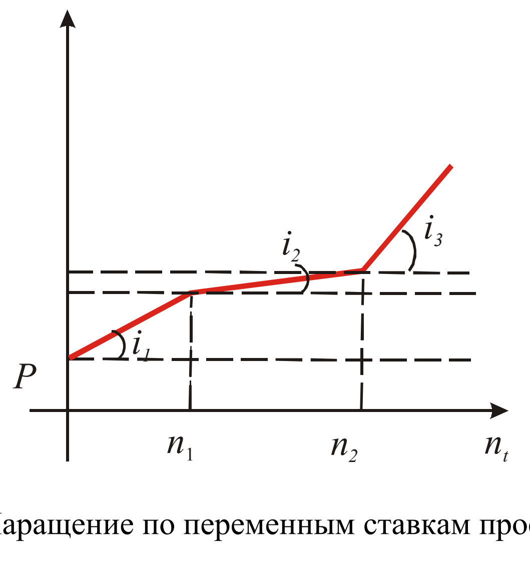


Рис. 7. Наращение по переменным ставкам простых процентов

Если это простые ставки, то наращенная на конец срока сумма определяется следующим образом:

$$S = P\left(1 + n_1 i_1 + n_2 i_2 + \cdots + n_m i_m\right) = P\left(1 + \sum_t n_t i_t\right) \tag{4}$$

где i_t – ставка простых процентов в периоде t,

n_t – продолжительность периода с постоянной ставкой,

$n = \sum_t n_t$.

Пример 3. Контракт предусматривает следующий порядок начисления процентов: первый год — 8%, в каждом последующем полугодии ставка повышается на 0,5%. Необходимо определить множитель наращения за 2,5 года.
Решение.

$$1 + \sum_t n_t i_t =$$
$$= 1 + 1 \times 0.08 + 0.5 \times 0.085 + 0.5 \times 0.09 + 0.5 \times 0.095 = 1.215$$

◆

Теперь рассмотрим случай, когда сумма, на которую начисляются проценты, изменяет свою величину во времени (размер вклада на сберегательном счете, текущий счет при периодическом его пополнении или снятии денег и т.п.).

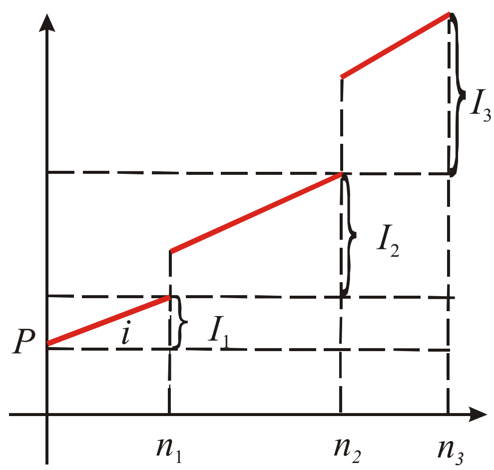


Рис. 8. Начисление процентов при изменении сумм депозита во времени

Пусть R_j - остаток средств на счете в момент j после очередного поступления или списания средств, n_j - срок хранения денег (в годах) до нового изменения остатка средств на счете. Тогда суммарные проценты за весь срок финансовой операции будут равны

$$I = \sum_j I_j = \sum_j R_j n_j i$$

$$\text{если } n_j = \frac{t_j}{K}, \text{ то} \tag{5}$$

$$I = \sum_j I_j = \sum_j R_j n_j i = \sum_j R_j \frac{t_j}{K} i = \frac{\sum R_j t_j}{100} : \frac{K}{i}$$

В формуле (5) K – временная база, т.е. число дней в году, t_j – срок в днях между последовательными изменениями остатков на счете.

Пример 4. Вкладчик положил на свой счет 50000руб. под 6% годовых. Через 2 месяца он пополнил вклад, добавив 10000 руб., а еще через 5 месяцев добавил 25000 руб. Какова будет сумма вклада через год после открытия счета?

Решение.

$$\underbrace{\left(\underbrace{\left(\underbrace{50(1+2\cdot 0{,}005)}_{\text{через 2 месяца}}+10\right)(1+5\cdot 0{,}005)+25}_{\text{через 7 месяцев}}\right)(1+5\cdot 0{,}005)}_{\text{через 1 год}} = 89187{,}81 \text{ руб.}$$

◆

На практике при инвестировании средств в краткосрочные депозиты иногда прибегают к неоднократному последовательному повторению наращения по простым процентам в пределах заданного общего срока. Это означает реинвестирование средств, полученных на каждом этапе наращения, с помощью постоянной или переменной ставок.

Пусть i_t – размер ставок, по которым производится реинвестирование, которые являются простыми постоянными на промежутке времени n_t процентными ставками. Тогда наращенная сумма для всего срока будет определена формулой (6)

$$S = P(1+n_1 i_1)(1+n_2 i_2)\cdots(1+n_t i_t)\cdots \qquad (6)$$

Если промежуточные сроки начисления и ставки не изменяются во времени, то формула (6) упрощается до (7):

$$S = P(1+ni)^m, \qquad (7)$$

где m – количество повторений реинвестирования.

Погашение задолженности частями

Необходимым условием финансовой или кредитной операции в любой ее форме является сбалансированность вложений и отдачи. Для лучшего понимания процесса, рассмотрим пример: выдана ссуда на срок T в размере P (см. рис. 9).

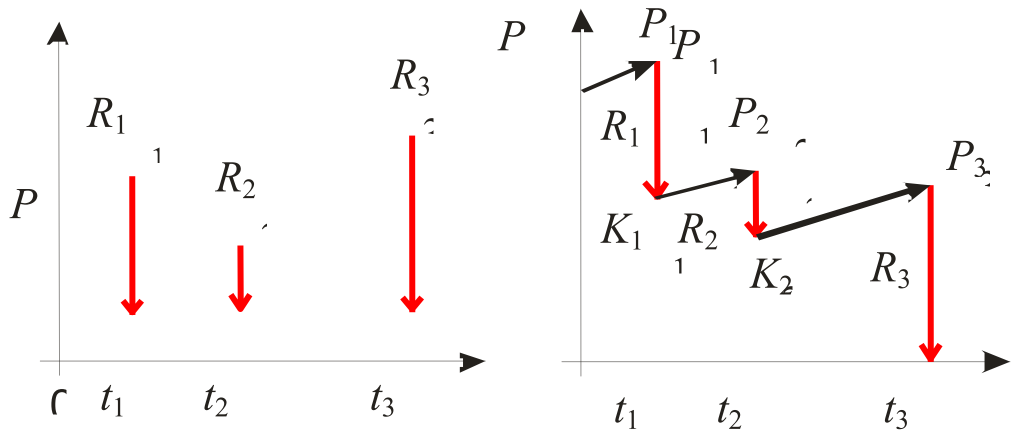


Рис. 9. Погашение задолженности частями

На протяжении этого срока в счет погашения задолженности производятся два платежа R_1 и R_2, а в конце срока выплачивается остаток задолженности в сумме R_3 (здесь не имеет значения, какая часть этой суммы идет на выплату процентов, а какая – на погашение долга).

Рассмотрим процесс погашения задолженности во времени. На первом интервале t_1 начисляются проценты, задолженность возрастает до величины P_1. В конце этого периода выплачивается в счет погашения задолженности сумма R_1, при этом долг уменьшается до величины K_1. Аналогичные процессы повторяются на последующих интервалах. Заканчивается операция тем, что кредитор получает в окончательный расчет сумму R_3 и задолженность становится равной нулю. График операции в этом случае называют контуром (рис. 9, справа). При этом сбалансированная операция обязательно имеет замкнутый контур, т.е. последняя выплата полностью покрывает остаток задолженности. В этом случае совокупность платежей точно соответствует условиям сделки.

Продолжая анализировать предыдущий пример, неизбежно встает вопрос о том, какую сумму надо брать за базу для расчета процентов и каким путем определять остаток задолженности по промежуточным платежам. Существуют два метода решения этой задачи. Первый, который применяется в основном в операциях со сроком более года, называют актуарным методом, второй – правилом торговца, он используется коммерческими фирмами в сделках со сроком не более года. Если иное не оговорено, то при начислении процентов в обоих методах используются обыкновенные проценты с приближенным числом дней (360/360).

Актуарный метод предполагает последовательное начисление процентов на фактические суммы долга. Частичный платеж идет в первую очередь на погашение процентов, начисленных на дату платежа. Если величина платежа превышает сумму начисленных процентов, то разница (остаток) идет на погашение основной суммы долга. Непогашенный остаток долга служит базой для начисления процентов за следующий период. Если частичный платеж меньше начисленных процентов, то никакие зачеты в сумме долга не делаются и поступление приплюсовывается к следующему платежу.

Актуарный метод для предыдущего примера будет иметь вид:

$$\begin{gathered}K_1 = P(1+t_1 i) - R_1\\ K_2 = K_1(1+t_2 i) - R_2\\ K_2(1+t_3 i) - R_3 = 0\end{gathered} \tag{8}$$

При использовании правила торговца возможны два варианта:

1. Срок ссуды не превышает год. В этом случае сумма долга с процентами остается неизменной до полного погашения. Накапливаются частичные платежи с начисленными на них до конца срока процентами. Последний взнос должен быть равен разности этих сумм.

2. Срок ссуды превышает год. Расчеты делаются для годового периода задолженности. В конце года из суммы задолженности вычитается наращенная сумма накопленных частичных платежей. Остаток погашается в следующем году.

Введем обозначения:

Q – остаток долга на конец срока или года,

S – наращенная сумма долга,

K – наращенная сумма платежей,

R_j – сумма частичного платежа,

n – общий срок ссуды,

t_j – интервал времени от момента платежа до конца срока ссуды или года.

Тогда правило торговца будет записано в форме (9):

$$Q = S - K = P(1+ni) - \sum R_j \left(1 + t_j i_j\right) \qquad (9)$$

Графическое изображение такой операции при выплате двух промежуточных платежей охватывает два параллельных контура (см. рис. 10). Первый характеризует наращение задолженности, второй – наращение на суммы поступлений.

Необходимо отметить, что для одних и тех же данных актуарный метод и правило торговца в общем случае дают разные результаты. Остаток задолженности по первому методу немного выше, чем по второму.

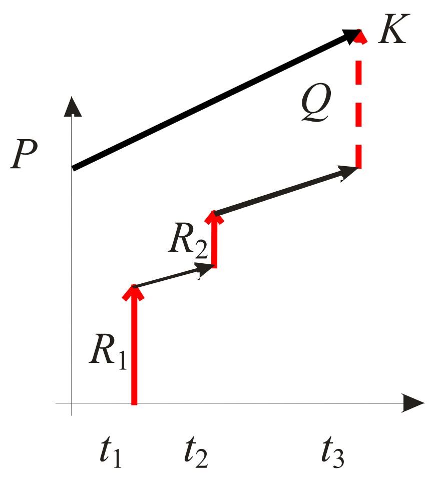


Рис. 10. Графическая иллюстрация правила торговца

Наращение процентов в потребительском кредите

В потребительском кредите проценты, как правило, начисляются на всю сумму кредита и присоединяются к основному долгу уже в момент открытия кредита.

Погашение долга с процентами производится частями, обычно равными суммами на протяжении всего срока кредита. В таком случае, наращенная сумма будет вычисляться по формуле (2), а величина разового погасительного платежа составит

$$R = \frac{S}{n \cdot m}, \tag{11}$$

где n – срок кредита в годах, m – число платежей в году.

В связи с тем, что проценты здесь начисляются на первоначальную сумму долга, а его фактическая величина систематически уменьшается во времени, действительная стоимость кредита заметно превышает договорную процентную ставку.

Дисконтирование по простым процентным ставкам

В финансовой практике часто сталкиваются с задачей, обратной наращению процентов: по заданной сумме S, которую следует уплатить через некоторое время n, необходимо определить сумму полученной ссуды P. Такая ситуация может возникнуть, например, при разработке условий контракта. Расчет P по S необходим и тогда, когда проценты с суммы S удерживаются вперед, т.е. непосредственно при выдаче кредита, ссуды. В этих случаях говорят, что сумма S дисконтируется или учитывается, сам процесс начисления процентов и их удержание называют учетом, а удержанные проценты — дисконтом. Необходимость дисконтирования возникает, например, при покупке краткосрочных обязательств, оплата которых должником произойдет в будущем.

Величину P, найденную с помощью дисконтирования, называют современной стоимостью, или современной величиной, будущего платежа S. Современная величина суммы денег является одним из важнейших понятий в количественном анализе финансовых операций. В большинстве случаев именно с помощью дисконтирования, а не наращения, удобно учитывать такой фактор, как время.

В зависимости от вида процентной ставки применяют два метода дисконтирования – математическое дисконтирование и банковский (коммерческий) учет. В первом случае применяется ставка наращения, во втором – учетная ставка.

Задача математического дисконтирования состоит в определении первоначальной суммы ссуды, которую надо выдать в долг, чтобы получить в конце срока сумму S, при условии, что на долг начисляются проценты по ставке i. Из формулы наращения по простым процентам (2), находим

$$P = \frac{S}{1+ni} = \frac{S}{1+\frac{t}{K}i}, \qquad (12)$$

где n – срок ссуды в годах, $n=t/K$, t – продолжительность финансовой операции в днях, K – временная база.

Пример 5. Инвестор намерен положить деньги в банк под 8% годовых с целью накопления через год 216 тыс. руб. Определить сумму вклада.
Решение.

$$P = 216\frac{1}{1+0,08} = 200 \text{ тыс.руб.}$$

◆

Банковский учет (учет векселей). Суть операции заключается в следующем. Банк или другое финансовое учреждение до наступления срока платежа по векселю или иному платежному обязательству приобретает его у владельца по цене, которая меньше суммы, указанной на векселе, т.е. покупает (учитывает) его с дисконтом. Получив при наступлении срока векселя деньги, банк реализует процентный доход в виде дисконта. В свою очередь владелец векселя с помощью его учета имеет возможность получить деньги хотя и не в полном объеме, однако ранее указанного на нем срока.

При учете векселя применяется банковский, или коммерческий, учет. Согласно этому методу проценты за пользование ссудой в виде дисконта начисляются на сумму, подлежащую уплате в конце срока. При этом применяется учетная ставка d.

Размер дисконта, или суммы учета равен Snd; если d – годовая учетная ставка, то n измеряется в годах.

$$P = S - Snd = S(1 - nd), \tag{13}$$

где n – срок от момента учета до даты погашения векселя. Дисконтный множитель равен $(1 - nd)$. При этом используется французская система подсчета срока начисления процентов.

Из формулы (13) следует, что при $n > 1/d$ величина дисконтного множителя и, следовательно, суммы P станет отрицательной. Иначе говоря, при относительно большом сроке векселя учет может привести к нулевой или даже отрицательной сумме P, что лишено смысла.

Учетная ставка в большей мере используется при временной базе K = 360 дней, число дней ссуды обычно берется точным.

Пример 6. Вексель на сумму 300 тыс.руб. предъявлен в банке за полгода до срока его погашения. Определить сумму, выплаченную владельцу векселя, и сумму дисконта, если банк использует простую учетную ставку 12% годовых.

Решение.

$$Snd = 300 \cdot \frac{1}{2} \cdot 0{,}12 = 18$$

$$P = S(1 - nd) = 300\left(1 - \frac{1}{2} \cdot 0{,}12\right) = 282 \text{ (тыс.руб.)}$$

◆

Простая учетная ставка иногда применяется и при расчете наращенной суммы. В частности, в этом возникает необходимость при определении суммы, которую надо проставить в векселе, если задана текущая сумма долга. Наращенная сумма в этом случае будет определяться по формуле (14)

$$S = P\frac{1}{1 - nd} \tag{14}$$

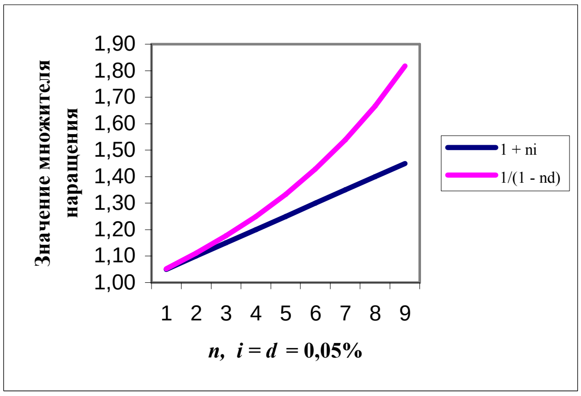
Наращение не пропорционально ни сроку, ни ставке. При $n > 1/d$ расчет лишен смысла, так как наращенная сумма становится бесконечно большим числом.

Ставки наращения и дисконтирования определяют два класса задач: прямые и обратные (см. табл. 1).

Таблица 1.

Определение прямых и обратных задач в финансовой математике

	Прямая задача	Обратная задача
Ставка наращения i	Определение наращения $S = P(1 + ni)$	Определение дисконтирования $P = \dfrac{S}{1+ni}$
Учетная ставка d	Определение дисконтирования $P = S(1 - nd)$	Определение наращения $S = P\dfrac{1}{1-nd}$



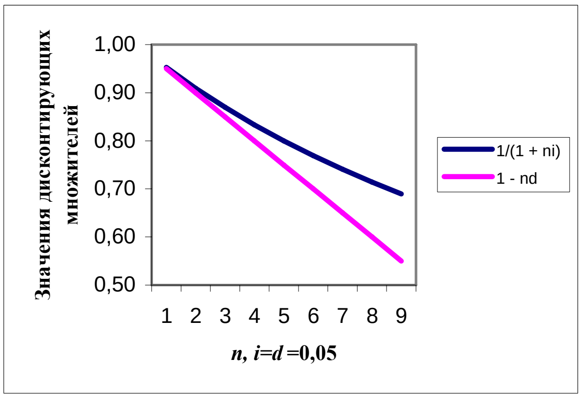


Рис. 11. Графики изменения значения множителя наращения (слева) и дисконтирующего множителя (справа)

Из графиков видно, что учетная ставка отражает фактор времени более жестко, значение множителя наращения и дисконтирующего множителя изменяются сильнее для учетной ставки. С ростом величины ставки эта тенденция усиливается. Учетная ставка дает более быстрый рост суммы задолженности, чем такой же величины ставка наращения. Из сказанного выше следует, что выбор конкретного вида процентной ставки заметно влияет на финансовые итоги операции. Однако возможен такой подбор величин ставок, при котором результаты наращения или дисконтирования будут одинаковыми. Такие ставки называются **эквивалентными**.

Для определения величины эквивалентной ставки необходимо приравнять соответствующие множители наращения или дисконтирования.

Приравняв множители наращения, найдем значение эквивалентной ставки наращения $i_э$ через учетную ставку d, и значение эквивалентной учетной ставки $d_э$ через ставку наращения i (15):

$$1+ni=\frac{1}{1-nd}\Rightarrow\begin{cases} i_э=\dfrac{d}{1-nd} \\ d_э=\dfrac{i}{1+ni} \end{cases} \tag{15}$$

При разработке условий контрактов или их анализе и сравнении возникает необходимость в определении срока ссуды и размера процентной ставки в том или ином ее виде при всех прочих заданных условиях.

Выразим срок ссуды n из формул (2), (13):

$$\begin{aligned} &S = P(1 + ni) \Rightarrow n=\frac{S-P}{Pi} \text{ – срок ссуды в годах} \\ &S = P(1 + \frac{t}{K}i) \Rightarrow t=\frac{S-P}{Pi}K \text{ – срок ссуды в днях} \end{aligned} \tag{16}$$

$$\begin{aligned} &P = S(1 - nd) \Rightarrow n=\frac{S-P}{Sd} \text{ – срок ссуды в годах} \\ &S = P(1 + \frac{t}{K}i) \Rightarrow t=\frac{S-P}{Sd}K \text{ – срок ссуды в днях} \end{aligned} \tag{17}$$

Здесь P – первоначальная стоимость,

S – наращенная стоимость,

i – процентная ставка,

d – учетная ставка,

n –срок уды в годах,

t – срок ссуды в днях,

K – временная база.

Необходимость в расчете процентной ставки возникает при определении финансовой эффективности операции и при сравнении контрактов по их доходности в случаях, когда процентные ставки в явном виде не указаны.

Выразим процентные ставки i и d из формул (2), (13):

$$\begin{aligned} &S = P(1 + ni) \Rightarrow i=\frac{S-P}{Pn} \text{ – срок ссуды в годах} \\ &S = P(1 + \frac{t}{K}i) \Rightarrow i=\frac{S-P}{Pt}K \text{ – срок ссуды в днях} \end{aligned} \tag{18}$$

$$P = S(1 - nd) \Rightarrow d = \frac{S-P}{Sn} \text{ – срок ссуды в годах}$$
$$S = P\left(1 + \frac{t}{K}i\right) \Rightarrow d = \frac{S-P}{St}K \text{ – срок ссуды в днях} \qquad (19)$$

Лекция 3. Наращение и дисконтирование по сложным процентам

Вопросы, рассматриваемые в лекции

1. *Формула наращения по сложным процентам.*
2. *Начисление процентов при дробном числе лет: общий и смешанный методы начисления процентов.*
3. *Сравнение роста по сложным и простым процентам.*
4. *Наращение процентов т раз в году. Номинальная и эффективная ставки процентов.*

5 *Дисконтирование по сложной ставке процентов.*

6 *Учет по сложной процентной ставке. Номинальная и эффективная учетные ставки.*

7 *Определение срока ссуды и размера сложной процентной ставки.*

8 *Непрерывное наращение и дисконтирование. Непрерывные проценты.*

В средне- и долгосрочных финансово-кредитных операциях, если проценты не выплачиваются сразу после их начисления, а присоединяются к сумме долга, применяют сложные проценты. База для начисления сложных процентов увеличивается с каждым шагом во времени. Для формализации такой модели финансовой операции будем использовать прежние обозначения:

P – первоначальная стоимость,
S – наращенная стоимость,
i – годовая процентная ставка,
d – учетная ставка,
n –срок уды в годах,
t – срок ссуды в днях,
K – временная база.

Тогда накопленный капитал по годам составит:

к концу первого года $S_1 = P + Pi = P(1+i)$,

к концу второго года $S_2 = S_1 + S_1 i = S_1(1+i) = P(1+i)^2$,

…

к концу n – го года $S_n = S_{n-1} + S_{n-1}i = S_{n-1}(1+i) = P(1+i)^n$.

Абсолютная сумма начисляемых процентов возрастает (см. рис. 12), и процесс увеличения суммы долга происходит с ускорением. Наращение по сложным процентам можно представить как последовательное реинвестирование средств, вложенных под простые проценты на один период начисления. Присоединение начисленных процентов к сумме, которая послужила базой для их начисления, часто называют *капитализацией процентов.*

Из сказанного выше следует, что наращенная сумма за весь период финансовой операции будет определена формулой (20), проценты за этот же срок в целом формулой (21), а проценты на проценты – (22):

$$S = P(1+i)^n \tag{20}$$

$$I = S - P = P\left[(1+i)^n - 1\right] \tag{21}$$

$$I_p = P(1+i)^n - P(1+in) = P\left[(1+i)^n - (1+in)\right] \tag{22}$$

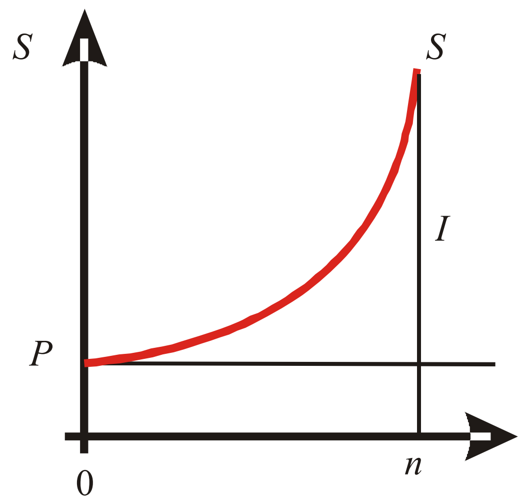


Рис. 12 Графическая иллюстрация наращения по сложным процентам

Величину $(1 + i)^n$ называют множителем наращения по сложным процентам. Если срок финансовой операции измеряется в днях, то для расчета наращенной суммы необходимо помнить, что время обычно измеряется как 365/365 и формула имеет вид:

$$S = P(1+i)^{t/K}, \qquad (23)$$

где t – срок в днях, K = 365 – временная база.

Пример 7. Предприниматель положил 8000 руб. в банк, выплачивающий 6% годовых (сложных). Какая сумма будет на счете этого клиента: 1) через 1 год, 2) через 4 года, 3) через 10 лет.
Решение. По формуле (20) имеем:

для n=1 $\quad S = 8\,000 \cdot (1+0,06)^1 = 8\,480$ (руб.)

для n=4 $\quad S = 8\,000 \cdot (1+0,06)^4 = 10\,099,82$ (руб.)

для n=10 $\quad S = 8\,000 \cdot (1+0,06)^{10} = 14\,326,78$ (руб.)

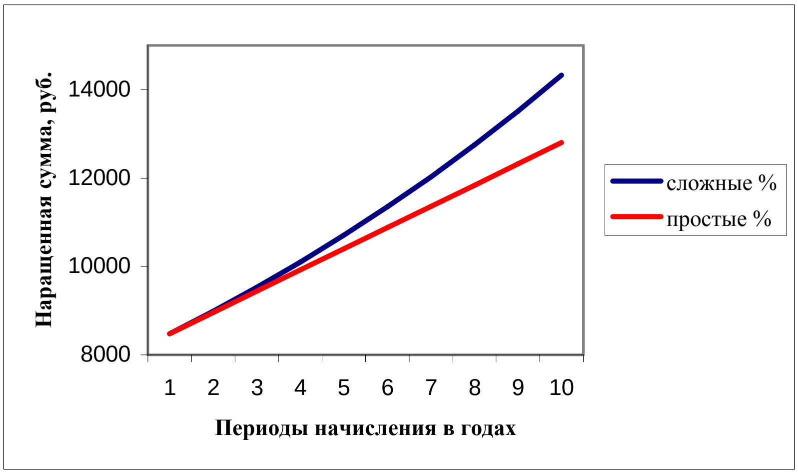


Рис. 13. Сравнение роста наращенной суммы по простым и сложным процентам

На рис. 13 показан сравнительный анализ роста наращенной суммы по простым и сложным процентам.

◆

Формулы (20) – (22) предполагают, что проценты на проценты начисляются по той же ставке, что и при начислении на основную сумму долга. В случае если проценты на основной долг начисляются по ставке i, а проценты на проценты по ставке r. Тогда

$$S = P + \underbrace{Pi}_{\substack{\text{проценты, начисленные} \\ \text{к концу 1-го года}}} + \underbrace{Pi(1+r)}_{\text{к концу 2-го года}} + \underbrace{Pi(1+r)^2}_{\text{к концу 3-го года}} + \cdots + \underbrace{Pi(1+r)^{n-1}}_{\text{к концу } n\text{-го года}} =$$

$$S = P + Pi\left[1 + (1+r) + (1+r)^2 + (1+r)^3 + \cdots + (1+r)^{n-1}\right] =$$

$$\left[\text{т.к. } 1 + k + k^2 + k^3 + \cdots + k^{n-1} = \frac{1-k^n}{1-k},\ k \neq 1\right]$$

$$= P + Pi\left[\frac{(1+r)^n - 1}{r}\right] = P\left(1 + i\frac{(1+r)^n - 1}{r}\right) \qquad (24)$$

Начисление процентов в смежных календарных периодах

Выше при начислении процентов не принималось во внимание расположение срока начисления процентов относительно календарных периодов. Вместе с тем, часто даты начала и окончания ссуды находятся в двух периодах: n_1 и n_2, $n=n_1+n_2$. Ясно, что начисленные за весь срок n проценты не могут быть отнесены только к последнему периоду n_2. Алгоритм деления общей массы процентов легко сформулировать на основе графика, построенного для двух смежных календарных периодов (см. рис. 14).

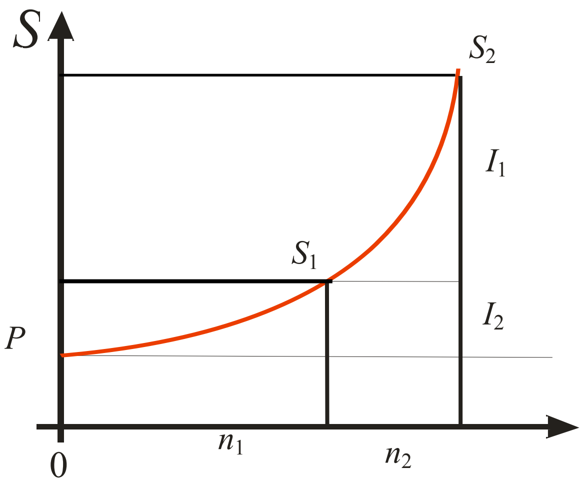


Рис. 14. Начисление сложных процентов в смежных периодах

$$S = P(1+i)^n = P(1+i)^{n_1+n_2} = \underbrace{P(1+i)^{n_1}}_{S_1}(1+i)^{n_2} = \underbrace{S_1(1+i)^{n_2}}_{S_2=S}$$

$$I_1 = S_1 - P = P(1+i)^{n_1} - P = P\left[(1+i)^{n_1} - 1\right]$$

$$I_2 = S_2 - S_1 = S_1(1+i)^{n_2} - S_1 = P(1+i)^{n_1}(1+i)^{n_2} - P(1+i)^{n_1} =$$
$$= P\left[(1+i)^{n_1+n_2} - (1+i)^{n_1}\right] = P\left[(1+i)^{n} - (1+i)^{n_1}\right] \qquad (25)$$

Пример 8. Ссуда была выдана на два года: с 1 июня 2006г. по 25 марта 2007г. Размер ссуды 50 тыс. руб. Необходимо распределить начисленные проценты (ставка 5% ACT/ACT) по календарным годам.

Решение. Вычислим количество дней в каждом периоде с помощью специальной встроенной функции в Microsoft Excel (см. рис. 15-16).

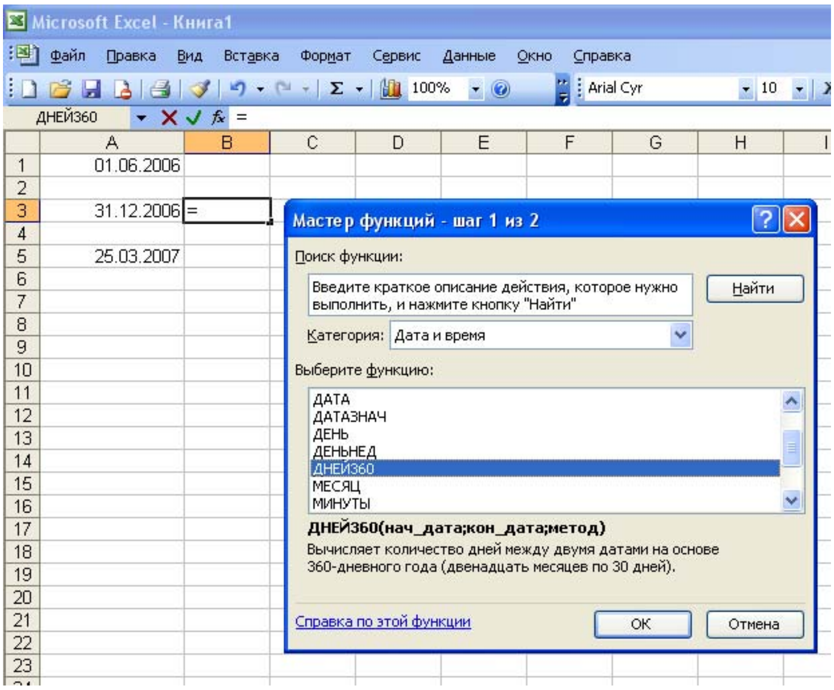


Рис. 15. Для вычисления количества дней между фиксированными датами используют функцию дней360(нач_дата;окон_дата,метод)

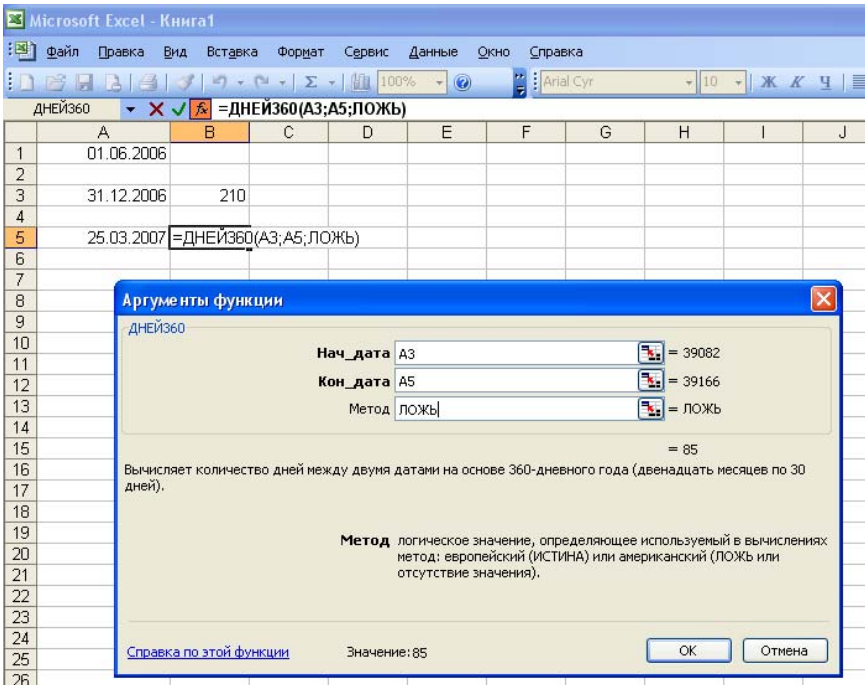


Рис. 16. В качестве параметров функции выступают даты начала и окончания финансовой операции и метод вычисления количества дней (при точном вычислении дней пишут атрибут метода –«ЛОЖЬ»)

В результате вычисления получим: n_1=210 дней, n_2=85 дней. Получим следующие суммы процентов (в тыс. руб.):
за период с 1 июня до конца года (210 дней)

$$I_1 = 50 \cdot \left(1{,}05^{\frac{210}{365}} - 1\right) = 1{,}423\ (\text{тыс.руб.})$$

$$S_1 = 51{,}423\ (\text{тыс.руб.})$$

за период с 1 января до конца периода (85 дней)

$$I_2 = 51{,}423 \cdot \left(1{,}05^{\frac{85}{365}} - 1\right) = 0{,}590\ (\text{тыс.руб.})$$

$$S_2 = 51{,}423 + 0{,}590 = 52{,}013\ (\text{тыс.руб.})$$

◆

Переменные ставки

Формула (20) предполагает постоянную ставку на протяжении всего срока начисления процентов. Неустойчивость кредитно-денежного рынка заставляет модернизировать "классическую" схему, например, с помощью применения *плавающих ставок.* В случае, когда изменения размеров ставок фиксируются в контракте, общий множитель наращения определяется как произведение частных, т.е.

$$S = P\left(1+i_1\right)^{n_1} \cdot \left(1+i_2\right)^{n_2} \cdots \left(1+i_k\right)^{n_k}, \qquad (26)$$

где, i_1, i_2,.. i_k - последовательные значения ставок; n_1, n_2,.. n_k - периоды, в течение которых используются соответствующие ставки.

Начисление процентов при дробном числе лет

Часто срок в годах для начисления процентов не является целым числом. В этом случае возможно применение одного из двух методов. Согласно первому, его называют *общим,* расчет ведется непосредственно по формуле (20). Второй – *смешанный* метод, предполагает начисление процентов за целое число лет по формуле сложных процентов и за дробную часть срока по формуле простых процентов (27):

$$S = P\left(1+i\right)^{a}\left(1+bi\right) \qquad (27)$$

где $n = a + b$ – срок ссуды, a – целое число лет, b – дробная часть года.

Аналогичный метод применяется и в случаях, когда периодом начисления является полугодие, квартал или месяц.

Сравнение роста по сложным и простым процентам

Для того чтобы сопоставить результаты наращения по сложным и простым процентным ставкам, достаточно сравнить соответствующие множители наращения. Нетрудно убедиться в том, что при одинаковых уровнях процентных ставок соотношения этих множителей существенно зависят от срока финансовой операции. Пусть временная база для начисления процентов будет одинаковой для обоих случаев, $i_{\text{пр}}$ – ставка простых процентов, $i_{\text{сл}}$ – ставка сложных процентов, тогда:

✓ для срока меньше года простые проценты больше сложных:

$$1+n\, i_{\text{пр}} > (1+ i_{\text{сл}})^n$$

✓ для срока больше года сложные проценты больше простых:

$$1+n\, i_{\text{пр}} < (1+ i_{\text{сл}})^n$$

✓ для срока, равного году, множители наращения равны друг другу.

Наиболее наглядное влияние вида ставки следует из сопоставляя числа лет, необходимые для удвоения первоначальной суммы. Из формул (2) и (20) получим *формулы удвоения:*

удвоение по простым процентам

$$2P = P(1+in)$$

$$2 = 1+in$$

$$n = \frac{1}{i}$$

удвоение по сложным процентам

$$2P = P(1+i)^n$$

$$\ln 2 = n\ln(1+i)$$

$$n = \frac{\ln 2}{\ln(1+i)}$$

Номинальная ставка

В современных условиях проценты капитализируются, как правило, не один, а несколько раз в году – по полугодиям, кварталам и т.д. При начислении процентов несколько раз в году можно воспользоваться формулой (20). Тогда параметр n будет означать число периодов начисления, а под ставкой i следует понимать ставку за соответствующий период. Например, при поквартальном начислении процентов за 3 года общее число периодов начисления составит 3 x 4 = 12. Множитель наращения по квартальной (сложной) ставке 8% равен в этом случае $1{,}08^{12} = 2{,}52$. На практике, как правило, в контрактах обычно фиксируется не ставка за период начисления, а годовая ставка, одновременно указывается период начисления процентов. Например, 8% годовых с поквартальным начислением процентов.

Пусть j – годовая ставка равна,

m – число периодов начисления в году,

N – общее количество периодов начисления.

Каждый раз проценты начисляются по ставке j/m. Ставку j называют *номинальной.* Тогда формулу наращения будет иметь вид:

$$S = P\left(1 + \frac{j}{m}\right)^N \qquad (28)$$

Эффективная ставка измеряет реальный относительный доход, который получают в целом за год, т.е. это годовая ставка сложных процентов, которая дает тот же результат, что и m-разовое начисление процентов по ставке j/m,

Обозначим эффективную ставку через i. По определению множители наращения по двум ставкам (эффективной и номинальной при m-разовом начислении) должны быть равны друг другу:

$$(1+i)^n = \left(1 + \frac{j}{m}\right)^{nm}$$

$$i = \left(1 + \frac{j}{m}\right)^m - 1 \qquad (29)$$

$$j = m\left(\sqrt[m]{1+i} - 1\right)$$

Эффективная ставка при $m > 1$ больше номинальной. Замена в договоре номинальной ставки j при m-разовом начислении процентов на эффективную ставку i не изменяет финансовых обязательств участвующих сторон. Обе ставки эквивалентны в финансовом отношении. Отсюда, следует, что разные по величине номинальные ставки оказываются эквивалентными, если соответствующие им эффективные ставки имеют одну величину.

Как и в случае простых процентов, для сложных процентов можно определить величину современной текущей стоимости P по значению будущей стоимости *8* при заданной ставке процента. В этом случае из (20) получим (30):

$$S = P(1+i)^n$$

$$P = \frac{S}{(1+i)^n} = S\left(\frac{1}{1+i}\right)^n = Sv^n \qquad (30)$$

Величину v называют *дисконтным* или *дисконтирующим множителем*. Для случаев, когда проценты начисляются m раз в году, получим (31):

$$P = \frac{S}{\left(1+\frac{j}{m}\right)^{mn}} = Sv^{mn} \qquad (31)$$

Разность ***S*** - ***P***, в случае, когда ***P*** определено дисконтированием, называют дисконтом и обозначают ***D***.

Пример 9. Сумма в 2 млн.руб. будет выплачена через 3 года. Необходимо определить ее современную величину, при условии, что применяется ставка сложных процентов, равная 13,5%

Решение.

$$P = \frac{2}{(1+0{,}135)^3} = \frac{2}{1{,}462} = 1{,}368 \text{ (млн. руб.)}$$

$$D = 2 - 1{,}368 = 0{,}632 \text{ (млн. руб.)}$$

◆

В практике учетных операций иногда применяют *сложную учетную ставку*. В этих случаях процесс дисконтирования происходит с замедлением, так как каждый раз учетная ставка применяется не к первоначальной сумме, а к сумме, дисконтированной на предыдущем шаге во времени:

$$S,\ Sd,\ Sd^2, \ldots, Sd^n$$

Дисконтирование по сложной учетной ставке осуществляется по формуле (32)

$$P = S(1-d)^n \qquad (32)$$

где d – сложная годовая учетная ставка.

Пример 10. Долговое обязательство на сумму 2 млн. руб., срок оплаты которого наступает через 3 года, продано с дисконтом по сложной учетной ставке 9% годовых. Каков размер полученной за долг суммы и величина дисконта?
Решение.

$$P = 2(1-0{,}09)^3 = 1{,}507142 \ (\text{млн. руб.})$$
$$D = 2 - 1{,}507142 = 0{,}492858 \ (\text{млн. руб.})$$

◆

Для должника дисконтирование по сложной учетной ставке выгоднее, чем по простой учетной ставке.

$$\text{если} \quad n = 1, \ \text{то} \quad w_{пр} = (1-nd) = w_{сл} = (1-d)^n,$$
$$\text{если} \quad n < 1, \ \text{то} \quad w_{пр} = (1-nd) > w_{сл} = (1-d)^n,$$
$$\text{если} \quad n > 1, \ \text{то} \quad w_{пр} = (1-nd) < w_{сл} = (1-d)^n.$$

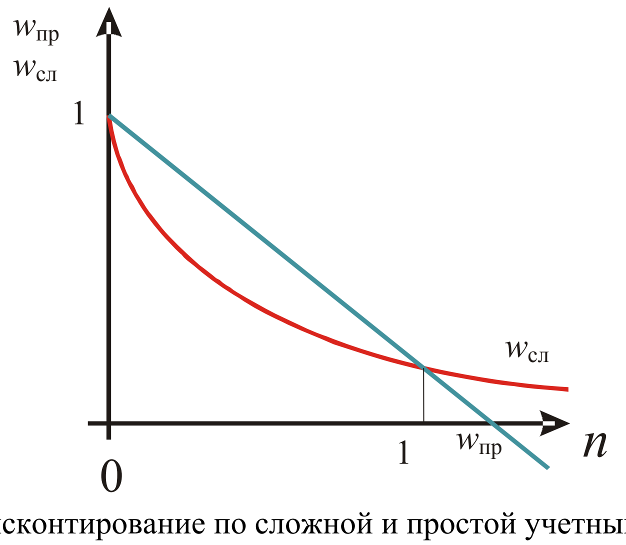


Рис. 17. Дисконтирование по сложной и простой учетным ставкам

Значение дисконтного множителя $w_{пр}$ равномерно уменьшается по мере роста $\boldsymbol{n}$ и достигает нуля при $\boldsymbol{n} = 1/\boldsymbol{d}$. Множитель $w_{сл}$ экспоненциально уменьшается и достигает нуля лишь в пределе, при $\boldsymbol{n} = \infty$.

Номинальная и эффективная учетные ставки

Дисконтирование может производиться не один, а *m* раз в году, т.е. каждый раз учет производится по ставке *f/m*. В этом случае

$$P = S\left(1 - \frac{f}{m}\right)^{mn} \tag{33}$$

где *f* – номинальная годовая учетная ставка.

Эффективная учетная ставка (*d*) характеризует степень дисконтирования за год. Определим ее на основе равенства дисконтных множителей

$$\underbrace{(1-d)^n = \left(1 - \frac{f}{m}\right)^{mn}}_{\swarrow\ \searrow}$$

$$d = 1 - \left(1 - \frac{f}{m}\right)^{m} \qquad f = m\left(1 - \sqrt[m]{1-d}\right) \tag{34}$$

Эффективная учетная ставка во всех случаях, когда $m > 1$, меньше номинальной.

Наращение по сложной учетной ставке

Иногда наращенную сумму получают с помощью сложной учетной ставки. Из формул (32) и (33) следует:

$$S=\frac{P}{\left(1-d\right)^{n}} \tag{34}$$

$$S=\frac{P}{\left(1-\frac{f}{m}\right)^{mn}} \tag{35}$$

Множитель наращения при использовании сложной ставки d равен $(1\text{-}d)^n$.

Определение срока ссуды и размера процентной ставки

При разработке условий финансовых операций часто сталкиваются с необходимостью решения обратных задач – расчетом продолжительности ссуды или уровня процентной ставки. Для сложных процентов эти задачи будут иметь вид (36) и (37).

$$S=P\left(1+i\right)^{n} \qquad\qquad S=P\left(1+\frac{j}{m}\right)^{nm}$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$n=\frac{\ln\left(\frac{S}{P}\right)}{\ln\left(1+i\right)} \qquad\qquad n=\frac{\ln\left(\frac{S}{P}\right)}{m\cdot\ln\left(1+\frac{j}{m}\right)}$$

$$S=\frac{P}{\left(1-d\right)^{n}} \qquad\qquad S=\frac{P}{\left(1-\frac{f}{m}\right)^{mn}}$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$n=\frac{\log\left(\frac{P}{S}\right)}{\log\left(1-d\right)} \qquad\qquad n=\frac{\log\left(\frac{P}{S}\right)}{m\cdot\log\left(1-\frac{f}{m}\right)} \tag{36}$$

$$i = \sqrt[n]{\frac{S}{P}} - 1 \qquad j = m \cdot \left(\sqrt[m \cdot n]{\frac{S}{P}} - 1 \right)$$
$$d = 1 - \sqrt[n]{\frac{P}{S}} \qquad f = m \cdot \left(1 - \sqrt[m \cdot n]{\frac{P}{S}} \right) \tag{37}$$

Непрерывное наращение и дисконтирование

В практических финансово-кредитных операциях непрерывное наращение, т.е. наращение за бесконечно малые отрезки времени, применяется крайне редко. Существенно большее значение непрерывное наращение имеет в анализе сложных финансовых проблем, например при обосновании и выборе инвестиционных решений, в финансовом проектировании. С помощью непрерывных процентов удается учесть сложные закономерности процесса наращения, например использовать изменяющиеся по определенному закону процентные ставки.

При непрерывном наращении процентов применяют особый вид процентной ставки – *силу роста.* Сила роста характеризует относительный прирост наращенной суммы за бесконечно малый промежуток времени. Она может быть постоянной или изменяться во времени.

При дискретном начислении процентов m раз в году по номинальной ставке j наращенная сумма имеет вид:

$$S = P\left(1 + \frac{j}{m}\right)^{nm}$$

если $m \to \infty$, то (38)

$$S = P\lim_{m \to \infty}\left(1 + \frac{j}{m}\right)^{nm} = P\left(\lim_{m \to \infty}\left(1 + \frac{j}{m}\right)^{m}\right)^{n} = Pe^{jn}$$

Для того чтобы отличить непрерывную ставку от дискретной, обозначим силу роста через δ. Тогда

$$S = Pe^{\delta n} \tag{39}$$

– постоянная сила роста.

Пусть сила роста изменяется во времени, следуя некоторому закону, представленному в виде непрерывной функции времени: $\delta_t = f\left(t\right)$. Тогда наращенная сумма и современная величина будут равны:

$$S = Pe^{\int_0^n \delta_t} \qquad P = Se^{-\int_0^n \delta_t}$$

Рассмотрим частный случай функции времени. Если функция времени будет иметь линейный вид, то

$$\text{если} \quad \delta_t = at + b, \quad \text{тогда}$$

$$\int_0^n \delta_t dt = \int_0^n \left(at + b\right)dt = \frac{an^2}{2} + bn \tag{40}$$

$$S = Pe^{\left(\frac{an^2}{2} + bn\right)} \qquad P = Se^{-\left(\frac{an^2}{2} + bn\right)}$$

Лекция 4. Производные процентные расчеты. Кривые доходности

Вопросы, рассматриваемые в лекции

1. *Средние процентные ставки.*
2, *Эквивалентные ставки.*
3. *Налоги и инфляция.*
4. *Кривые доходности.*

Средние процентные ставки

Если в финансовой операции размер процентной ставки изменяется во времени (кусочно-постоянна), то все значения ставки можно обобщить с помощью средней. Переход к усредненным процентным ставкам не меняет результатов наращения и дисконтирования.

Пусть за последовательные периоды

$$n_1,\ n_2,\ \ldots, n_k, \quad \text{причем} \quad \sum_{l=1}^{k} n_l = n$$

начисляются простые проценты по ставкам i_1, i_2, $\ldots i_k$. Приравняем множители наращения, полученные для усредненной ставки процентов $\overline{i}$ и взвешенных процентных ставок i_1, i_2, $\ldots i_k$ по периодам начисления n_1, n_2, $\ldots n_k$,

$$1+\overline{i}n = 1+\sum_{l=1}^{k} n_l i_l$$

$$\overline{i} = \frac{\sum_{l=1}^{k} n_l i_l}{\sum_{l=1}^{k} n_l} \qquad (41)$$

Аналогичный результат получим, если приравняем соответствующие простые множители наращения по учетным ставкам:

$$1-\overline{d}n = 1-\sum_{l=1}^{k} n_l d_l$$

$$\overline{d} = \frac{\sum_{l=1}^{k} n_l d_l}{\sum_{l=1}^{k} n_l} \qquad (42)$$

Найденный в (41) показатель представляет собой средне взвешенную простую процентную ставку с весами, равными продолжительности отдельных периодов.

Пример 11. Контракт предусматривает переменную по периодам ставку простых процентов: 5, 7 и 9 %. Продолжительность последовательных периодов начисления процентов: 6, 12 и 24 месяцев. Какой размер ставки приведет к аналогичному наращению исходной суммы?

Решение. Находим усредненную ставку:

$$\overline{i} = \frac{6\cdot 5+12\cdot 7+24\cdot 9}{6+12+24} = \frac{30+84+216}{42} \approx 7{,}86\%$$

◆

Теперь рассмотрим случаи со сложными процентами. Если усредняются сложные переменные во времени ставки сложных процентов, то из равенства множителей наращения имеем формулу (43) для процентной ставки и (44) для учетной ставки:

$$\left(1+\bar{i}\right)^{\sum_{l=1}^{k} n_l} = \prod_{l=1}^{k}\left(1+i_l\right)^{n_l}$$

$$\bar{i} = \sqrt[\sum_{l=1}^{k} n_l]{\prod_{l=1}^{k}\left(1+i_l\right)^{n_l}} \qquad (43)$$

$$\left(1-\bar{d}\right)^{\sum_{l=1}^{k} n_l} = \prod_{l=1}^{k}\left(1-d_l\right)^{n_l}$$

$$\bar{d} = \sqrt[\sum_{l=1}^{k} n_l]{\prod_{l=1}^{k}\left(1-d_l\right)^{n_l}} \qquad (44)$$

В этом случае средняя вычисляется как взвешенная средняя геометрическая.

Рассмотрим теперь усреднение ставок, применяемых в нескольких однородных операциях, которые различаются суммами ссуд и процентными ставками. Искомые средние ставки находим из условия равенства соответствующих сумм после наращения процентов. Так, если применяются простые ставки и сроки этих операций одинаковы (*n*), то из равенства

$$\sum_{l=1}^{k} P_l\left(1+\bar{i}n\right) = \sum_{l=1}^{k} P_l\left(1+ni_l\right)$$

$$\sum_{l=1}^{k} P_l + \sum_{l=1}^{k} P_l\bar{i}n = \sum_{l=1}^{k} P_l\left(1+ni_l\right)$$

$$\bar{i} = \frac{\sum_{l=1}^{k} P_l\left(1+ni_l\right) - \sum_{l=1}^{k} P_l}{n\sum_{l=1}^{k} P_l} = \frac{\cancel{\sum_{l=1}^{k} P_l} + \sum_{l=1}^{k} P_l n i_l - \cancel{\sum_{l=1}^{k} P_l}}{n\sum_{l=1}^{k} P_l}$$

$$\bar{i} = \frac{\sum_{l=1}^{k} P_l i_l}{\sum_{l=1}^{k} P_l} \qquad (45)$$

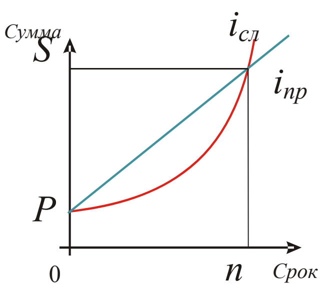
Перейдем к усреднению сложных ставок для однородных ссудных операций. Пусть сроки операций одинаковы (n). Тогда получим выражение (46):

$$\sum_{l=1}^{k} P_l \left(1+\bar{i}\right)^n = \sum_{l=1}^{k} P_l \left(1+i_l\right)^n$$

$$\left(1+\bar{i}\right)^n \sum_{l=1}^{k} P_l = \sum_{l=1}^{k} P_l \left(1+i_l\right)^n$$

$$\bar{i} = \sqrt[n]{\frac{\sum_{l=1}^{k} P_l \left(1+i_l\right)^n}{\sum_{l=1}^{k} P_l}} - 1 \qquad (46)$$

Эквивалентность процентных ставок



Для процедур наращения и дисконтирования могут применяться различные виды процентных ставок. Замена одного вида ставки на другой при соблюдении принципа эквивалентности не изменяет финансовых отношений сторон в рамках одной операции. Для участвующих в финансовой операции сторон в общем безразлично, какой вид ставки фигурирует в контракте. Такие ставки называются *эквивалентными.*

Формулы эквивалентности ставок получают исходя из равенства взятых попарно множителей наращения. Приведем в табл. 2 соотношения между эквивалентными процентными ставками.

Таблица 2.

Эквивалентные процентные ставки

Эквивалентность простых и сложных процентов	$i_{пр}=\frac{(1+i_{сл})^n}{n}-1$	$i_{сл}=\sqrt[n]{1+i_{пр}n}-1$
Эквивалентность простой ставки наращения и учетной ставки	$i_{пр}=\frac{d_{пр}}{1-nd_{пр}}$	$d_{пр}=\frac{i_{пр}}{1+ni_{пр}}$
Эквивалентность простой ставки наращения и номинальной ставки	$i_{пр}=\frac{\left(1+\frac{j}{m}\right)^{mn}-1}{n}$	$j=m\left(\sqrt[mn]{1+ni_{пр}}-1\right)$
Эквивалентность простой учетной ставки и сложной ставки наращения	$d_{пр}=\frac{1-(1+i_{сл})^n}{n}$	$i_{сл}=\sqrt[n]{1-nd_{пр}}-1$
Эквивалентность простой учетной ставки и номинальной ставки	$d_{пр}=\frac{1-\left(1+\frac{j}{m}\right)^{mn}}{n}$	$j=m\left(\sqrt[mn]{1-nd_{пр}}-1\right)$
Эквивалентность сложной ставки наращения и номинальной ставки	$i_{сл}=\left(1+\frac{j}{m}\right)^m-1$	$j=m\left(\sqrt[m]{1+i_{сл}}-1\right)$
Эквивалентность сложных учетной ставки и ставки наращения	$i_{сл}=\frac{d_{сл}}{1-d_{сл}}$	$d_{сл}=\frac{i_{сл}}{1+i_{сл}}$

Финансовая эквивалентность обязательств

На практике нередко возникают случаи, когда необходимо заменить одно денежное обязательство другим, например с более отдаленным сроком платежа, объединить несколько платежей в один (консолидировать платежи) и т.п. Такие изменения не могут быть произвольными, в основе их лежит принцип финансовой эквивалентности обязательств. Эквивалентными считают такие платежи, которые будучи "приведенными" к одному моменту времени, оказываются равными.. Приведение осуществляется путем дисконтирования или, наоборот, наращения суммы платежа.

По существу, принцип эквивалентности в наиболее простом проявлении следует из формул наращения и дисконтирования, связывающих величины P и S. Сумма P эквивалентна S при принятой процентной ставке и методе ее начисления. Две суммы денег S_1 и S_2, выплачиваемые в разные моменты времени, считаются эквивалентными, если их современные величины, рассчитанные по одной и той же процентной ставке и на один момент времени, одинаковы. Замена S_1 на S_2 в этих условиях формально не изменяет отношения сторон.

Пример 12. Имеются два обязательства. Условия первого: выплатить 100 тыс. руб. через 4 месяца; условия второго: выплатить 101,923 тыс. руб. через 6 месяцев. Можно ли считать их равноценными при ставке в 12%?

Решение. Так как платежи краткосрочные, то необходимо применять простую ставку.

$$P_1 = \frac{100}{1 + \frac{4}{12} 0.12} = 96,154 \text{ (тыс. руб.)}$$

$$P_1 = \frac{101,923}{1 + \frac{6}{12} 0.12} = 96,154 \text{ (тыс. руб.)}$$

Сравниваемые обязательства являются эквивалентными при заданной ставке и могут заменять друг друга.

◆

С ростом процентной ставки i размеры современных стоимостей уменьшаются, причем при $i = i_0$ наблюдается равенство $P_1 = P_2$. Для любой ставки $i < i_0$ имеем $P_1 < P_2$. Таким образом, результат сравнения зависит от размера ставки, равного i_0. Ставку i_0 называют *критической* или *барьерной.*

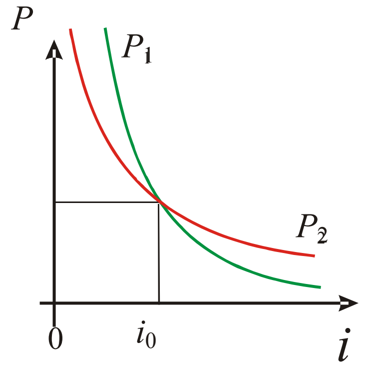


Рис. 18. Определение барьерной ставки

Формула, задающая значение барьерной ставки для случая простых процентов имеет вид (47):

$$S_1 = P_1\left(1 + i_0 n_1\right) \quad S_2 = P_2\left(1 + i_0 n_2\right)$$

$$P_1 = \frac{S_1}{1 + i_0 n_1} \qquad P_2 = \frac{S_2}{1 + i_0 n_2}$$

$$\frac{S_1}{1 + i_0 n_1} = \frac{S_2}{1 + i_0 n_2} \tag{47}$$

$$i_0 = \frac{1 - \frac{S_1}{S_2}}{\frac{S_1}{S_2} n_2 - n_1}$$

Если дисконтирование производится по сложной ставке, то аналогичные действия приводят к формуле (48)

$$i_0 = \sqrt[n_2 - n_1]{\frac{S_2}{S_1}} - 1 \tag{48}$$

Консолидирование задолженности

Одним из распространенных случаев изменения условий контрактов является *консолидация* (объединение) платежей. Пусть платежи .S_1, S_2,…S_m, со сроками n_1, n_2,…n_m заменяются одним в сумме S_0 и сроком n_0.

Определим размер консолидированного платежа. Пусть $n_1 < n_2 < \ldots < n_m$, тогда из уравнения эквивалентности при использовании простых процентных ставок получим формулу (49):

$$S_0 = \sum_j S_j\left(1 + t_j i\right) + \sum_k S_k\left(1 + t_k i\right)^{-1}, \qquad t_j = n_0 - n_j, \quad t_k = n_k - n_0 \tag{49}$$

где S_j – размеры объединяемых платежей со сроками $n_j < n_0$,

S_k – размеры платежей со сроками $n_k > n_0$

Консолидацию платежей можно осуществить и на основе сложных процентных ставок (50).

$$S_0 = \sum_j S_j \left(1+i\right)^{t_j} + \sum_k S_k \left(1+i\right)^{-t_k} \qquad (50)$$

Теперь перейдем к задаче определения срока консолидированного платежа. Если при объединении платежей задана величина консолидированного платежа S0, то возникает проблема определения его срока n0. В этом случае уравнение эквивалентности удобно представить в виде равенства современных стоимостей соответствующих платежей.

При применении простой ставки это равенство имеет вид

$$S_0 \left(1+n_0 i\right)^{-1} = \sum_j S_j \left(1+n_j i\right)^{-1}$$

$$n_0 = \frac{1}{i}\left(\frac{S_0}{\sum_j S_j \left(1+n_j i\right)^{-1}} - 1\right) \qquad (51)$$

При применении сложной ставки получим

$$S_0 \left(1+i\right)^{-n_0} = \sum_j S_j \left(1+i\right)^{-n_j}$$

$$Q = \sum_j S_j \left(1+i\right)^{-n_j} \qquad (52)$$

$$n_0 = \frac{\ln\left(\frac{S_0}{Q}\right)}{\ln\left(1+i\right)}$$

Налоги и инфляция

В ряде стран полученные (юридическими, а иногда и физическими лицами) проценты облагаются налогом, что, естественно, уменьшает реальную наращенную сумму и доходность депозитной операции.

Пусть

S – наращенная сумма до выплаты налогов,

S' – наращенная сумма с учетом налогов**,**

g – ставка налога на проценты равна*,*

G – общая сумма налога*.*

При начислении налога на проценты возможны два варианта: налог начисляется за весь срок сразу, т.е. на всю сумму процентов, или последовательно по периодам, например в конце каждого года.

При начислении простых процентов за весь срок находим

$$G = P \cdot n \cdot i \cdot g$$
$$S' = S - (S - P) \cdot g = P \cdot \left[1 + n(1 - g)i\right] \tag{53}$$

Из формулы (53) следует, что учет налога при определении наращенной суммы сводится к соответствующему сокращению процентной ставки – вместо ставки i фактически применяется ставка $(1 - g)i$*.* Размер налога пропорционален сроку.

В случае долгосрочных операций со сложными процентами будем иметь:

Определение налога за весь срок

Сумма налога равна

$$G = (S - P) \cdot g = P \cdot \left[(1 + i)^n - 1\right] \cdot g\,. \tag{54}$$

Наращенная сумма после выплаты налога составит

$$S' = S - G = P\left[(1 - g) \cdot (1 + i)^n + g\right]. \tag{55}$$

Определение налога за каждый истекший год

В этом случае сумма налога величина переменная - с ростом наращенной суммы растет и сумма налога. Налог на проценты за ***t***-й год составят

$$G_t = \left(S_t - S_{t-1}\right) \cdot g = P \cdot \left[\left(1+i\right)^t - \left(1+i\right)^{t-1}\right] \cdot g = \\ = P \cdot \left(1+i\right)^{t-1} \cdot i \cdot g \tag{56}$$

За весь срок сумма налогов будет равна

$$\sum_t G_t = \sum_t P\left(1+i\right)^{t-1} ig = P\left[\left(1+i\right)^n - 1\right]g = G \tag{57}$$

Из формул (54)-(57) следует, что метод взыскания налога не влияет на общую его сумму. Однако, для плательщика налога далеко небезразлично, когда он его выплачивает.

Инфляция

В современных условиях инфляция в денежных отношениях играет заметную роль, и без ее учета конечные результаты часто представляют собой условную величину. Инфляцию необходимо учитывать, по крайней мере, в двух случаях: при расчете наращенной суммы денег и при измерении реальной эффективности (доходности) финансовой операции.

Введем обозначения:

S - наращенная сумма денег, измеренная по номиналу,

S_α - наращенная сумма с учетом ее обесценения,

I_p - индекс цен,

I_α - индекс, характеризующий изменение покупательной способности денег за период;

h - темп инфляции.

Очевидно, что

$$S_\alpha = S\,I_\alpha \tag{58}$$

Индекс покупательной способности денег равен обратной величине индекса цен - чем выше цены, тем ниже покупательная способность:

$$I_{\alpha} = \frac{1}{I_p} \qquad (59)$$

Указанные индексы должны относиться к одинаковым интервалам времени.

Можно связать индекс цен и темп инфляции. Под темпом инфляции h понимается относительный прирост цен за период, который обычно измеряется в процентах и определяется (60):

$$h = 100\,(I_p - 1). \qquad (60)$$

Инфляция является цепным процессом. Следовательно, индекс цен за несколько периодов равен *произведению* цепных индексов цен:

$$I_p = \prod_{t=1}^{n}\left(1 + \frac{h_t}{100}\right) \qquad (61)$$

где h_t – темп инфляции в периоде t.

Если h – постоянный ожидаемый (или прогнозируемый) темп инфляции за один период, то за n таких периодов получим

$$I_p = \left(1 + \frac{h}{100}\right)^n \qquad (62)$$

Если наращение производится по простой ставке, то наращенная сумма с учетом покупательной способности равна

$$S_{\alpha} = \frac{S}{I_p} = P\frac{1+in}{I_p} = P\frac{1+in}{\left(1+\frac{h}{100}\right)^n} \qquad (63)$$

Из (63) следует, что увеличение наращенной суммы с учетом ее инфляционного обесценивания имеет место только тогда, когда $1 + in > I_p$.

Наращенная сумма с учетом инфляционного обесценивания находится как

$$S_\alpha = \frac{S}{I_p} = P\frac{(1+i)^n}{I_p} = P \cdot \underbrace{\left(\frac{1+i}{1+\dfrac{h}{100}}\right)^n}_{\substack{\text{множитель наращения,}\\ \text{учитывающий ожидаемый}\\ \text{уровень инфляции}}} \qquad (64)$$

Величины, на которые умножаются P в формулах (63) и (64), представляют собой множители наращения, учитывающие ожидаемый уровень инфляции. Если среднегодовой темп инфляции равен процентной ставке, то роста реальной суммы не произойдет – наращение будет поглощаться инфляцией, и следовательно, $S_\alpha = P$.

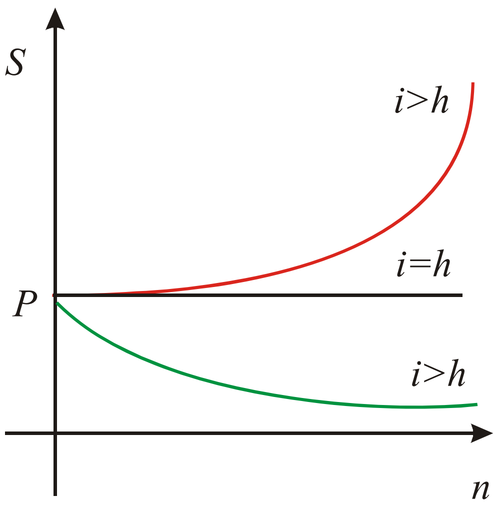


Рис. 19. Связь процентной ставки и уровня инфляции

Если $h/100 > i$, то наблюдается "эрозия" капитала – его реальная сумма будет меньше первоначальной.

Если $h/100 < i$, то происходит реальный рост, реальное накопление (см. рис. 19).

Очевидно, что при начислении простых процентов ставка, компенсирующая влияние инфляции, соответствует величине

$$i' = \frac{I_p - 1}{n}. \qquad (65)$$

Ставку, превышающую критическое значение i' (при начислении сложных процентов $i' = h$), называют положительной ставкой процента

Кривые доходности

Как уже говорилось выше, процентная ставка является измерителем доходности финансовой операции. Для практика важно представить себе закономерность изменения величины доходности (или процентных ставок, используемых в однородных по содержанию операциях), в зависимости от некоторых фундаментальных факторов. Вероятно, наиболее важным из них является *риск невозврата* вложенных средств. Очевидно также, что подобного рода риск существенно зависит от срока ссуды. Компенсировать риск владельцу денег может повышение ожидаемой доходности, договорной процентной ставки. Таким образом, зависимость "доходность – риск" приближенно можно охарактеризовать с помощью зависимости "доходность – срок" (см. рис. 20), получить которую существенно проще.

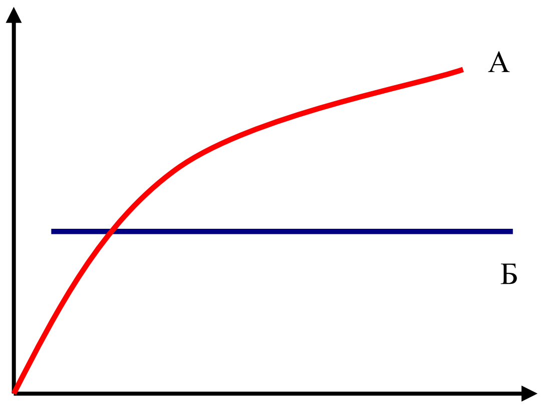


Рис. 20. Кривые доходности

На графике по вертикали откладывают доходность (Y), по горизонтали – срок (n) (см. рис. 20). Кривые доходности обычно строятся раздельно для кратко-, средне- и долгосрочных операций и однородных кредитно-ссудных операций и финансовых инструментов.

Наблюдаемые значения доходности обычно находятся около кривой или непосредственно на ней. Конкретная кривая доходности отвечает реальной ситуации, сложившейся на денежно-кредитном рынке, и характерна для короткого временного периода. Изменение ситуации меняет форму кривой и ее положение на графике.

Для нормальных экономических условий кривая доходности имеет форму кривой *А:* доходность (Y) здесь растет по мере увеличения срока. Причем каждая следующая единица прироста срока дает все меньшее увеличение доходности. Такую кривую называют *положительной,* или *нормальной,* кривой доходности. Нормальная форма кривой наблюдается в условиях, когда инвесторы в своей массе учитывают такие факторы, как рост неопределенности финансовых результатов (риска) при увеличении срока.

Кривая доходности, близкая к горизонтальной прямой (линия *Б*), указывает на то, что инвесторы не принимают во внимание или в малой степени учитывают риск, связанный со сроком.

Иногда встречаются "отрицательные" и "сгорбленные" кривые доходности. Отрицательные кривые соответствует уменьшению доходности финансового инструмента по мере роста срока (высокая нестабильность рынка, ожидание повышения процентных ставок), сгорбленные – падению доходности после некоторого ее роста.

ТЕМА 2. ПОТОКИ ПЛАТЕЖЕЙ. ПОСТОЯННЫЕ И ПЕРЕМЕННЫЕ ФИНАНСОВЫЕ РЕНТЫ

Лекция 5. Постоянные финансовые ренты

Вопросы, рассматриваемые в лекции

1. *Постоянные и переменные финансовые ренты.*
2. *Потоки платежей.*
3. *Общие вопросы и классификация.*
4. *Наращенная сумма ренты.*
5. *Современная величина ренты.*

Виды потоков платежей и их основные параметры

В предыдущих лекциях были рассмотрены различные случае разовых платежей. Однако, современные финансовые операции часто предполагают не отдельные платежи, а некоторую их последовательность во времени, например, погашение задолженности в рассрочку, периодическое поступление доходов от инвестиций, выплаты пенсии и т. д. Такого рода последовательность, или ряд платежей, называют потоком платежей. Отдельный элемент такого ряда платежей назовем *членом потока*. На рис. 21 представлена классификация основных потоков платежей.

Потоки платежей могут быть *регулярными* (размеры платежей постоянные или следуют установленному правилу, предусматривающему равные интервалы между платежами) и *нерегулярными*.

Члены потоков могут быть как положительными (поступления), так и отрицательными величинами (выплаты).

Поток платежей, все члены которого – положительные величины, а временные интервалы между платежами одинаковы называют *финансовой рентой*, или просто рентой. Рента описывается следующими параметрами: член ренты (размер отдельного платежа), период ренты (временной интервал между двумя последовательными платежами), срок ренты (время от начала первого периода ренты до конца последнего), процентная ставка.

В практике применяют разные по своим условиям ренты. В основу их классификации может быть положен ряд признаков:

✓ по количеству выплат членов ренты на протяжении года ренты делятся на годовые (выплата раз в году) и *p*-срочные;

✓ по числу раз начислений процентов на протяжении года различают: ренты с ежегодным начислением, с начислением *m* раз в году, с непрерывным начислением;

✓ по величине своих членов ренты делятся на постоянные (с одинаковыми размерами члена ренты) и переменные. Члены переменных рент изменяют свои размеры во времени, следуя какому-либо закону. Постоянные ренты – наиболее распространенный вид ренты;

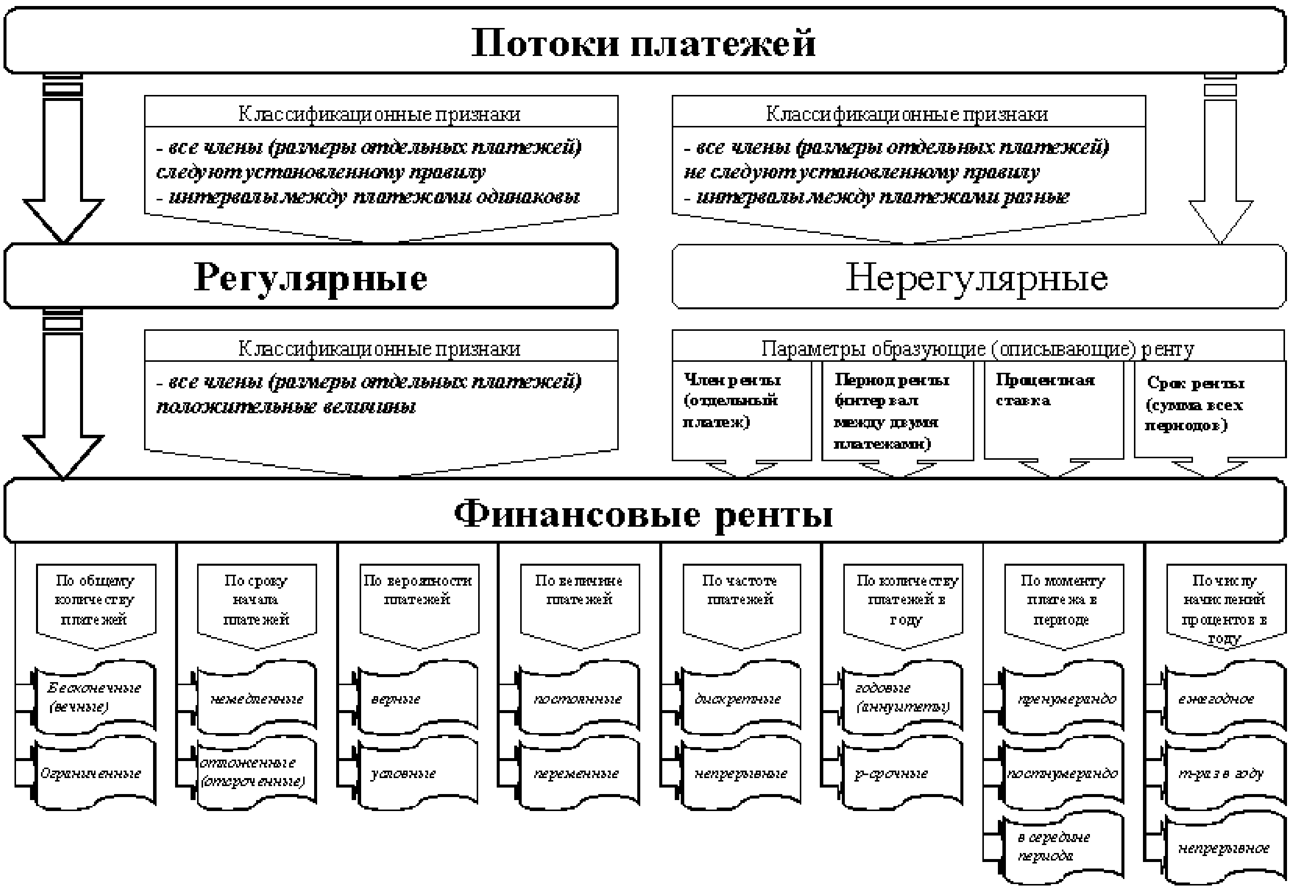


Рис. 21. Классификация потоков платежей

✓ по вероятности выплат ренты делятся на *верные* и *условные*. Верные ренты подлежат безусловной уплате, например, при погашении кредита. Число членов такой ренты заранее известно. В свою очередь выплата условной ренты ставится в зависимость от наступления некоторого случайного события, число ее членов заранее неизвестно. Примером такой ренты является *страховой аннуитет* — последовательные платежи в имущественном и личном страховании.

✓ по количеству членов различают ренты с конечным числом членов, или *ограниченные* ренты (их срок заранее оговорен), и *бесконечные,* или *вечные* ренты*;*

✓ по соотношению начала срока ренты и какого-либо момента времени, упреждающего начало ренты (например, начало действия контракта или даты его заключения), ренты делятся на *немедленные* и *отложенные,* или *отсроченные.* Пример отсроченной ренты: погашение долга в рассрочку после льготного периода;

✓ по моменту выплат платежей в пределах периода ренты деляться на. *постнумерандо* и *пренумерандо.* Иногда контракты предусматривают платежи или поступления денег в середине периодов.

Анализ потока платежей обычно предполагает расчет одной из двух обобщающих характеристик: наращенной суммы или современной стоимости потока. *Наращенная сумма потока платежей* – сумма всех членов потока платежей с начисленными на них к концу срока процентами. *Современная стоимость потока платежей* – сумма всех его членов, дисконтированных на начало срока ренты или некоторый упреждающий момент времени.

Прямой метод расчета наращенной суммы и современной стоимости потока платежей

Рассмотрим общую постановку задачи (см. рис. 22). Допустим, имеется ряд платежей R_i выплачиваемых спустя время n_i после некоторого начального момента времени. Общий срок выплат n лет. Необходимо определить наращенную на конец срока потока платежей сумму, если проценты начисляются раз в году по сложной ставке i.

Сумма первого платежа S_1 с наращенными на него за весь срок процентами будет иметь вид:

$$S_1 = R_1\left(1+i\right)^{n-n_1}$$

Для второго платежа, для которого про центы начисляются на один год меньше, получим:

$$S_2 = R_1\left(1+i\right)^{n-n_2} \text{ и т.д.}$$

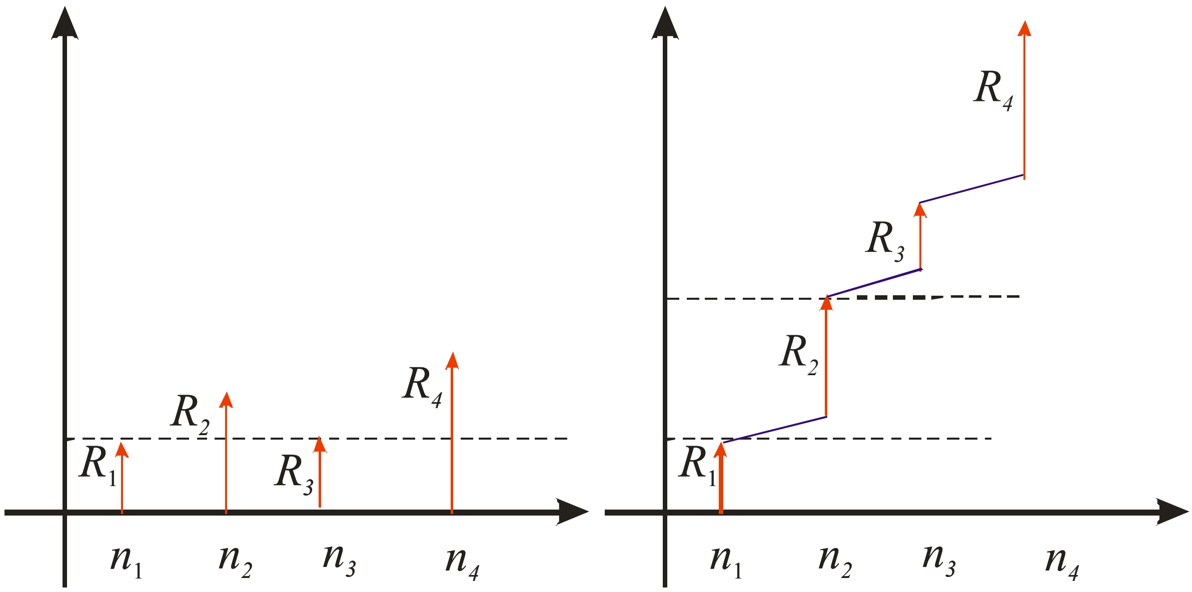


Рис. 22. Прямой метод расчета наращенной суммы потока платежей

На последний платеж, произведенный в конце последнего n-го года, проценты не начисляются

$$S_n = R_1(1+i)^{n-n} = R.$$

Тогда для всей наращенной суммы ренты S получим:

$$S = \sum_{t=1}^{n} S_t = \sum_{t=1} R_t (1+i)^{n-n_t} \qquad (66)$$

Современную стоимость такого потока также находим прямым счетом как сумму дисконтированных платежей:

$$A = \sum_{t=1} R_t (1+i)^{n_t} \qquad (67)$$

где A – современная стоимость потока платежей.

Наращенная сумма постоянной ренты постнумерандо

Годовая рента. Пусть в течение n лет в банк в конце каждого года вносится по R рублей. На взносы начисляются сложные проценты по ставке i% годовых. Таким образом, имеется рента, член которой равен R , а срок — n. Все члены ренты, кроме последнего, приносят проценты — на первый член проценты начисляются n - 1 год, на второй n - 2 и т.д. На последний взнос проценты не начисляются. Эти условия определяют задачу годовой ренты постнумерандо.

Наращенная к концу срока каждого взноса сумма составит:

$$R(1+i)^{n-1},\ R(1+i)^{n-2},\ \cdots, R(1+i),\ R$$

Эта последовательность определяет геометрическую прогрессию со знаменателем (1 + i) и первым членом R. Число членов прогрессии равно n. Тогда

$$S = R\frac{(1+i)^n - 1}{(1+i)-1} = R\underbrace{\frac{(1+i)^n - 1}{i}}_{\substack{s_{n,i}-\text{коэффициент} \\ \text{наращения ренты}}} \qquad (68)$$

Для каждого платежа современное значение стоимости определяется формулой:

$$A_t = R\frac{1}{(1+i)^t},$$

Откуда современная приведенная величина всей ренты будет определяться выражением

$$A = \sum_{t=1}^{n} A_t = R\underbrace{\sum_{t=1}^{n}\frac{1}{(1+i)^t}}_{\substack{\textit{сумма n членов геометрической} \\ \textit{прогрессии, у которой} \\ \textit{первый член и знаменатель} \\ \textit{равны } \frac{1}{1+i}}} = R\frac{1-(1+i)^{-n}}{i} \qquad (69)$$

Зная наращенную сумму и современную стоимость ренты, можно найти величину разового платежа

$$R = \frac{S}{\textit{коэф.наращ.}} = \frac{Si}{(1+i)^n - 1}$$
$$R = \frac{A}{\textit{коэф.приведения}} = \frac{Ai}{1-(1+i)^{-n}} \qquad (70)$$

Срок ренты n и ставку наращения i можно получить как решение системы уравнений, составленную из (68), (70). Это решение для срока ренты будет иметь вид (71):

$$n = \frac{\ln\left[\left(\frac{S}{R}\right)i + 1\right]}{\ln(1+i)} \quad \textit{или} \quad n = -\frac{\ln\left[1-\left(\frac{A}{R}\right)i\right]}{\ln(1+i)} \qquad (71)$$

Годовая рента, начисление процентов *m* раз в году

Рассмотрим годовую ренту постнумерандо, у которой проценты начисляются m раз в году. Число членов ренты равно nm. Члены ренты с начисленными к концу срока процентами образуют ряд:

$$R\left(1+\frac{j}{m}\right)^{(n-1)m},\ R\left(1+\frac{j}{m}\right)^{2m},\ \cdots, R\left(1+\frac{j}{m}\right)^{m},\ R$$

Сумма членов этой прогрессии составляет

$$S = R\frac{\left(1+\frac{j}{m}\right)^{nm} - 1}{\left(1+\frac{j}{m}\right)^{m} - 1} = R \cdot s_{nm,\frac{j}{m}},$$

где $s_{n;i} = \dfrac{(1+i)^n + 1}{i}$ – коэффициент наращения ренты.

p-срочная рента

Пусть рента выплачивается p раз в году равными суммами, процент начисляется m раз в конце года. Тогда будем иметь варианты, представленные в табл. 3:

Таблица 3

Постоянные финансовые ренты

Условия ренты	Наращенная сумма ренты	Современная величина ренты
$p=1;$ $m=1$	$S = R\dfrac{(1+i)^{nm}-1}{i}$	$A = R\dfrac{1-(1+i)^{-n}}{i}$
$p=1;$ $m>1$	$S = R\dfrac{\left(1+\dfrac{j}{m}\right)^{nm}-1}{\left(1+\dfrac{j}{m}\right)^{m}-1}$	$A = R\dfrac{1-\left(1+\dfrac{j}{m}\right)^{-nm}}{\left(1+\dfrac{j}{m}\right)^{m}-1}$
$p=1;$ $m\to\infty$	$S = R\dfrac{e^{\delta n}-1}{e^{\delta}-1}$	$A = R\dfrac{1-e^{-\delta n}}{e^{\delta}-1}$
$p>1;$ $m=1$	$S = R\dfrac{(1+i)^{n}-1}{p\left[(1+i)^{1/p}-1\right]}$	$A = R\dfrac{1-(1+i)^{-n}}{p\left[(1+i)^{1/p}-1\right]}$
$p>1;$ $m>1;$ $m\neq p$	$S = R\dfrac{\left(1+\dfrac{j}{m}\right)^{nm}-1}{p\left[\left(1+\dfrac{j}{m}\right)^{m/p}-1\right]}$	$A = R\dfrac{1-\left(1+\dfrac{j}{m}\right)^{-nm}}{p\left[\left(1+\dfrac{j}{m}\right)^{m/p}-1\right]}$
$p>1;$ $m\to\infty$	$S = R\dfrac{e^{\delta n}-1}{p\left(e^{\delta/p}-1\right)}$	$A = R\dfrac{1-e^{-\delta n}}{p\left(e^{\delta/p}-1\right)}$

Лекция 6. Переменные финансовые ренты

Вопросы, рассматриваемые в лекции

1. *Ренты с постоянным абсолютным приростом платежей.*
2. *Ренты с постоянным относительным приростом платежей*
3. *Постоянная непрерывная рента*

Ренты с постоянным абсолютным приростом платежей

В практике встречаются случаи, когда размеры членов потока платежей изменяются во времени. Такие изменения могут быть связаны с какими-либо обстоятельствами объективного порядка (например, условиями производства и сбыта продукции), а иногда и случайными факторами. Частным случаем такого потока является *переменная рента.* Члены переменной ренты изменяются по каким-то установленным (принятым, оговоренным и т.д.) законам или условиям развития.

Ренты с постоянным абсолютным изменением членов во времени

Изменения размеров членов ренты происходят согласно арифметической прогрессии с первым членом R и разностью a, иначе говоря, они образуют последовательность

$$R,\ R + a,\ R + 2a, \ldots,\ R + (n - 1)a.$$

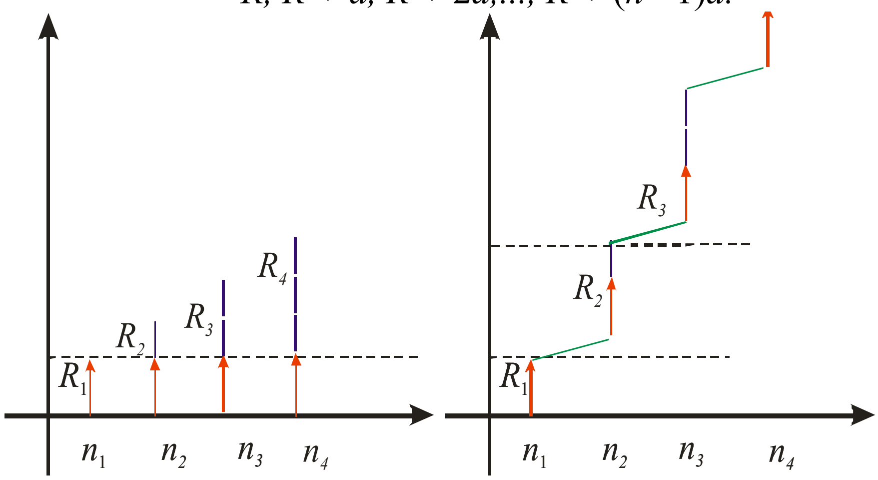


Рис. 23. Ренты с постоянным абсолютным изменением членов во времени

Величина i-го члена ренты равна R + (t -1)a. Для ренты годовой постнумерандо формулы современной стоимости и наращенной стоимости будут иметь вид:

$$A = Rv + (R+a)v^2 + (R+a)v^3 + \cdots + (R+(n-1)a)v^n = \\ = \left(R + \frac{a}{i}\right)a_{n;i} - \frac{nav^n}{i} \tag{72}$$

$$S = \left(R + \frac{a}{i}\right)s_{n;i} - \frac{na}{i}, \tag{73}$$

$a_{n;i} = \dfrac{1-(1+i)^{-n}}{i}$ – коэффициент приведения ренты,

$s_{n;i} = \dfrac{(1+i)^n - 1}{i}$ – коэффициент наращения ренты,

v – дисконтный множитель по ставке i

При решении задач может возникнуть необходимость нахождения первого члена ренты R или ее прироста a при условии, что все остальные параметры ренты известны. Например, когда известна сумма, которую необходимо аккумулировать за n лет, и необходимо разработать конкретный план реализации этой задачи. Для этого из (72), (73) выразим R:

$$R = \frac{A + \frac{nav^n}{i}}{a_{n;i}} - \frac{a}{i} \qquad R = \frac{S + \frac{na}{i}}{s_{n;i}} - \frac{a}{i} \tag{74}$$

Зная значение первого члена ренты R, можно определить размер прироста a.

Переменная p-срочная рента с постоянным абсолютным приростом

Пусть R — базовая величина разовой выплаты, a – годовой прирост выплат. В этом случае последовательные выплаты равны

$$R,\ R+\frac{a}{p},\ R+\frac{2a}{p}\ ,...,\ R+\frac{(np\text{-}1)a}{p},$$
$$R_t = R+(t-1)\frac{a}{p},\ t=\overline{1,pn} \tag{75}$$

По определению для ренты постнумерандо при начислении процентов p раз в году получим

$$A=\sum_{t=1}^{np}\left(R+\frac{at}{p}\right)v^{t/p}$$
$$S=\sum_{t=1}^{np}\left(R+\frac{a}{p}(t-1)\right)(1+i)^{(n-t)/p} \tag{76}$$

Ренты с постоянным относительным приростом платежей

Рассмотрим ситуацию, когда платежи изменяют свои размеры во времени с постоянным относительным ростом, т.е. следуют геометрической прогрессии. Поток таких платежей состоит из членов

$$R,\ Rq,\ Rq^2,...,\ Rq^{n-1},$$

q – знаменатель прогрессии или темп роста. Пусть этот ряд представляет собой ренту постнумерандо. Тогда ряд дисконтированных платежей состоит из величин

$$Rv,\ Rqv^2,\ ...,\ Rq^{n-1}v^n.$$

Это геометрическая прогрессия с первым членом Rv и знаменателем qv. Сумма членов этой прогрессии равна

$$A = Rv\frac{q^n v^n - 1}{qv - 1} = R\frac{(qv)^n - 1}{q - (1+i)} \qquad (77)$$

Пусть теперь $q = 1 + k,$ где k – темп прироста платежей. Тогда (77) после преобразований будет иметь вид:

$$A = R\frac{1 - \left(\dfrac{1+k}{1+i}\right)^n}{i - k}, \quad \text{где } k \text{ любого знака.} \qquad (78)$$

Наращенная сумма ренты в этом случае будет иметь вид:

$$S = A(1+i)^n = R\frac{q^n - (1+i)^n}{q - (1+i)} = R\frac{(qv)^n - 1}{q - (1+i)} = R\frac{(1+k)^n - (1+i)^n}{k - i} \qquad (78)$$

Постоянная непрерывная рента

Во всех рассмотренных выше рентах предполагалось, что члены потока платежей поступают дискретно – через фиксированные интервалы времени (периоды ренты). Вместе с тем иногда более адекватное описание потока платежей достигается, когда он воспринимается как *непрерывный процесс.* Например, когда отдача от инвестиций происходит так часто, что в целом этот поток можно рассматривать как непрерывный. Предположение о непрерывности в определенных условиях увеличивает возможности количественного анализа, особенно при анализе сложных производственных долгосрочных инвестиций.

Рассмотрим методы расчета наращенной суммы и современной стоимости в этом случае, а также расчет некоторых параметров, характеризующих постоянную непрерывную ренту, при условии, что применяется годовая дискретная процентная ставка.

Для непрерывной ренты $p \to \infty$. Найдем коэффициент приведения такой ренты, обозначив его как $\tilde{a}_{n;i}$. Для этого необходимо найти предел коэффициента приведения p-срочной ренты при $p \to \infty$

$$\tilde{a}_{n;i} = \lim_{p\to\infty} a_{n;i}^{(p)} = \lim_{p\to\infty} \frac{1-(1+i)^{-n}}{p\left[(1+i)^{1/p}-1\right]} =$$

$$\left[\begin{array}{ll} \text{неопределенность типа } \dfrac{1}{\infty}, & \\ \text{раскрываем по правилу} & \lim\limits_{p\to\infty} \dfrac{1}{p\left[(1+i)^{1/p}-1\right]} = \dfrac{1}{\ln(1+i)} \\ \text{Лопиталя} & \end{array}\right] \quad (79)$$

$$= \frac{1-(1+i)^{-n}}{\ln(1+i)}$$

Аналогично получим коэффициент наращения непрерывной ренты

$$\tilde{s}_{n;i} = \frac{(1+i)^{n}-1}{\ln(1+i)} \quad (80)$$

Формулы (79), (80) предполагают непрерывное поступление платежей и дискретное начисление процентов. Более "естественным" является положение, когда оба процесса (поступление денег и наращение процентов) непрерывны. Для получения формул соответствующих коэффициентов необходимо воспользоваться формулами эквивалентности между непрерывными и дискретными ставками:

$$\delta = \ln(1+i) \qquad i = e^{\delta} - 1, \quad (81)$$

где δ – сила роста. В этом случае формулы (79), (80) будут преобразованы к виду

$$\tilde{a}_{n;i} = \frac{1 - e^{-\delta n}}{\delta} \tag{82}$$

$$\tilde{s}_{n;i} = \frac{e^{\delta n} - 1}{\delta} \tag{83}$$

Следует отметить, что формулы (79), (80) и (82), (83) будут давать одинаковые результаты только в случае эквивалентных ставок.

Конверсия рент

В практике иногда сталкиваются со случаями, когда на этапе разработки условий контракта или даже в ходе его выполнения необходимо в силу каких-либо причин изменить условия выплаты ренты, т.е. произвести *конвертировании условий, предусматриваемых при выплате финансовой ренты.* Простейшими случаями конверсии являются: замена ренты разовым платежом *(выкуп ренты),* или наоборот, замена разового платежа рентой *(рассрочка* платежа). К более сложному случаю относится объединение нескольких рент с разными характеристиками в одну – *консолидация рент.* Общий случай конверсии – замена ренты с одними условиями на ренту с другими условиями.

Конверсия рент широко применяется при *реструктурировании задолженности.* При этом нередко условия погашения долга смягчаются, однако принцип эквивалентности соблюдается и в этих случаях.

Рассмотрим основные случаи конверсии рент.

Выкуп ренты. Этот вид конверсии сводится к замене ренты единовременным платежом. Искомый размер выкупа должен быть равен современной стоимости выкупаемой ренты. Применяемая при расчете современной стоимости процентная ставка должна удовлетворять обе участвующие стороны.

Рассрочка платежей. Если есть обязательство уплатить некоторую крупную сумму и стороны согласились, что задолженность будет погашена частями – в рассрочку, то последнюю удобно осуществить в виде выплаты постоянной ренты. Задача обычно заключается в определении одного из параметров этой ренты – члена ренты или ее срока – при условии, что остальные параметры заданы. Для этого приравнивают современную стоимость ренты к сумме долга инаходят неизвестные параметры.

Объединение (консолидация) рент. Объединение рент заключается в замене нескольких рент одной, параметры которой необходимо определить.

Из принципа финансовой эквивалентности следует равенство современных стоимостей заменяющей и консолидированных рент.

ТЕМА 3. ПРИКЛАДНЫЕ АСПЕКТЫ ФИНАНСОВОЙ МАТЕМАТИКИ

Лекция 7. КРЕДИТНЫЕ РАСЧЕТЫ

Вопросы, рассматриваемые в лекции

1. *Кредитные расчеты.*
2. *Расходы по обслуживанию долга*
3. *Погасительный фонд*
4. *Постоянные взносы в фонд. Изменяющиеся взносы*
5. *Погашение долга в рассрочку*
6. *Переменные расходы по займу*

Расходы по обслуживанию долга

Разработка плана погашения займа заключается в составлении графика (расписания) периодических платежей должника. Такие расходы должника обычно называют *расходами по обслуживанию долга.* Расходы по обслуживанию долга включают как текущие процентные платежи, так и средства, предназначенные для погашения основного долга.

Методы определения размера срочных уплат существенно зависят от условий погашения долга, которые предусматривают:

- ✓ срок займа,
- ✓ продолжительность *льготного периода,*
- ✓ уровень и вид процентной ставки,
- ✓ методы уплаты процентов и способы погашения основной суммы долга.

В льготном периоде основной долг не погашается, обычно выплачиваются проценты. Впрочем, не исключается возможность присоединения процентов к сумме основного долга. В долгосрочных займах проценты обычно выплачиваются на протяжении всего срока займа. Значительно реже они начисляются и присоединяются к основной сумме долга. Основная сумма долга иногда погашается одним платежом, чаще она выплачивается частями – в рассрочку.

Все методы планирования погашения долга использует результаты, полученные при анализе финансовых рент.

Введем обозначения:

D – сумма задолженности;

Y – срочная уплата;

I – проценты по займу;

R – расходы по погашению основного долга;

g – ставка процента по займу;

n – общий срок займа;

L – продолжительность льготного периода.

Расходы по обслуживанию долга (срочная уплата) находятся как

$$Y = I + R. \tag{84}$$

Если в льготном периоде выплачиваются проценты, то расходы по долгу в этом периоде сокращаются до $Y = I$.

Погасительный фонд

Если по условиям займа должник обязуется вернуть сумму долга в конце срока в виде разового платежа, то он должен предпринять меры для обеспечения этого. При значительной сумме долга обычная мера заключается в создании *погасительного фонда.* Необходимость формирования такого фонда иногда оговаривается в договоре выдачи займа в качестве гарантии его погашения. Разумеется, создание фонда необязательно надо связывать с погашением долга. На практике возникает необходимость накопления средств и по другим причинам, например, для накопления амортизационных отчислений на закупку изношенного оборудования и т.п.

Погасительный фонд создается из последовательных взносов Должника (например, на специальный счет в банке), на которые начисляются проценты. Таким образом, должник имеет возможность последовательно инвестировать средства для погашения долга. Сумма взносов в фонд вместе с начисленными процентами, накопленная в погасительном фонде к концу срока, должна быть равна его сумме. Взносы могут быть как постоянными, так и переменными во времени.

Постоянные взносы в фонд

Задача разработки способа погашения долга, в том числе и в виде плана создания погасительного фонда, заключается в определении размеров срочных уплат и составляющих их элементов в зависимости от конкретных условий займа.

Пусть накопление производится путем регулярных ежегодных взносов *R,* на которые начисляются сложные проценты по ставке *i,* которая определяет темп роста погасительного фонда. Одновременно происходит выплата процентов за долг по ставке *g,* которая определяет сумму выплачиваемых за заем процентов. В этом случае срочная уплата составит

$$Y = Dg + R. \tag{85}$$

Обе составляющие срочной уплаты постоянны во времени, при этом первое слагаемое определяется величиной долга и процентной ставкой по займу. Найдем вторую составляющую – *R*.

Пусть фонд должен быть накоплен за N лет. Тогда взносы, уплачиваемые в фонд, образуют постоянную ренту с параметрами: *R, N, i*. Пусть рента будет постнумерандо, тогда

$$R = \frac{D}{s_{N;i}}, \tag{86}$$

где $s_{N.i}$ – коэффициент наращения постоянной ренты со сроком *N*.

В целом срочная уплата будет определяться формулой:

$$Y = Dg + \frac{D}{s_{N;i}}. \tag{87}$$

Если условия контракта предусматривают присоединение процентов к сумме основного долга, то срочная уплата определяется формулой (88):

$$Y = D\frac{\left(1+g\right)^N}{s_{N;i}} \tag{88}$$

Данный способ погашения долга через создание фонда является выгодным должнику только тогда, когда $i > g$, так как в этом случае должник на аккумулируемые в погасительном фонде средства получает больше процентов, чем сам выплачивает за заем. Чем больше разность $i - g$, тем, больше экономия средств должника, направляемая на покрытие долга. В противном случае, преимущества создания фонда пропадают – финансовые результаты для должника оказываются такими же, как и при погашении долга частями.

Накопленные за t лет средства фонда определяются по формулам наращенных сумм постоянных рент или рекуррентно:

$$S_{t+1} = S_t\left(1+i\right) + R. \tag{89}$$

Изменяющиеся взносы

При решении задачи о создании погасительного фонда возможен другой подход – использование изменяющихся во времени сумм взносов. В этом случае следует использовать результаты, полученные для переменных рент.

Рассмотрим частный случай, когда взносы в фонд следуют арифметической прогрессии:

$$R,\ R + a,\ R + 2a, \ldots, R + (t\text{-}1)a;$$
$$R_t = R + a(t\text{-}1),\ t = 1, \ldots, N.$$

Тогда срочные уплаты изменяются во времени по правилу (90):

$$Y_t = Dg + R_t. \tag{90}$$

Так как взносы в фонд следуют арифметической прогрессии, разность которой равна a, первый член – R, то последняя величина определяется следующим образом:

$$R = \frac{1}{s_{N;i}}\left[D - a\frac{(1+i)^N - (1+Ni)}{i^2}\right]. \qquad (91)$$

Погашение долга в рассрочку

В практической финансовой деятельности, особенно при значительных размерах задолженности, долг обычно погашается в рассрочку, частями. Такой метод часто называют *амортизацией долга.* Он осуществляется различными способами:

- ✓ погашением *основного долга* равными суммами,
- ✓ погашением *всей задолженности* равными или переменными суммами по обслуживанию долга.

Погашение основного долга равными суммами

Пусть долг в сумме D погашается в течение n лет. В этом случае сумма, ежегодно идущая на его погашение, составит

$$d = \frac{D}{n}.$$

Размер долга последовательно сокращается:

$$D,\ D - d,\ D - 2d \ldots \text{ и т.д.}$$

При этом уменьшаются и выплачиваемые проценты, так как они начисляются на остаток долга. Пусть проценты выплачиваются раз в конце года по ставке g. Тогда за первый год и последующие годы они равны

$$Dg,\ (D - d)g,\ (D - 2d)g \ \ldots \text{и т.д.}$$

Процентные платежи образуют убывающую арифметическую прогрессию с первым членом Dg и разностью —dg.

Срочная уплата в конце первого года находится как

$$\mathrm{Y}_1 = Dg + d.$$

Для конца года t срочная уплата будет равна

$$Y_t = D_{t\text{-}1}g + d,\ t=1,\dots n, \tag{92}$$

где D_t – остаток долга на конец года t.

Остаток долга можно определять последовательно через реккурентное соотношение:

$$D_t = D_{t\text{-}1}\frac{n-1}{n}. \tag{93}$$

Если долг погашается p раз в году постнумерандо и с такой же частотой выплачиваются проценты, каждый раз по ставке g/p, то срочная уплата составит:

$$Y_t = \frac{D_{t\text{-}1}g}{p} + \frac{D_0}{pn},\ t=1,\dots np, \tag{94}$$

Остаток задолженности на конец года t в этом случае составит

$$D_t = D_{t\text{-}1}\frac{np-1}{np}. \tag{95}$$

У рассмотренного метода амортизации задолженности есть одно положительное свойство – простота расчетов. Однако, как мы только что убедились, в начале срока срочные уплаты погашения выше, чем в конце его, что часто является нежелательным для должника.

Погашение долга равными срочными уплатами

В соответствии с этим методом расходы должника по обслуживанию долга постоянны на протяжении всего срока его погашения. Из общей суммы расходов должника часть выделяется на уплату процентов, остаток идет на погашение основного долга. Так же как и при предыдущем методе, величина долга здесь последовательно сокращается, в связи с этим уменьшаются процентные платежи и увеличиваются платежи по погашению основного долга. По определению

$$Y_t = D_{t-1}g + R_t = const. \tag{92}$$

План погашения обычно разрабатывается при условии, что задается срок погашения долга. Тогда первый этап разработки плана погашения – определение размера срочной уплаты. Далее полученная величина разбивается на процентные платежи и сумму, идущую на погашение долга. После чего легко найти остаток задолженности.

Периодическая выплата постоянной суммы Y равнозначна ренте с заданными параметрами. Приравняв сумму долга к современной величине этой ренты, находим

$$Y = \frac{D}{a_{n;g}}, \tag{93}$$

где $a_{n;g}$ – коэффициент приведения годовой ренты со ставкой g и сроком $n.$

Все величины, необходимые для разработки плана, можно рассчитать на основе величины Y и данных финансового контракта. Сумма первого погасительного платежа будет равна

$$D_1 = Y - Dg \tag{94}$$

Суммы, идущие на погашение долга, увеличиваются во времени:

$$d_t = d_{t-1}\,(1 + g), \tag{95}$$

В связи с этим рассматриваемый метод погашения называют *прогрессивным.* Платежи по погашению долга образуют ряд

$$d_1,\, d_1\left(1+g\right), \ldots, d_1\left(1+g\right)^{n-1}.$$

В этом случае, сумма погашенной задолженности на конец года t после очередной выплаты будет равна:

$$W_t = \sum_{k=0}^{t-1} d_1\left(1+g\right)^k = d_1 s_{t;g}, \tag{96}$$

где $s_{t;g}$ – коэффициент наращения постоянной ренты постнумерандо.

Аналогичным образом разрабатываются планы погашения и для случаев, когда выплата процентов и погашение основного долга производятся не один, а несколько раз в году.

Переменные расходы по займу

Погашение долга может быть связано с поступлением средств из каких-либо источников и зависеть от ряда обстоятельств, в этом случае условие Y=const становится не удобным. Выход из этой ситуации лежит в использовании переменных расходах по займу. Срочные уплаты в этом случае образуют ряд, члены которого либо задаются заранее (график погашения), либо следуют какому-либо формальному закону.

Рассмотрим случай, когда изменении расходов происходит по геометрической прогрессии. Ряд срочных уплат в этом случае будет представлять собой геометрическую прогрессию со знаменателем q:

$$Y,\ Yq,\ Yq^2,\ \ldots,\ Yq^{n-1}.$$

Приравняв современную стоимость этой ренты сумме первоначального долга, находим:

$$Y = D\frac{q-(1+g)}{\left(\frac{q}{1+g}\right)^n - 1}, \qquad (97)$$

где q – заданный годовой темп роста платежей, g – процентная ставка по займу. Далее находятся срочные уплаты и разрабатывается детальный план погашения.

В ряде случаев размеры срочной уплаты связываются с ожидаемыми поступлениями средств и задаются заранее в виде графика погашения. Размер последней срочной уплаты не задается. Она определяется как сумма остатка долга на начало последнего периода.

Лекция 8. АНАЛИЗ ИНВЕСТИЦИОННЫХ ПРОЦЕССОВ

Вопросы, рассматриваемые в лекции

1. *Анализ эффективности инвестиций в облигации.*
2. *Анализ инвестиционных процессов.*

Виды облигаций и их рейтинг

Наиболее распространенным видом ценных бумаг с фиксированным доходом являются облигации. При необходимости привлечения значительных денежных ресурсов (для финансирования крупных проектов, покрытия текущих расходов и т.д.) государство, муниципалитеты, банки и другие финансовые институты, а также отдельные компании или их объединения, часто прибегают к выпуску и продаже облигаций. Под облигацией понимается ценная бумага, свидетельствующая о том, что ее держатель предоставил заем эмитенту этой бумаги.

Облигация обеспечивает ее владельцу некоторый доход – в большинстве случаев он регулярно получает проценты по купонам и в конце срока выкупную цену. Доход от облигаций обычно ниже, чем от других видов ценных бумаг, в то же время он более надежен, так как в меньшей степени зависит от конъюнктурных колебаний и фазы экономического цикла, чем, например, доход от акций. В связи со сказанным в облигации инвестируют свободные резервы пенсионные фонды, страховые компании, различного рода инвестиционные фонды и т.д.

Основные параметры облигации:

- ✓ номинальная цена или номинал,
- ✓ выкупная цена или правило ее определения, если она отличается от номинала,
- ✓ дата погашения,
- ✓ норма доходности или купонная процентная ставка,
- ✓ даты выплат процентов.

В современной практике выкуп по номиналу является преобладающим.Выплаты процентов производятся ежегодно, по полугодиям или поквартально, а иногда в конце срока.

Виды облигаций

В практике применяются облигации различных видов, которые можно классифицировать по следующим признакам:

По методу обеспечения:

- ✓ государственные и муниципальные облигации, выплаты по которым обеспечиваются гарантиями государства или муниципалитета;
- ✓ облигации частных корпораций, которые обеспечиваются залогом имущества, передачей прав на недвижимость, доходами от различных программ и проектов;
- ✓ облигации частных корпораций без специального обеспечения.

По сроку:

- ✓ облигации с фиксированной датой погашения*;*
- ✓ без указания даты погашения или бессрочные*,* эмитент не связывает себя конкретным сроком, соответственно облигации могут быть выкуплены в любой момент.

По способу погашения номинала (выкупа облигации):

- ✓ разовым платежом*;*
- ✓ распределенными во времени погашениями оговоренных долей номинала;
- ✓ последовательным погашением доли общего количества облигаций, соответствующие облигации называются *серийными;* часто этот метод по-

гашения осуществляется с помощью лотерей — *лотерейные* или *тиражные займы.*

По методу выплаты дохода:

- ✓ выплачиваются только проценты, срок выкупа не оговаривается;
- ✓ выплата процентов не предусматривается – *облигации с нулевым купоном;*
- ✓ проценты выплачиваются вместе с номиналом в конце срока;
- ✓ периодически выплачиваются проценты, а в конце срока – номинал или выкупная цена;

Процентная ставка обычно используется постоянная, правда, иногда выпускают облигации с "плавающей" процентной ставкой, уровень которой ставится в зависимость от каких-либо внешних условий.

Облигации являются важным объектом долгосрочных инвестиций. С момента их эмиссии и до погашения они продаются и покупаются на кредитно- денежном рынке по *рыночным ценам.* Рыночная цена в момент выпуска может быть равна номиналу, ниже номинала, или с дисконтом, и выше номинала, или с премией.

Различают два вида рыночных цен. Облигации реализуются по так называемой *грязной* или *полной* цене, которая включает не только собственно цену облигации, но и сумму процентов за период после последней их выплаты и до момента продажи. Рыночная цена за вычетом суммы начисленных процентов называется *чистой*. Далее будем рассматривать чистые цены.

Определенное значение имеет наличие оговорки о запрете досрочного выкупа облигации эмитентом. Наличие права досрочного выкупа снижает качество облигации, так как повышает степень неопределенности для инвестора. Более того, в случаях, когда облигация куплена с премией, досрочный ее выкуп заметно сокращает доходность инвестиции.

Поскольку номиналы у разных облигаций существенно различаются между собой, то возникает необходимость в сопоставимом измерителе рыночных цен. Таким показателем является *курс облигации.* Под курсом понимают цену одной облигации в расчете на 100 денежных единиц номинала:

$$K = \frac{P}{N}100, \tag{98}$$

где K – курс облигации, P – рыночная цена, N – номинал облигации.

Доход от облигации состоит из двух основных слагаемых:

- ✓ периодически получаемых по купонам процентов;
- ✓ разности между номиналом и ценой приобретения облигации.

Количественный анализ облигаций заключается в следующем:

- ✓ расчет доходности облигаций и ряда дополнительных характеристик,

- ✓ определение расчетной цены облигации на любой момент ее "жизни",
- ✓ измерение динамики дисконта или премии по облигации.

Рейтинг облигаций

Инвестиции в ценные бумаги сопряжены с определенным риском. Выделим два основных вида риска — *кредитный* (возможность отказа в выплате процентов и основной суммы долга) и *рыночный* (связан с колебаниями рыночной цены облигации, в значительной мере определяемыми изменениями процентной ставки на денежно-кредитном рынке).

Кредитный риск характеризует кредитоспособность и надежность самого эмитента. Качество облигаций в отношении кредитного риска оценивается специальными агентствами путем отнесения конкретных облигаций к той или другой категории. Такая операция называется *рейтингом.*

Доходность облигаций

Доходность облигаций характеризуется несколькими показателями. Различают *купонную, текущую* и *полную доходности.*

Купонная доходность определена при выпуске облигации и, следовательно, нет необходимости ее рассчитывать. Текущая доходность характеризует отношение поступлений по купонам к цене приобретения облигации. Этот параметр не учитывает второй источник дохода – получение номинала или выкупной цены в конце срока. Поэтому он непригоден при сравнении доходности разных видов облигаций.

Наиболее информативным является показатель полной доходности, который учитывает оба источника дохода. Именно этот показатель пригоден для сравнения доходности инвестиций в облигации и в другие ценные бумаги. Итак, полная доходность или, применив старую коммерческую терминологию, *ставка помещения,* измеряет реальную эффективность инвестиций в облигацию для инвестора в виде годовой ставки сложных процентов. Иначе говоря, начисление процентов по ставке помещения на цену приобретения облигации строго эквивалентно выплате купонного дохода и сумме погашения облигации в конце срока.

Рассмотрим методику определения показателей доходности различных видов облигаций.

Облигации без обязательного погашения с периодической выплатой процентов

Пусть выплата номинала в необозримом будущем во внимание не принимается. Введем обозначения:

g – объявленная норма годового дохода (купонная ставка процента);

i_t – текущая доходность;

i – полная доходность (ставка помещения).

Текущая доходность в этом случае равна:

$$i_t = \frac{gN}{P} = \frac{g}{K}100. \tag{99}$$

Если по купонам выплата производится p раз в году, каждый раз по ставке g/p, то и в этом случае на практике применяется формула (99).

Поскольку купонный доход постоянен, то текущая доходность продаваемых облигаций изменяется вместе с изменением их рыночной цены. Для владельца облигации, который уже инвестировал в нее некоторые средства, эта величина постоянна.

Теперь рассмотрим случай полной доходности. Так как доход по купонам является единственным источником текущих поступлений от данного вида облигации, то очевидно, что полная доходность у рассматриваемых облигаций равна текущей в случае, когда выплаты по купонам ежегодные, т.е. $i=i_t$,. Если же проценты выплачиваются раз в году, каждый раз по норме g/p, то получим:

$$i = \left(1 + \frac{g}{p}\frac{100}{K}\right)^p - 1 = \left(1 + \frac{i_t}{p}\right)^p - 1. \tag{100}$$

Облигации без выплаты процентов

Данный вид облигации обеспечивает ее владельцу в качестве дохода разность между номиналом и ценой приобретения. Курс такой облигации всегда меньше 100. Для определения ставки помещения приравняем современную стоимость номинала цене приобретения:

$$v^n = \frac{P}{N} = \frac{K}{100}, \tag{101}$$

где n – срок до выкупа облигации,

$$i = \frac{1}{\sqrt[n]{\frac{K}{100}}} - 1. \tag{102}$$

Облигации с выплатой процентов и номинала в конце срока

Проценты в этом случае начисляются за весь срок и выплачиваются одной суммой вместе с номиналом. Купонного дохода нет. Поэтому текущую доходность условно можно считать нулевой, поскольку соответствующие проценты получают в конце срока.

Найдем полную доходность, приравняв современную стоимость дохода цене облигации:

$$\begin{gathered}(1+g)^n N v^n = P, \\ \left(\frac{1+g}{1+i}\right)^n = \frac{K}{100}, \\ i = \frac{1+g}{\sqrt[n]{\frac{K}{100}}} - 1.\end{gathered} \tag{103}$$

Облигации с периодической выплатой процентов и погашением номинала в конце срока

Для такой облигации можно получить все три показателя доходности – купонную, текущую и полную. Текущая доходность рассчитывается по формуле (99). Для определения полной доходности необходимо знать современную стоимость всех поступлений, приравненных цене облигации. Поскольку поступления по купонам представляют собой постоянную ренту постнумерандо, то член такой ренты равен *gN,* а современная ее стоимость составит gNan;I, если купоны оплачиваются ежегодно; и $gNa_{n;i}^{(p)}$, если эти выплаты производятся *p* раз в году, каждый раз по ставке *g/p.* Дисконтированная величина номинала равна *Nv".* В итоге получим равенство для облигации с годовыми купонами

$$P = Nv^n + gN\sum_{t=1}^{n} v^t = Nv^n + gNa_{n;i},$$
$$\frac{K}{100} = v^n + ga_{n;i}. \tag{104}$$

Для облигации с погашением купонов по полугодиям и поквартально применяют

$$\frac{K}{100} = v^n + ga_{n;i}^{(p)}, \tag{105}$$

где $a_{n;i}^{(p)}$ – коэффициент приведения *p*-срочной ренты (*p* = 2, *p* = 4).

Измерители финансовой эффективности производственных инвестиций

Финансовый анализ производственных инвестиций в основном заключается в *измерении конечных финансовых результатов инвестиций – их доходности для инвестора.* С такой задачей сталкиваются как на этапе первоначального анализа финансовой привлекательности проекта, так и при разработке бизнес-плана. Отрицательный вывод обычно дает основание отказаться от дальнейшего, более основательного и углубленного, изучения проекта. Без расчета такого рода измерителей нельзя осуществить и сравнение альтернативных инвестиционных проектов. В анализе используют в основном четыре основных показателя (см. рис. 24), которые отражают результат сопоставления обобщенных, суммарных отдач от инвестиций со стоимостью самих инвестиций.

Рис. 24. Основные показатели эффективности инвестиционного проекта

Под *чистым приведенным доходом* понимается разность дисконтированных показателей чистого дохода (положительные величины) и инвестиционных затрат (отрицательные величины). Чистый приведенный доход представляет обобщенный конечный результат инвестиционной деятельности в абсолютном измерении.

Относительной мерой эффективности реализации инвестиционного проекта является *внутренняя норма доходности*. Этот параметр характеризует такую расчетную процентную ставку, которая при ее начислении на суммы инвестиций обеспечит поступление предусматриваемого (ожидаемого) чистого дохода, т.е. эта ставка "уравновешивает" инвестиции и доходы, распределенные во времени.

По определению современная стоимость доходов, полученных за *дисконтный срок окупаемости* $n_{ок}$, должна быть равна сумме инвестиций, т.е. окупить инвестиции с учетом разновременности получаемых доходов. Срок окупаемости можно рассматривать как характеристику риска – чем он больше, тем выше вероятность изменения условий для получения ожидаемого дохода.

Индекс доходности равен отношению современной стоимости поступлений к стоимости инвестиций. Он близок к показателю *рентабельности.*

Потоки платежей в инвестиционном анализе

Основная задача при разработке модели, с помощью которой намереваются проанализировать долгосрочный инвестиционный проект заключается в формировании ожидаемого *потока платежей.* Первым шагом в этом направлении является разработка структуры потока во времени — разбивка его на этапы, различающиеся своим содержанием и закономерностями в изменении доходов и затрат. При этом должны быть приняты во внимание как ожидаемые внешние условия (например, динамика цен на продукцию), так и производственные параметры (объемы производства, уровень производственных затрат и т.д.).

Отдельные отрезки потока платежей могут быть представлены в виде постоянных или переменных дискретных, или непрерывных рент. Сформированные таким путем показатели затрат и поступлений дают возможность определить члены потока платежей для каждого временного интервала.

В общем виде *член потока платежей* для каждого временного интервала определяется следующим образом при условии, что выплачивается налог на прибыль:

$$R = (G- C)-(G- C- D)T- K+ S, \quad (106)$$

где R – член потока платежей;

G – ожидаемый общий доход от реализации проекта , сумма выручки за период;

C – текущие расходы;

D – расходы, на которые распространяются налоговые льготы;

T– налоговая ставка;

K – инвестиционные расходы;

S – различные компенсации, сокращающие текущие затраты.

Уравнение (106) характеризует общий подход к определению члена потока, однако оно разработано без учета источников финансирования. Если в потоке платежей учитывается привлечение заемных средств, их погашение и выплата процентов, то член потока будет иметь вид:

$$R = (G – C – I) – (G – C – D – I)T – K + B – P + S, \quad (107)$$

где I – сумма выплаченных процентов за заемные средства;

B – полученные в текущем году заемные средства;

P – погашение основного долга.

В (107) предполагается, что суммы выплат процентов за кредит освобождаются от налогов.

Нетрудно видеть, что в первые годы реализации проекта члены потока – отрицательные величины, так как затраты превышают поступления. Нельзя исключить ситуации, когда отрицательными оказываются потоки платежей и в отдельные годы периода эксплуатации, например, в связи с модернизацией технологического процесса.

Инвестиционные затраты включают все виды расходов, необходимых для реализации проекта: проектно-изыскательские работы, закупка лицензий, заказ и оплата оборудования, строительство, монтаж и наладка оборудования и т.д. Что касается поступлений от инвестиций, то в расчет принимаются только чистые доходы. Причем, под чистым доходом понимается не бухгалтерская прибыль, а доход, полученный в каждом временном отрезке, за вычетом всех реальных расходов, связанных с его созданием. При этом в расходах *амортизационные списания не учитываются,* так как соответствующие затраты сделаны раньше – при инвестировании средств.

Ставка приведения

Важным является выбор уровня процентной ставки для дисконтирования – *ставка приведения.* Чем ставка выше, тем в большей мере отражается фактор времени, так как отдаленные платежи оказывают все меньшее влияние на современную стоимость потока.

В западной литературе ставку, принятую для дисконтирования потоков платежей в инвестиционном проекте, рассматривают как *минимально привлекательную ставку доходности.* В отечественной практике эту ставку, вероятно, следует рассматривать как *норматив доходности,* приемлемый для инвестора. Если финансовый анализ предполагает расчет внутренней нормы доходности, то этот норматив рассматривается как некоторое пороговое (минимально допустимое) значение данного параметра.

При выборе ставки приведения нельзя не учитывать финансовое положение инвестора, его способность учесть будущие условия и т.п. Важным моментом является учет риска. Риск в инвестиционной деятельности независимо от ее конкретных форм в конечном счете проявляется в виде возможного сокращения отдачи от вложенного капитала по сравнению с ожидаемой ее величиной. В качестве общей рекомендации по учету риска сокращения отдачи, инфляционного обесценения дохода, изменения конъюнктуры и других отрицательных факторов обычно предлагается вводить *рисковую надбавку* к уровню ставки приведения.

Чистый приведенный доход

Чистый приведенный доход применим при решении широкого круга финансовых проблем, в том числе при определении различных показателей эффективности. Рассмотрим методы его расчетов.

Пусть капиталовложения и доходы представлены в виде потока платежей, тогда искомая величина находится как современная стоимость этого потока, определенная на начало действия проекта:

$$N = \sum_{t=1}^{n} R_t v^t, \tag{108}$$

где R_t – размер члена потока платежей в году t, v — дисконтный множитель по ставке i (ставка приведения, принятая норма доходности).

Так как члены потока платежей являются как положительными, так и отрицательными, то положительной или отрицательной может быть и величина N.

Пусть теперь поток платежей представлен как поток инвестиций и поток чистых доходов, тогда чистый приведенный доход определяется разностью

$$N = \sum_{j=1}^{n_2} R_j v^{j+n_1} - \sum_{t=1}^{n_1} K_t v^t,, \tag{109}$$

где K_t – инвестиционные расходы в году t, $t = 1,2,...,n_1$; R_j – чистый доход в году j, $j = 1,2,..., n_2$; n_1 — продолжительность инвестиционного периода; n_2 — продолжительность периода поступлений дохода.

В случаях, когда поток доходов можно описать как постоянную или переменную ренту, расчет N упрощается. Так, если доходы поступают в виде постоянной годовой ренты, причем ожидается, что они равномерно распределены в пределах года, то

$$N = Ra_{n_2;i} v^{n_1 - 0.5} - \sum_{t=1}^{n_1} K_t v^t, \tag{110}$$

где R — годовая сумма дохода.

Если капиталовложения мгновенны, а доходы регулярно поступают сразу после инвестирования, то

$$N = Ra_{n;i} - K. \qquad (111)$$

Внутренняя норма доходности

Под *внутренней нормой доходности* понимают такую расчетную ставку приведения, при которой капитализация получаемого дохода дает сумму, равную инвестициям, и, следовательно, капиталовложения только окупаются. Иначе говоря, при начислении на сумму инвестиций процентов по ставке, равной внутренней норме доходности – *J,* обеспечивается получение распределенного во времени дохода, эквивалентного инвестициям. Чем выше эта норма, тем, разумеется, больше эффективность инвестиций. Этот параметр может быть как положительной, так и отрицательной величиной.

Пусть i – приемлемый для инвестора уровень ставки процента. Разность ставок $J - i$ характеризует эффективность инвестиционной деятельности. С чисто финансовых позиций инвестиции имеют смысл только тогда, когда $J > i$. При $J < i$ нет оснований для осуществления инвестиций, так как доходность ниже принятого норматива, если же под i понимается стоимость заемных средств, то они просто убыточны.

Расчет внутренней нормы доходности часто применяют в качестве первого шага анализа инвестиций. Для дальнейшего анализа отбираются только те проекты, которые обеспечивают некоторый приемлемый для данной компании уровень доходности.

В общем случае, когда инвестиции и доходы задаются в виде потоков платежей, искомая ставка J определяется на основе решения уравнения относительно v:

$$\sum_{t=1}^{n} R_t v^t = 0. \qquad (112)$$

где v – дисконтный множитель по искомой ставке J;

t – время от начала реализации проекта;

R_t – член потока платежей (вложения и чистые доходы).

Затем, зная v, можно найти искомую ставку J.

Срок окупаемости

Срок окупаемости, как уже отмечено выше, определяется в двух вариантах — на основе дисконтированных членов потока платежей – $n_{ок}$ и без дисконтирования. Обозначим первый как, второй как m. Величина $n_{ок}$ характеризует число лет, которое необходимо для того, чтобы сумма дисконтированных на момент окончания инвестиций чистых доходов была равна размеру инвестиций (барьерная точка для срока). Т.е. это расчетное время, необходимое для полной компенсации инвестиций поступающими доходами с дисконтированием обоих потоков по ставке приведения. Второй показатель не учитывает время получения доходов и доходы не дисконтируются.

В предельно простом случае срок окупаемости m определяется как отношение суммы инвестиций к средней ожидаемой величине поступаемых доходов:

$$m = \frac{K}{R}.$$

Такой расчет имеет смысл при относительно незначительных колебаниях годовых доходов относительно средней. В финансовом отношении более обоснованным является дисконтный срок окупаемости $n_{ок}$.

Пусть размеры капитальных вложений к концу срока инвестирования составляют величину K. Доходы поступают в виде нерегулярного потока платежей R_r. Необходимо найти такой срок, при котором будет выполнено равенство

$$\sum_{t=1}^{n_{ок}} R_t v^t = K. \tag{113}$$

Пусть теперь капиталовложения заданы одной суммой, а поток доходов постоянен и дискретен (постоянная ограниченная рента). Тогда из условия полной окупаемости за срок $n_{ок}$ при заданной процентной ставке i и ежегодных поступлений постнумерандо следует:

$$K = R\frac{1-(1+i)^{-n_{ок}}}{i} \Rightarrow$$
$$n_{ок} = \frac{-\ln\left(1-\frac{K}{R}i\right)}{\ln(1+i)}. \qquad (114)$$

Заметим, что дисконтный срок окупаемости существует, если не нарушаются определенные соотношения между доходами и размером инвестиций, а именно: если постоянные доходы поступают ежегодно, то $R > iK$. Это вытекает из формулы (114).

Индекс доходности

Рентабельность инвестиций может быть измерена двумя способами: бухгалтерским и с учетом фактора времени (с дисконтированием членов потока платежей). В обоих случаях доход сопоставляется с размером инвестиций. На основе бухгалтерского метода получим два варианта индекса доходности:

$$u = \frac{\sum_j R_j}{K} \quad \text{или} \quad u = \frac{\sum_j R_j - K}{K} = \underbrace{\frac{\sum_j R_j}{K} - 1}_{\text{индекс рентабельности}} \qquad (115)$$

При дисконтировании членов потока платежей индекс доходности определяется через приведение капиталовложений к одной сумме K, т.е.

$$U = \frac{\sum_{j=1} R_j v^j}{K}. \qquad (116)$$

Если же капитальные затраты распределены во времени, то

$$U = \frac{\sum_j R_j v^{j+n_1}}{\sum_t K_t v^t}. \qquad (117)$$

Если поток доходов представляет собой постоянную ренту постнумерандо, а капиталовложения мгновенны, то

$$U = \frac{R}{K} a_{n;i}. \qquad (118)$$

Лекция 9. СТРАХОВЫЕ АННУИТЕТЫ

Вопросы, рассматриваемые в лекции

1. *Финансовая эквивалентность в страховании*
2. *Таблицы смертности*
3. *Страховые вероятности*
4. *Стоимость страхового аннуитета*

Финансовая эквивалентность в страховании

В преобладающем числе областей финансовой деятельности объектами приложения методов количественного анализа являются детерминированные процессы, описываемые верными рентами. Однако в страховании и при анализе некоторых инвестиционных проектов возникает необходимость в использовании условных рент, в которых важную роль играют вероятности наступления соответствующих событий. Выплата члена ренты в страховании зависит от наступления страхового события. Заранее число платежей в страховых аннуитетах, а часто и их срок, остаются неизвестными.

Согласно договору страхования страхователь уплачивает вперед страховщику некоторую сумму — *премию.* В свою очередь он имеет право получить страховую сумму S после наступления страхового события. Если вероятность наступления этого события q заранее известна, то теоретически, без учета всех прочих факторов (в том числе и фактора времени), премия P определяется как

$$P = Sq. \tag{119}$$

Премия, которая поступает страховой организации, обычно превышает величину нетто-премии, т.к. включает помимо нетто-премии и *нагрузку, которая* охватывает все расходы по ведению дела и некоторую прибыль страховой организации. *Брутто-премии* равна нетто-премии плюс нагрузка.

Пусть P – размер премии, q_n — вероятность страхового события. Если страховое событие произойдет на первом году страхования, то страховщик получит сумму P, если же это событие наступит во втором году, то сумма премий составит $2P$ и т.д. Математическое ожидание такого ряда премий составит:

$$Pq_1 + 2Pq_2 + \ldots + nPq_n \tag{120}$$

Однако, если учесть, что премии выплачиваются в разные моменты времени, то математическое ожидание современной стоимости взносов составит:

$$M(A) = P\left[q_1 + (1+v)q_2 + \ldots + \left(1 + v + \ldots + v^{n-1}\right)q_n\right], \tag{121}$$

где v –дисконтный множитель по ставке i.

Теперь рассмотрим выплату страховой суммы. Пусть она выплачивается в конце года, в котором имел место страховой случай. Тогда математическое ожидание выплаты составит:

в первом году – Sq_1,

во втором году – Sq_1, и т.д.

Математическое ожидание с учетом фактора времени (актуарная стоимость) выплат будет равна:

$$M(S) = S\left[vq_1 + v^2 q_2 + \ldots + v^n q_n\right]. \tag{122}$$

Исходя из принципа эквивалентности обязательств страховщика и страхователя можно написать равенство:

$$M(A)=M(S),$$

которое позволяет найти искомое значение нетто-премии P.

Пусть теперь речь идет об имущественном страховании. Если можно полагать, что вероятность наступления страхового случая статистически устойчива (постоянна), то актуарная стоимость премий за n лет составит

$$M(A) = P\left[q + (1+v)q + \ldots + \left(1 + v + \ldots + v^{n-1}\right)q\right] = PqK, \quad K = n + \sum_{t=1}^{n-1}(n-t)v^t. \qquad (123)$$

В свою очередь актуарная стоимость выплат страховых сумм находится как

$$M(S) = Sq\sum_{t=1}^{n} v^t. \qquad (124)$$

Из равенства актуарных стоимостей взносов и выплат находим искомый размер нетто-премии.

В практике страховых, или как их часто называют, *актуарных расчетов* разработаны специальные приемы формирования упомянутых выше потоков платежей (страховых аннуитетов) и расчета их актуарных стоимостей.

Таблицы смертности

Для осуществления актуарных расчетов, в том числе расчетов стоимостей страховых аннуитетов, необходимы исходные данные, характеризующие совокупность застрахованных по полу и возрасту, а также система нормативных демографических показателей, отражающих статистические закономерности дожития до того или иного возраста. Последние содержатся в *таблицах смертности.*

Таблица смертности представляет собой *числовую модель процесса вымирания по возрастам некоторой абстрактной совокупности людей.* Такая таблица показывает, как последовательно с увеличением возраста уменьшается эта совокупность, достигая нуля сразу после предельного возраста ω. Она является обобщением данных демографической статистики за некоторый период времени.

Основной показатель таблицы смертности — число людей ℓ_x в возрасте ровно x лет, оставшихся в живых из первоначальной совокупности ℓ_0, обычно равной 100 тыс. человек.

Величины ℓ_x определяются расчетным путем на основе заданных вероятностей смерти (q_x), или, что реже, количества умерших (d_x). В современных таблицах смертности исходным показателем обычно служит вероятность смерти, т.е доля умерших в возрасте от x до x + 1 лет из числа доживших до возраста x-лет. Указанные вероятности получают на основе данных статистики населения с последующим их усреднением и сглаживанием.

Помимо показателей ℓ_x таблица смертности содержит число умерших за год в каждой возрастной группе (d_x). Никакие иные факторы выбытия, кроме повозрастных вероятностей умереть, при разработке таблицы во внимание не принимаются.

Показатели таблицы смертности ℓ_x и d_x связаны соотношениями:

$$\begin{gathered} \ell_{x+1} = \ell_x - d_x \\ d_x = \ell_x \cdot q_x \\ q_x = 1 - p_x = 1 - \frac{\ell_{x+1}}{\ell_x} = \frac{d_x}{\ell_x} \end{gathered} \tag{125}$$

Страховые вероятности

На основе данных таблицы смертности нетрудно получить систему вероятностей дожития, необходимую для расчета соответствующих страховых показателей. Рассмотрим наиболее важные из таких вероятностей.

Вероятность прожить от возраста x до $x + n$:

$$_n p_x = \frac{\ell_{x+n}}{\ell_x}. \tag{126}$$

Вероятность прожить еще один год после возраста x лет:

$$p_x = 1 - q_x = 1 - \frac{d_x}{\ell_x} = \frac{\ell_{x+1}}{\ell_x}. \qquad (127)$$

Вероятность умереть через m лет (на протяжении года $m + 1$) для лица в возрасте x лет составит:

$$_m q_x = {}_m p_x \cdot q_{x+m} = \frac{\ell_{x+m}}{\ell_x} \cdot \frac{d_{x+m}}{\ell_{x+m}} = \frac{d_{x+m}}{\ell_x}.$$

В свою очередь вероятность для лица в возрасте x лет умереть в возрастном интервале от $x + m$ до $x + m + n$ лет равна:

$$_{m|n} q_x = \frac{\ell_{x+m} - \ell_{x+m+n}}{\ell_x} = \frac{\ell_{x+m}}{\ell_x} \cdot \frac{\ell_{x+m} - \ell_{x+m+n}}{\ell_{x+m}} =$$

$$= {}_m p_x \cdot {}_n q_{x+m}$$

Стоимость страхового аннуитета

Основой актуарного анализа является определение стоимости страхового аннуитета. Введем обозначения для стоимостей годовых аннуитетов постнумерандо:

a_x – для немедленного пожизненного аннуитета,

$a_{x;t|}$ – для немедленного ограниченного аннуитета,

$_{n|}a_x$ – для отложенного пожизненного аннуитета,

$_{n|}a_{x;t|}$ – для отложенного ограниченного аннуитета.

Пусть лицу, начиная с возраста x лет, пожизненно в конце каждого года выплачивается по 1 рублю (аннуитет пожизненный, постнумерандо, немедленный). Тогда

$$ax = p_x v + {}_2 p_x v^2 + \ldots + {}_{\omega-x} p_x v^{\omega-x} =$$

$$= \frac{\ell_{x+1} v}{\ell_x} + \frac{\ell_{x+2} v^2}{\ell_x} + \ldots + \frac{\ell_\omega v^{\omega-x}}{\ell_x}$$

Предметный указатель

СПИСОК ЛИТЕРАТУРЫ

Основная литература

1. Башарин Г.П. Начала финансовой математики. М.: ИНФРА - М, 1998. - 160с.
2. Бухвалов А.В., Бухвалова В.В., Идельсон А.В. Финансовые вычисления для профессионалов / Под общ. ред. А.В. Бухвалова. - СПб.: БХВ-Петербург, 2001.-320с: ил.
3. Ващенко Т.В. Математика финансового менеджмента. - М.: Перспектива, 1996.-82 с.
4. Вебер М. Коммерческие расчеты от А до Я. Формулы, примеры расчетов и практические советы: Пер. с нем. - М.: Дело и Сервис, 1999. - 384 с.
5. Гармаш А.Н., Орлова И.Н., Федосеев В.В., Половников В.А. "Экономико-математические методы и прикладные модели". Учебное пособие для вузов - 2-е издание. М.: ЮНИТИ-ДАНА, 2005.
6. Ершов Ю.С. Финансовая математика в вопросах и ответах: Учебное пособие. - Новосибирск: Сибирское соглашение, 1999. - 160с.
7. Капитоненко В.В. Финансовая математика и ее приложения: Учебно-практ. пособие для вузов. - М.: ПРИОР, 1999. - 144с.
8. Ковалев В.В. Сборник задач по финансовому анализу: Учеб. пособие. -М.: Финансы и статистика, 2000. 128с.
9. Ковалев В.В., Уланов В.А. Курс финансовых вычислений. -М.: Финансы и статистика, 1999. - 328с:
10. Количественные методы финансового анализа / Под ред. С. Дж. Брауера, М.П. Криумена: Пер. с англ. - М.: Инфра-М, 1996.
11. Кочович Е. Финансовая математика: Теория и практика финансово-банковских расчетов: Пер. с серб. / Предисл. Е.М. Четыркина. - М.: Финансы и статистика, 1994. - 268с
12. Кутуков В.Б. Основы финансовой и страховой математики. М.: Дело, 1998.
13. Лукасевич И.Я. Анализ финансовых операций. Методы, модели, техника вычислений. - М.: Финансы, ЮНИТИ, 1998. - 400с
14. Лукашин Ю.П. Основы финансовой математики: Учебное пособие —М.: Изд-во МЭСИ, 1999. - 60с
15. Малыхин В.И. Финансовая математика: Учеб. Пособие для вузов. - М.: ЮНИТИ-ДАНА, 2000.
16. Медведев Г.А. Начальный курс финансовой математики: Учеб. пособие. - М.: ТОО «Остожье», 2000. - 267с.
17. Морошкин В.А., Морошкина СВ. Простые и сложные проценты. Методическое пособие по расчету: вкладов, кредитов, платежей. - М.: АКАЛИС -Бизнес-книга, 1996. - 80с.
18. Статистика финансов: Учебник / Под ред. проф. В.Н. Салина. - М.:Финансы и статистика, 2000. - 816 с.

19. Финансовая математика: Метод, указания для студентов дневного отделения эконом, фак-та/ Краснояр. гос. ун-т.; Сост. И.С. Пыжев, В.П. Зуев. -Красноярск, 2001. 24с
20. Фомин Г.П. Финансовая математика: 300 примеров и задач: Учеб. пособие. - М.: Гном-Пресс, 2000. - 120с.
21. Четыркин Е.М. Финансовая математика: Учебник. - М.: Дело, 2002 (2000,2001).- 400с.
22. Ширяев А.Н. Основы стохастической финансовой математики. -М.: Фазис, 1998.Т.1,2.

Дополнительная литература на русском языке

23. Бирман Г., Шимдт С. Экономический анализ инвестиционных проектов: Пер. с англ. под ред. Л.П. Белых. – М.: Банки и биржи, ЮНИТИ, 1997.
24. Богатин Ю.В., Швандар В.А. Инвестиционный анализ: Учеб. пособие для вузов. – М., ЮНИТИ-ДАНА, 2000.
25. Ван Хорн Дж. К. Основы управления финансами: Пер. с англ. / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 1996.
26. Ковалев В.В. Методы оценки инвестиционных проектов. – М.: Финансы и статистика, 2001.
27. Уотшем Т. Дж., Паррамоу К. Количественные методы в финансах: Учебное пособие для вузов / Пер. с англ. Под ред. М.Р. Ефимовой. – М.: Финансы, ЮНИТИ, 1999. – 527с.
28. Хелферт Э. Техника финансового анализа / Пер. с англ. Под ред. Л.П. Белых. – М.: Аудит, ЮНИТИ, 1996. – 663с.
29. Четыркин Е.М. Методы финансовых и коммерческих расчетов. – 2-е изд., испр. и доп. - М.: «Дело Лтд», 1995. – 320с.
30. Четыркин Е.М., Васильева Н.Е. Финансово-экономические расчеты. М.: Финансы и статистика, 1990.

Дополнительная литература на русском языке

31. Baxter M., Rennie A. Financial Calculus. Cambridge: Cambridge Univ. Press 1996.
32. Bosch K. Finanzmamematik.R. Oldenburg Verlag: Munchen, Wien. 1992.
33. Jones R., Mackey J. Business Mathematics and Information Technology Longmann Group, UK Ltd, 1988.
34. Karatzas I., Shreves S. Methods of Mathematical Finance. New York: Columbic Univ. Press, 1995.
35. Rogers L.C.G., Talay D. (Eds.) Numerical Methods in Finance. Cambridge Cambridge Univ. Press, 1997.

36. Varian H.R. (Ed.) Economic and Financial Modelling with Mathematica (TELOS - The Electronic Library of Science). Berlin: Springer-Verlag, 1993.

37. Willmott P., Howison S. Dewynne J. The Mathematics of Financial Derivatives (A student instruction). Cambridge: Cambridge Univ. Press, 1996.

Информационные ресурсы

Интернет-ресурсы в поддержку курса «Финансовая математика» можно найти на сайте:
www.cfin.ru/finanalysis

В том числе:

Доработка алгоритма прогнозирования объема продаж
http://www.cfin.ru/finanalysis/math/add_to_kosh.shtml
Развитие алгоритма прогнозирования, предложенного Кошечкиным С.А.

Учебник "Техника финансовых вычислений на Excel"
http://www.cfin.ru/finanalysis/smirnova/index.shtml
Базовые понятия финансовой математики и рекомендации по выполнению расчетов

Анализ операций с ценными бумагами с Microsoft Excel
http://www.cfin.ru/finanalysis/inexcel/index.shtml
Учебник по финансовой математике с рекомендациями по применению Excel в сложных финансовых расчетах

Алгоритм прогнозирования объёма продаж в MS Excel
http://www.cfin.ru/finanalysis/sales_forecast.shtml
Прогноз объёма реализации для продуктов с сезонным характером продаж

Математический аппарат для инвестора
http://www.cfin.ru/press/afa/97_3_164-219.shtml
1. Методика статистического анализа и прогнозирования
2. Дескриптивная статистика
3. Анализ временных рядов
4. Прогнозирование временных рядов
5. Корреляционный анализ
6. Регрессионный анализ
7. Факторный и компонентный анализ
8. Кластерный анализ
9. Частотный анализ
10. Работа с математическим аппаратом на компьютере

Соросовский образовательный журнал

РАЗДЕЛ 4 (МОДУЛЬ 4). ТЕОРИЯ ИГР

ТЕМА 4.1. ТЕОРИЯ ИГР КАК ИНСТРУМЕНТ КОЛИЧЕСТВЕННОГО ОПИСАНИЯ ЭКОНОМИЧЕСКИХ СУБЪЕКТОВ В УСЛОВИЯХ СУБЪЕКТИВНОЙ НЕОПРЕДЕЛЕННОСТИ

Лекция 4.1.1. Основные положения теории игр. Классификация игр

Теория игр рассматривает задачи принятия решений в ситуациях с несколькими участниками, когда достижение цели для каждого из субъектов зависит о решения, принимаемых всеми остальными участниками.

Теория игр впервые систематически изложена Нейманом и Моргенштерном только в 1944 г., в работе «Теория игр и экономическое поведение»[1], хотя отдельные результаты были опубликованы в еще в 20-е годы. Нейман и Моргенштерн главным образом приводили экономические примеры, поскольку экономическому конфликту легче всего придать численную форму. Во время второй мировой войны и сразу после нее теорией игр серьезно заинтересовались военные, которые увидели в ней математический аппарат для исследования стратегических решений. Среди социальных наук аппарат теории игр используется, например, в психологии для анализа торговых сделок и переговоров.

Математическая теория игр является инструментом исследования экономических процессов и принятия решений в условиях неопределенности, когда участники не обладают полной информацией о процессе. В случае с несколькими участниками неопределенность для каждого участника связана с неполной информацией о поведении других участников, которые стремятся оптимизировать свои решения за счет других, при этом используется общее допущение, что все участники ведут себя рационально, т.е. не выбирают не оптимальных.

Терминология теории игр в значительной степени заимствована из терминологии общественных игр: спортивные игры, шахматы, карточные игры (преферанс, покер, бридж), шашки и другие.

[1] Фон Нейман Дм., Моргенштейн О., Теория игр и экономическое поведение. – М.: Наука, 1970 (Von Neuman J., Morgenstern., Teory of Games and Economic Behavior, Princeton University Press. Prenciton, Wew Jersey, 1944 г.)

В теории игр участников, принимающих решение, называют **игроками**. Это могут быть отдельные лица или группы лиц.

Каждый игрок имеет некоторое множество возможных выборов, называемых **стратегиями**. Целевые функции игроков называются **платежными функциями**, которые определяют **выигрыш (платеж)** каждого игрока в зависимости от выбранной им стратегии.

Каждый игрок должен учитывать зависимость получаемого им платежа от остальных игроков при выборе стратегии в общем случае. Стратегия – это набор правил, формулируемых до игры, которые определяют выбор варианта в любой из могущих возникнуть ситуаций. Так как понятие стратегии является в теории игр центральным, то эту дисциплину нередко называют «стратегическими играми»

Таким образом, игра представляет собой совокупность известных всем игрокам правил, которые определяют, что может делать игрок и каковы последствия и выигрыши в результате каждого отдельного их действия.

Ход – это момент игры, когда игроки должны произвести выбор одного из возможных вариантов. Партией игры называется некоторая определенная совокупность ходов и выборов. Подчеркнем еще раз, что чертой любой игры является то, что выигрыш каждого игрока зависит обычно не только от сделанного им самим выбора, но и от выбора других игроков.

Различные виды игр можно **классифицировать**, основываясь на том или ином принципе: **по числу игроков** или по **числу стратегий**; по **свойствам платежной функции** или по **характеру предварительных переговоров между игроками до игры**; по **способу списания игры**.

В зависимости от числа игроков различают игры двух и n игроков. Первые из них наиболее изучены. Игры трех и более игроков менее исследованы из-за возникающих принципиальных трудностей и технических возможностей получения решения. Чем больше игроков – тем больше проблем.

Наиболее простой пример выбора стратегий в условиях неопределенности можно представить как игру 2-х лиц, где один из игроков – «природа», т.е. игроку, принимающему решение, не противостоит разумный противник, стремящийся сознательно изменить выигрыш игрока.

При наличии двух разумных игроков могут возникнуть и конфликтные ситуации, и необходимость в координированных действиях (кооперация). Когда число игроков не меньше трех, могут создаваться коалиции – группы из двух или более игроков, имеющих общую цель и координирующих свои стратегии.

Согласно другому признаку классификации – по количеству стратегий – различают конечные и бесконечные игры. Если в игре все игроки имеют конечное число возможных стратегий, то она называется **конечной**. В данном модуле, как правило, будут рассматриваться игры, в которых каждый из игроков располагает конечным числом возможных стратегий.

Рассматриваемые примеры, включают число стратегий равно двум или трем, однако эта же математическая теория пригодна и для игр со сколь угодно большим числом стратегий. Если хотя бы один из игроков имеет бесконечное количество возможных стратегий, игра также является **бесконечной**.

По характеру платежей игры делятся на: игры с **нулевой суммой** (общий выигрыш всех игроков не меняется, а перераспределяется между игроками; сумма выигрышей всех игроков равна нулю) и игры с **ненулевой суммой**. Данный признак лежит в основе деления игры на: антагонистические и неантагонистические. Игры с нулевой суммой пример антагонистических игр.

По виду функций выигрыша игры делятся на : матричные, биматричные, непрерывные, выпуклые и др.

Матричная игра – это конечная игра двух игроков с нулевой суммой, в которой задается выигрыш игрока 1 в виде матрицы (строка матрицы соответствует номеру применяемой стратегии игрока 2, столбец – номеру применяемой стратегии игрока 2; на пересечении строки и столбца матрицы находится выигрыш игрока 1, соответствующий применяемым стратегиям).

Для матричных игр доказано, что любая из них имеет решение и оно может быть легко найдено путем сведения игры к задаче линейного программирования.

Биматричная игра – это игра двух игроков с ненулевой суммой, в которой выигрыши каждого игрока задаются матрицами отдельно для соответствующего игрока (в каждой матрице строка соответствует стратегии игрока 1, столбец – стратегии игрока 2, на пересечении стройки и столбца в первой матрице находится выигрыш игрока 1, во второй матрице – выигрыш игрока 2).

Непрерывной считается игра, в которой функция выигрышей каждого игрока является непрерывной в зависимости от стратегии. Доказано, что игры этого класса имеют решение, однако не разработано практически приемлемых методов их нахождения.

Если функция выигрышей является выпуклой, то такая игра называется **выпуклой**. Для них разработаны приемлемые методы решения, состоящие в отыскании чистой оптимальной стратегии (определенного числа) для одного игрока и вероятностей применения чистых оптимальных стратегий другого игрока.

Если функция выигрышей является выпуклой, то такая игра называется выпуклой. Для них разработаны приемлемые методы решения, состоящие в отыскании чистой оптимальной стратегии (определенного числа) для одного игрока и вероятностей применения чистых оптимальных стратегий другого игрока.

В зависимости от характера предварительной договоренности между игроками различают кооперативные (коалиционные) и некооперативные (бескоалиционные) игры.

В игре двух участников с нулевой суммой выигрыш одного из партнеров равен проигрышу другого, т.е. налицо прямой конфликт между игроками. Прямой противоположностью играм такого типа являются игры двух участников с постоянной разностью, в которых игроки и выигрывают, и проигрывают одновременно, так что им выгодно действовать сообща. В общей игре с ненулевой суммой имеют место, как правило, и конфликты, и согласованные действия игроков. Игра называется кооперативной (коалиционной), если до начала игры игроки образуют коалиции и принимают взаимообязывающие соглашения о своих стратегиях. Игра, в которой игроки не могут координировать свои стратегии подобным образом, называется некооперативной (бескоалиционной).

В экономической практике с помощью коалиционных игр можно описать ситуации, связанные с формированием ряда монопольных объединений, например, картельные соглашения.

Способ классификации игр по способу описания и анализа игры, когда указывается какие ходы могут делать игроки, какой информацией во время игры они располагают, какие варианты стратегий можно выбирать и какими могут быть предельные размеры платежей в конце игры. Игра, описанная таким образом, называется **игрой в развернутой** (экстенсивной) форме, а само описание составляется обычно в виде дерева игры. Такой тип игр получил название – **позиционные игры**.

Игра в развернутой форме является игрой с полной информацией, если в ней нельзя делать одновременно несколько ходов, и если участникам известны выборы, сделанные при предшествующих ходах, включая и случайные ходы. Примерами игр с полной информацией являются шахматы и игра в крестики и нолики. Покер, напротив, представляет собой игру с неполной информацией, так как игрокам не известны некоторые выборы, сделанные при случайных ходах, - прежде всего им не известно, какие карты находятся на руках у противника.

Другой способ описания игры состоит в том, что рассматриваются все возможные стратегии каждого игрока и определяются платежи, соответствующие любой возможной комбинации стратегий двух игроков.

Описанная таким образом игра называется игрой в нормальной форме. Зная развернутую форму игры, можно получить и ее нормальную форму. Нормальная форма игры двух участников состоит из двух платежных матриц, показывающих, какую сумму получает каждый из игроков при любой из возможных пар стратегий.

Позиционные игры и нормальная форма игры.

В позиционной игре $\underline{K}$ лиц разрешенные ходы указаны в их логической последовательности. Каждый ход производится либо игроком (личный ход), либо выбирается случайным образом (случайный ход). В первом случае игрок выбирает ход, имея полную информацию, во втором случае задается распределение вероятностей ходов. В каждой окончательной позиции игры значение платежа или исхода можно выразить при помощи вектора ($П_1$, … , $П_к$), где $П_к$ – выигрыш к-го игрока при данном исходе.

В общем случае, как уже упоминалось, позиционная игра может быть представлена в виде дерева с выделенном узлом (называемым начальной позицией игры). Каждый узел представляет определенную позицию игры; каждая дуга – ход в игре. Информация задается при помощи информационных множеств. Две позиции принадлежат одному и тому же информационному множеству, если один и тот же игрок, которому следует ходить в каждой из этих позиций, не может отличить одну позицию от другой.

Стратегия представляет собой некоторое правило, описывающее действие игрока, т.е. указывает, какую альтернативу следует выбирать в каждом информационном множестве.

Рассмотрим в качестве примера очень упрощенный вариант карточной игры в покер. В начале игры каждый игрок делает единичную ставку на «кон». Колода карт (содержащая m «картинок» и n без «картинок») тасуется, и одна из карт сдается игроку 1. Посмотрев на свою карту, игрок может спасовать или поставить некоторую сумму денег a .

Если игрок 1 пасует, игра на этом прекращается; игрок выигрывает (ставку) единицу, если у него картинка, и проигрывает, если у него карта без картинки. Если игрок 1 увеличивает ставку, игрок 2, пренебрегая картой игрока 1, должен решить: пасовать ему или поднимать ставку (объявить игру)?

Если игрок 2 пасует, игрок 1 выигрывает ставку (единицу). Если игрок 2 поднимает ставку, то карты открываются и игрок 1 выигрывает (1+a), если у него «картинка», и проигрывает (1+a) при карте без картинки.

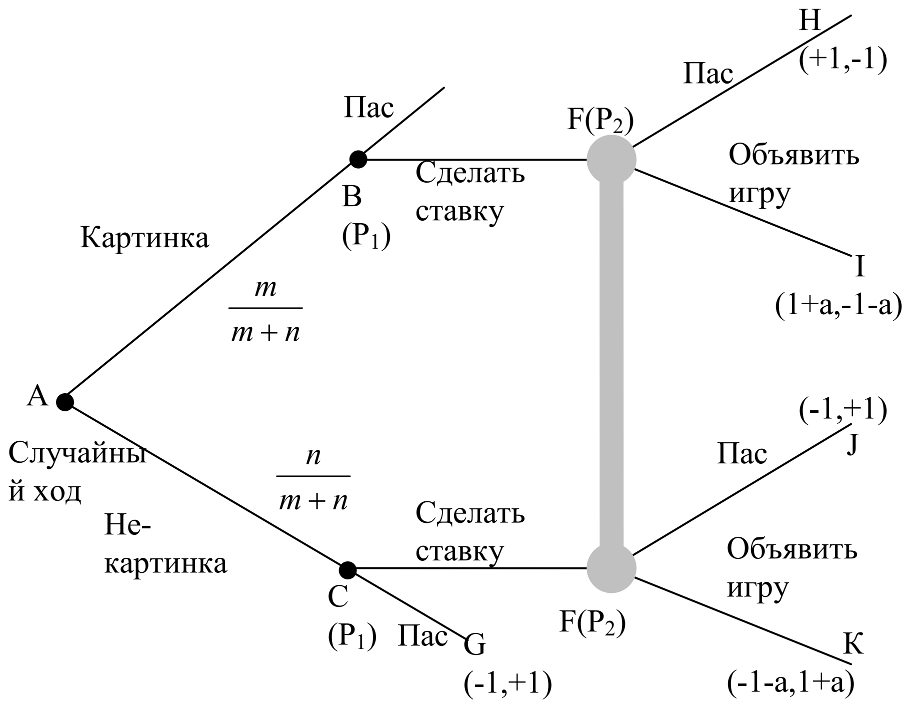


Рис. 4.1.1. Дерево игры в покер (упрощенный вариант)

На рис. 4.1.1 эта игра представлена в виде дерева. Узлы от A до K обозначают позиции игры. Позиция A является случайной (получение карт из перетасованной колоды), B и C принадлежат игроку 1, а E и F – игроку 2. Другие узлы (H,I,J,K) являются окончательными вершинами, для каждой из которых установлен вектор платежей, представляющий выигрыш каждого из игроков.

Игрок 1 имеет два информационных множества – две вершины: B и C. Игрок 2 имеет только одно информационное множество, содержащее как E, так и F, поскольку для игрока 2 эти вершины неразличимы. (Напомним, что игрок 1 делает ставку, зная исход случайного хода, т.е. знает карту, а игрок 2 должен сделать ставку или спасовать, не зная этой карты.)

Поскольку игрок 1 имеет два информационных множества и по два выбора (пасовать или поднять ставку) для каждого случая, то он имеет 4 - стратегии, которые можно обозначить через ВП (всегда поднимать ставку), ПК (поднимать, имея картинку), ПН (поднимать, не имея картинки) и П (всегда пасовать). Его противник имеет только одно информационное множество и два выбора в этом множестве. Следовательно, у него две стратегии, а именно ПС (поднимать ставку) или П (пасовать).

Другой способ описания игры состоит в том, что рассматриваются все возможные стратегии каждого игрока и определяются платежи, соответствующие любой возможной комбинации стратегий двух игроков.

Описанная таким образом игра называется **игрой в нормальной форме**. Зная развернутую форму игры, можно получить и ее нормальную форму. Нормальная форма игры двух участников описывается платежными матрицами, показывающими какую сумму получает каждый из игроков при любой из возможных стратегий.

Для примера игры в покер (рис. 4.1.1 можно записать нормальную форму игры. Для этого для каждой стратегии подсчитывает ожидаемые платежи. Так, если игрок 1 выбирает стратегию ВП (всегда поднимать ставку) и игрок 2 выбирает стратегию ПС (поднимать ставку), то игрок 1 выигрывает сумму (1+a) с вероятностью $p = m/(m+n)$ и с вероятностью (1- p) потеряет сумму (1 – a). Таким образом, математическое ожидание выигрыша игрока 1 равно $(2p – 1)(1+a)$. Таблицу математических ожиданий выигрыша для 1 можно представить в матричной форме (табл. 4.1.1)

Таблица 4.1.1

Математическое ожидание выигрыша игры в покер игрока 1

Стратегия игрока 1	Стратегия игрока 2	
	Поднимать ставку	Пасовать
Всегда поднимать	$(2p – 1)(1+a)$.	1

ставку (ВП)		
Поднимать ставку, имея картинку (ПК)	2р-1 + ар	2р - 1
Поднимать ставку, не имея картинки (ПН)	(2р – 1) а + р - 1.	1
Пасовать (П)	2р - 1	2р - 1

Очевидно, что математические ожидания выигрышей игрока 2 те же, что и у игрока 1, только со знаком «минус». Это нормальная форма игры.

Лекция 4.1.2 Описание игр в нормальной форме.

Как уже рассмотрено в предшествующей лекции (пример три в покер), имея представление игры в развернутой форме, мы можем перейти к описанию игры в нормальной форме. Нормальной формой игры называется функция, которая задает вектор ожидания выигрыша для каждой из стратегий каждого игрока. Нормальная форма игры двух лиц с конечным числом стратегии состоит из двух платежных матриц, элементы которых показывают выигрыш каждого из игроков при любой из возможных пар стратегий.

Пусть игрок 1 имеет $m(i = 1_i\overline{m})$ стратегий $S^1 = \{S_1^{\ 1}, S_2^{\ 1} .. S_m^{\ 1}\}$, а игрок 2 имеет $n(j = 1_i\overline{n})$ стратегий $S^2 = \{S_1^{\ 2}, S_2^{\ 2} .. S_n^{\ 2}\}$, тогда платежные матрицы игрока 1 (Π^1) и игрока 2 (Π^2) имеют вид (табл.4.1.2(1) и (табл.4.1.2(2).

$$\Pi^1 = \left(\Pi \frac{1}{\mathrm{ij}}\right) \begin{pmatrix} \Pi^1_{11} & \Pi^1_{12} & \Pi^1_{1j} & \Pi^1_{1n} \\ \Pi^1_{21} & \Pi^1_{22} & \Pi^1_{2j} & \Pi^1_{2n} \\ \Pi^1_{i1} & \Pi^1_{i2} & \Pi^1_{ij} & \Pi^1_{in} \\ \Pi^1_{m1} & \Pi^1_{m1} & \Pi^1_{mj} & \Pi^1_{mn} \end{pmatrix} \text{, где}$$

$\Pi \frac{1}{ij}$ - выигрыш первого игрока, если он использует i- ую стратегию, а игрок 2 использует j –ую стратегию.

$$\Pi^2 = \left(\Pi \frac{2}{\mathrm{ij}}\right) \begin{pmatrix} \Pi^2_{11} & \Pi^2_{12} & \Pi^2_{1j} & \Pi^2_{1n} \\ \Pi^2_{22} & \Pi^2_{22} & \Pi^2_{2j} & \Pi^2_{2n} \\ \Pi^2_{i1} & \Pi^2_{i2} & \Pi^2_{ij} & \Pi^2_{in} \\ \Pi^2_{m1} & \Pi^2_{m2} & \Pi^2_{mj} & \Pi^2_{mn} \end{pmatrix} \text{, где}$$

$\Pi \frac{2}{ij}$ - выигрыш игрока 2, если он использует j-устойчивую стратегию, а игрок 1 использует i-ую стратегию.

Объединенная матрица игры, элементы которой являются пара чисел, первое из которых определяет выигрыш игрока 1, второе выигрыш игрока 2 представлены в табл. 4.1.2 (3)

Таблица 4.1.2 (3)

Платежная объединенная матрица для игры двух лиц

Выбор стратегии игроком 1	Выбор стратегии игроком 2			
	S_1^2	$S_2^2 \dots$	$S_j^2 \dots$	S_n^2
S_1^1	$(П^1_{11}, П^2_{11})$	$(П^1_{12}, П^2_{12}) \dots$	$(П^1_{1j}, П^2_{1j}) \dots$	$(П^1_{1n}, П^2_{1n}$
S_2^1	$(П^1_{21}, П^2_{21})$	$(П^1_{22}, П^2_{22}) \dots$	$(П^1_{2j}, П^2_{2j}) \dots$	$(П^1_{2n}, П^2_{2}$
…	……..			
S_i^1	$(П^1_{i1}, П^2_{i})$	$(П^1_{i2}, П^2_{i2}) \dots$	$(П^1_{1j}, П^2_{1j}) \dots$	$(П^1_{1n}, П^2_{1n}$
…	……...			
S_m^1	$(П^1_{m1}, П^2_{m1})$	$(П^1_{m2} П^2_{m2}) \dots$	$(П^1_{mj}, П^2_{mj})$	$(П^1_{mn}\ П^2_{m}$

В матрице игры табл. 4.1.2(3) i- стратегия игрока S_i есть соответствующая i- строка платежной матрицы, а S_j–стратегия игрока 2-j столбец платежной матрицы. Пара чисел $(П^1_{ij}\ П^2_{ij})$ на пересечение строки i и столбца j показывает величину выигрыша каждого из них.

Рассмотрим пример игры в «чет – нечет», где каждый из двух игроков имеют по две стратегии, именуемая «чет» и «нечет». Если игроки выбирают одинаковые стратегии, т. е. в случаях, если оба говорят «чет» или оба говорят «нечет».

Игрок 1 выигрывает 1 рубль, игрок 2 проигрывает 1 рубль. В случае, когда игроки выбирают различные стратегии: игрок 1 проигрывает 1 рубль, а игрок 2 соответственно 1 рубль выигрывает.

Платежная матрица к– выигрышей для обоих игроков, которая дает представление о всей игре имеет вид (табл.4.1.2(4))

Таблица 4.1.2 (4)

Платежная матрица игры «чет» - «нечет»

Стратегия игрока 2	Стратегия игрока 2	
	Чет	Нечет
Чет	(1, -1)	(-1, 1)
Нечет	(1, -1)	(-1, 1)

В каждой клетке матрицы указаны: слева – выигрыш игрок 1, справа - выигрыш игрок 2

Игры в нормальной форме с тремя и более игроками требуют построения более сложной нормализованной формы, чем игры с двумя игроками.

Рассмотрим следующую задачу о рекламе, с которой могут столкнуться три предприятия. Предположим, что каждое из них выпускает по одному виду некоторой однотипной продукции (например. Открытки), идущей под девизом: «лучший подарок к празднику». Среди способов рекламы этой продукции фирменные магазины наших предприятий могут воспользоваться, скажем, рекламой по городскому телевидению, либо вечером накануне, последнего перед праздником рабочего дня магазина, либо утром в самый этот день. Далее будем касаться лишь эффективности этих видов рекламы и при том применительно к контингенту тех покупателей, которые смотрят вечерние телепередачи в среднем вдвое чаще, чем утренние, и. кроме того, понимают слова «лучший подарок» настолько буквально, что, услышав одновременно о двух «лучших подарках», не склонятся к покупке одного из них.

Таким образом, с одной стороны, каждое предприятие заинтересовано в вечерней рекламе (ибо она по сравнению с утренней завербует большее число покупателей), а с другой, – в том, чтобы участвовать одновременно с другими предприятиями в какой-либо из передач (ибо тогда оба одновременно участвующие в рекламе предприятия только дезориентируют покупателей, и реклама останется без ответа). Обратим внимание на то, что здесь для каждого предприятия в отдельности имеется объективно наилучший способ действий: участвовать в вечерней передаче. Однако как только этот способ действий избирает более чем одно предприятие, он уже становится бесполезным.

Подчеркнем то, что не предусмотрено в условиях задачи, не должно использоваться и в ее решении. В частности, условия задачи не предусматривают какой-либо договоренности с рекламным отделом телевидения и получения от него информации об уже поступивших заказах на рекламу. Равным образом мы имеем в виду вполне определенных покупателей выпускаемого товара.

Формализуем поставленную задачу, представив ее как игру в нормальной форме.

Участниками игры (игроками) являются здесь предприятия, которые мы для удобства обозначим через 1,2 и 3. Напротив, ни покупатели, ни телевидение, хотя их интересы прямо или косвенно и отражены в условиях данной задачи, действуют в условиях задачи вполне определенным образом и поэтому игроками считаться не могут.

Каждый игрок имеет здесь две стратегии: участвовать в утренней рекламной передаче (стратегия 1) и в вечерней (стратегия 2). Поэтому в игре имеется 2x2x2 – 8 ситуаций.

Ясно, что ситуациям соответствуют вектора-тройки числе, состоящие из нулей и единиц. Как известно, множество всех векторов-троек естественно вложить в «единичный» куб. Тогда, рассмотренные ситуации соответствуют вершинам куба (рис. 4.1.2) для этого перейдем к выигрышам игроков. В стратегиях, описываемых, как говорилось, тройками из нулей и единиц, только тот игрок получит какой-либо положительный выигрыш, чья компонента тройки отличается от всех остальных. При этом если его компонента оказывается равной нулю, то он получает 2, а если единице, то 1. Так:

1 в ситуации (1, 0,0)

игрок 1 получит

2 в ситуации (0,1,1)

Выигрыши игроков в каждой ситуации также составляют тройки чисел. Эти тройки выигрышей для 8-ми в ситуации определяют координаты соответствующих вершин куба (рис. 4.1.2)

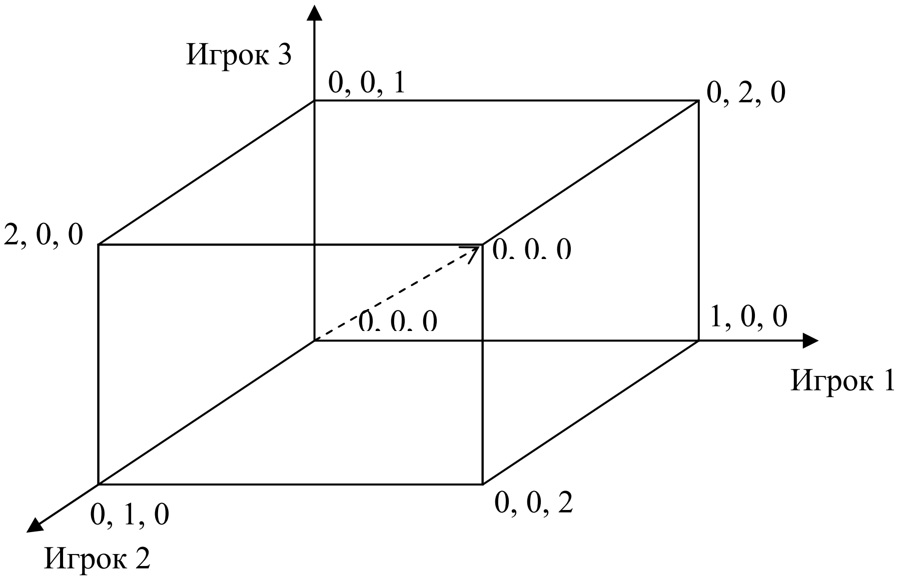


Рис. 4.1.2 Распределение выигрыша игроков в задаче о рекламе

Одно из основных понятий теорий игр – равновесные стратегии. Результатом каждого хода игры является пара стратегий $((S_i^1,\ S_j^2)$, выбранных игроками. Свой выбор стратегии игроки делают с целью максимизировать свой индивидуальный выигрыш (платеж).

Некоторый набор стратегий $\left\{S_i^1,\ S_j^2\right\}$- для игры двух лиц либо $\left\{S_i^1, S_j^2 - \ldots S_h^k\right\}$ - для игры с К участниками является равновесным в том лишь случае, когда ни один из игроков не может увеличить свой выигрыш, изменяя стратегию в одностороннем порядке.

В следующих лекциях мы рассмотрим более подробно проблемы поиска равновесных стратегий.

Тема 4.2 Выбор стратегий игроком: подходы и оценка выигрыша

Лекция 4.2.1 Статистические критерии выбора стратегии игроками. Игра с природой

Сейчас остановимся на вопросе каким образом игроки выбирают ту или иную стратегию, ориентируясь на максимизацию выигрыша (платежа). Данную задачу выбора можно проиллюстрировать на примере игры с «природой». Выбор стратегии игроком при игре с «природой» опишем платежной матрицей $\Pi = \{\Pi_{ij}\}$, где

i=$\overline{1,m}$ - индекс стратегии игрока S_i;

j=$\overline{1,n}$ - индекс возможного состояния «природы»;

Π_{ij} – выигрыш игрока при выбранной I-ой стратегии при j-ом состоянии «природы» (см. табл.4.2.1 (5))

Таблица 4.2.1(5)

Платежная матрица игры с природой

Стратегия игрока	Состояние природы					
	A_{11}	A_2	…..	A_j	…..	A_n
S_1	Π_{11}	Π_{12}	…..	Π_{ij}	…..	Π_{1n}
S_2	Π_{21}	Π_{22}	…..	Π_{2j}	…..	Π_{2n}
………	…..	…..	…..	…..	…..	…..
S_i	Π_{i1}	Π_{i2}	…..	Π_{ij}	…..	Π_{in}
……..	…..	…..	…..	…..	…..	…..
S_m	Π_{m1}	Π_{m2}	…..	Π_{mj}	…..	Π_{mn}

Выбор стратегии игроком будет осуществляться в условиях неопределенности реализации того или иного состояния «природы». А одним из инструментов выбора решения (стратегии) является использование статистических критериев выбора. Мы рассмотрим следующие критерии:

2) критерий Лапласа; 1) критерий Вальда; 4) критерий Гурвица; 5) критерий Сэвиджа; 3) критерий крайнего оптимизма.

Не существует общих правил применимости того или иного критерия в играх с «природой», выбор конкретного критерия игроком зачастую связан с субъективной оценкой ситуации, когда осуществляется процедура выбора субъективной ментальностью лица, принимающего решение.

Логика применения каждого из статистических критериев основывается на 2-х шаговой процедуре:

- на первом шаге для каждой S_i стратегии оценивается наиболее вероятный выигрыш;
- на втором шаге выбирается стратегия, которая обеспечивает наибольший выигрыш.

Опираясь на платежную матрицу (табл.4.1.2 (5)) рассмотрим применение четырех однозначных статистических критерия.

Критерий Вальда. (критерий глубокого пессимизма).

Этот критерий является наиболее осторожным, поскольку он основывается на выборе наилучшей стратегии из наихудших возможностей состояний природы A_j. Тогда оценка лучшего гарантированного выигрыша равна:

$\Pi^* = \max_i \min_j \Pi_{ij}$, ($S_i^*$ – выбранная стратегия, соответствующая Π^*)

Критерий Лапласа.

Этот критерий ориентируется на среднюю оценку ожидаемого выигрыша для каждой стратегии при гипотезе, при состоянии природы A_j. Тогда оценка выигрыша равна:

$$\overline{\Pi} = \max_i \left(\frac{1}{n} \sum_{j=1}^{n} \Pi ij \right),$$ ($\overline{S_i}$, выбранная стратегия, соответствующая $\overline{\Pi}$).

Критерий крайнего оптимизма.

Этот критерий основывается на выборе наилучшей стратегии при наиболее благоприятных состояниях «природы». Тогда оценка выигрыша равна:

$\tilde{\Pi} = \max_i \max_j \Pi_{ij}$, ($\tilde{S}_i$ – выбранная стратегия, соответствующая $\tilde{\Pi}$)

Критерий Гурвица:

Этот критерий строится, как охват ряда различных подходов к принятию решений: от крайне оптимистического до наиболее пессимистического. Иными словами для каждой S_i стратегии ожидания выигрыша есть выпуклая комбинация худшего и лучшего выигрыша. Тогда оценка выигрыша в игре равна:

$\Pi' = \max\limits_{i}\left\{\alpha \max\limits_{j} \Pi ij + (1-\alpha)\min\limits_{j} \Pi ij\right\}$, где $0 \le \alpha \le 1$., ($\tilde{S}_i$ – выбранная стратегия, соответствующая $\tilde{\Pi}$)

Критерий Гурвица есть некий баланс между крайним оптимизмом и крайним пессимизмом, установленный путем взвешивания обоих подходов в поведении игрока с соответствующими весами α и $(1-\alpha)$. При $\alpha=1$ мы имеем критерий крайнего оптимизма; при $\alpha=0$ имеем критерий Вальда.

Критерий Сэвиджа.

Данный критерий ориентирует игрока на выбор стратегии не с позиций оценки выигрыша, а с позиций оценки возможных потерь в результате выбора.

Для оценки потерь игрока при выборе не лучшей стратегии S_i для состояния «природы» A_j строится матрица сожаления (потерь) $C = (Cij)$;

$Cij = \max\limits_{i} \Pi ij - \Pi ij$, для каждого $j = \overline{1n}$.

Выбор стратегии Si ориентируется на минимизацию потерь в худших условиях состояния природы Aj, тогда оценка потерь в игре равна

$\overline{C} = \min\limits_{i} \max\limits_{j} Cij$, ($\overline{S_i}$ выбранная стратегия, соответствующая $\overline{C}$)

Рассмотрим пример пусть предприятие имеет 4 стратегии выпуска продукции. Точное число клиентов неизвестно, но возможно оценить 4 варианта объема спроса на продукцию. Платежная матрица игры описывает объем получаемой прибыли предприятия от реализации продукции по каждой из стратегий в зависимости от вариантов прогноза спроса (табл. 4.2.2 (6)).

Таблица 4.2.2

Платежная матрица прибыли предприятия

Стратегия предприятия	Состояние спроса							
	A_1	A_2	A_3	A_4	$\max\limits_{j}$ Πij	$\min\limits_{j}$ Πij	$1/\sum n \sum \Pi ij$ 1/n	Средняя оценка, при $\alpha=0{,}3$
S_1	50	100	180	250	250	50	145	110

S_2	80	70	80	230	230	70	115	118
S_3	210	180	120	210	210	120	180	147
S_4	300	220	190	150	300	150	215	205
$\max\limits_i$ Пij	300	220	190	250				

Решение приведенного примера включает несколько этапов.

Этап 1. Для каждой стратегии S_1, S_2, S_3, S_4 рассчитываем оценки получаемой прибыли и заносим значения в дополнительные столбцы матрицы табл. 4.2.2 (6):

а) максимальная оценка прибыли для каждой стратегии $= \max\limits_j \text{Пij}$

б) минимальная оценка для каждой стратегии $= \min\limits_j \text{Пij}$

в) средняя оценка прибыли для каждой стратегии $= \frac{1}{4}\sum\limits_{j=1}^{4} \text{Пij}$.

г) средняя оценка прибыли по Гурвичу при ($\alpha = 0{,}3$) $\alpha \cdot \max_j \Pi_{ij} + (1-\alpha)\min_j \Pi_{ij}$

Этап 2. Применяем критерии для оценки выигрыша.

а) по критерию Вальда объем, получаемой прибыли $\Pi^* = \max\limits_j \{50;70;120;150\} = 150$, что соответствует стратегии S_4;

б) по критерию Лапласа объем, получаемой прибыли $\overline{\overline{\Pi}} = \max\limits_j \{145;115;180;215\} = 215$, что соответствует стратегии S_4;

в) по критерию крайнего оптимизма объем получаемой прибыли $\tilde{\Pi} = \max\limits_j \{250;230;210;300\} = 300$, что соответствует стратегии S_4;

г) по критерию Гурвица (при $\alpha = 0{,}3$) объем, получаемой прибыли $\Pi' = \max\limits_j \{110{,}118{,}147{,}205\}$, что соответствует стратегии S_4.

Результат этапа 2 показывает, что необходимо выбирать только стратегию S_4, а значения критериев дает различные оценки, получаемой прибыли.

Этап 3. Для применения критерия выбора Сэвиджа строим матрицу ущерба С будет иметь вид:

$$C = \begin{pmatrix} 250 & 120 & 10 & 0 \\ 220 & 150 & 110 & 20 \\ 90 & 40 & 70 & 40 \\ 0 & 0 & 0 & 100 \end{pmatrix},$$

а оценка потерь прибыли по критерию Сэвиджа будет равна:

$\overline{C} = \min\{250,220,90,100\}$=90, что соответствует стратегии S_3.

Как мы видим выбор стратегии по критерию Сэвиджа (S_3) не совпадает с выбором по другим критериям (S_4). Окончательный выбор из S_3 и S_4 остается за игроком.

В антагонистических играх, когда все участники ведут себя рационально, каждый игрок, осуществляя выбор стратегии, как правило, ориентируется на исследование критерия Вальда, этот критерия гарантирует ему получение лучшего выигрыша из худших исходов, которые ему создает рациональный второй игрок.

Тема 4.3 Антагонистические игры. Матричные игры двух лиц с нулевой суммой.

Лекция 4.3.1 Антагонистические игры двух лиц с нулевой суммой в матричном виде в чистых стратегиях.

Данный тип игр достаточно хорошо и полно разработан. Если игра представлена в нормальной форме то выбор стратегий может быть описан для каждого игрока платежными матрицами:

для игрока 1: $П^1 = \{Пij^1\}, i = 1,\overline{m}; j = 1,\overline{n}$

для игрока 2: $П^2 = \{Пij^2\}, i = 1,\overline{m}; j = 1,\overline{n}$

Каждый игрок может выбирать априорно заданный вектор из своей платежной матрицы (чистую стратегию):

а) первый игрок выбирает вектор строку $П_i$ (i=$\overline{1,m}$ из матрицы $П^1$ (чистую i-ую стратегию)

б) второй игрок выбирает вектор столбец $П_j^2 (j = 1,\overline{n})$ из матрицы $П^2$ (чистую j-ую стратегию).

Нулевая сумма игры, означает, что

$$П_{ij}^1 + П_{ij}^2 = 0, \qquad (4.3.1)$$

следовательно

$$П_{ij}^1 = -П_{ij}^2, \qquad (4.3.2)$$

т.е. платежная матрица игрока 2 равна платежной матрице игрока 2 умноженной на (-1). Следовательно, в игре выбор чистых стратегий игроков можно заменить выбором стратегий по одной матрице $П$:

$$П_{ij} = П_{ij}^1 = -П_{ij}^2, \text{ где } П_{ij} = \begin{vmatrix} П_{11}.....П_{ij}.....П_{1n} \\ П_{i1}.....П_{ij}......П_{in} \\ П_{m1}....П_{mj}....П_{mn} \end{vmatrix}, \text{ где } П_{ij} = П_{ij}^{\ 1} = -П_{ij}^{\ 2}, \qquad (4.3.3)$$

При этом игрок 1 стремится выбирать строку матрицы так, чтобы максимизировать свой выигрыш (платеж), а игрок 2 – такой столбец матрицы, при котором его проигрыш минимален.

В качестве основного допущения в теории игр двух лиц с нулевой суммой принимается, что каждый игрок стремится обеспечить себе максимально возможный выигрыш при любых действиях противника, Однако наибольший гарантированный выигрыш определяется при том условии, что избранная данным игроком стратегия становится известной противнику, который затем выбирает свою оптимальную стратегию. Пусть игрок 1 считает, что какую бы строку он не выбрал, игрок 2 выберет столбец, максимизирующий выигрыш и тем самым минимизирующий выигрыш игрока 1. Тогда можно исключить из платежной матрицы все элементы, оставив в каждой строке только по одному элементу, соответствующему минимальному платежу. Оптимальная стратегия игрока 1, которая обеспечить ему наибольший из возможных выигрышей вне зависимости от стратегии противника (гарантированный выигрыш), будет состоять в выборе строки с самым высоким значением из таких минимальных платежей. Таким образом, игрок 1 выбирает i-ю стратегию, которая является решением задачи для игрока 1:

$$П^{1*} = \max_{j} \min_{i} \qquad (4.3.4)$$

$П^{1*}$ - (нижняя цена игры);

i - стратегия 1 игрока , соответствующая максимальному значению минимумов по строкам ($П^{1*}$), называется максиминной стратегией.

Игрок 2 точно так же стремится обеспечить себе наивысшую величину выигрыша (т.е. наименьшее значение платежа своему противнику) вне зависимости от стратегии, избираемой партнером. Следовательно, игрок 2 может исключить из платежной матрицы все элементы, оставив в столбце только по одному элементу, соответствующему максимальному платежу. Его оптимальной стратегией будет столбец с наименьшим значением максимального платежа. Таким образом, игрок 2 выбирает j-ю стратегию, которая является решением задачи для игрока 2:

$$П^{2*} = \min_{j} \max_{i} П_{ij} \qquad (4.3.5)$$

$П^{2*}$ - верхняя цена игры;

j^* стратегия игрока 2, соответствующая минимальному значению максимумов столбцов ($П^{2*}$), называется минимаксной стратегией.

Если игрок 1 придерживается максиминной стратегии, то его выигрыш будет не меньше максиминного значения, т.е.

$$П_{ij} \geq \max_{i} \min_{j} П_{ij} \qquad (4.3.6)$$

Если игрок 2 избирает минимаксную стратегию, то его проигрыш будет не больше минимаксного значения, т.е.

$$\Pi_{ij} \leq \min_j \max_i \Pi_{ij} \tag{4.3.7}$$

Если

$$\max_i \min_j \Pi_{ij} = \max_j \min_i \Pi_{ij} = \Pi^*, \tag{4.3.8}$$

то игроки получают свои гарантированные платежи. В этом случае их стратегии оказываются совместимыми, а платежная матрица имеет седловую точку на пересечении i^*-й строки и j^*-го столбца, т.е. элемент Π^*_{ij} является одновременно минимальным в своей строке и максимальным в своем столбце.

Платеж Π^* называется **ценой игры**. Игры, имеющие седловую точку, называются **вполне определенными играми**.

Рассмотрим пример вполне определенной игры 2-х лиц с нулевой суммой с платежной матрицей П:

$$\Pi = \begin{pmatrix} 2 & 1 & 4 \\ -1 & 0 & 6 \end{pmatrix} \tag{4.3.9}$$

В данном пример игры, в которой игрок 1 располагает двумя стратегиями. А игрок 2 – тремя. Игрок 1 рассчитывает, что если он выберет первую строку, то противник может выбрать второй столбец, так что выигрыш будет равен 1. Если же он выберет вторую строку, то противник может взять первый столбец, так что выигрыш будет равен -1. Эти два числа есть не что иное, как минимумы строк, показанные в табл. 4.2.1.

Взяв максимальное значение этих минимумов, игрок 1 останавливается на своей первой стратегии, которая обеспечивает ему выигрыш, равный 1 (или больший выигрыш, если игрок 2 выберет первый или третий столбец). Точно так же и игрок 2 рассматривает наихудшие варианты, когда противник выбирает вторую строку, в то время как у него выбран третий столбец. Эти варианты соответствуют максимальным значениям столбцов 2, 1 и 6. Взяв минимальное значение этих максимумов, игрок 2 останавливается на своей второй стратегии, при которой его проигрыш не превосходит 1. следовательно, в этой игре существуют совместимые выборы, т.е.

$$V=\max_i \min_j \Pi_{ij}=\min_j \max_i \Pi_{ij}=1 \tag{4.3.10}$$

И, следовательно, существует седловая точка матрицы,, являющаяся ценой игры (V) для игроков. Седловая точка соответствует положению равновесия, если один из игроков использует стратегию, соответствующую седловой точке, то другому выгоднее всего избрать свою стратегию, также отвечающую седловой точке. Так, если в данном примере игрок 1 использует первую стратегию, то для игрока 2 оптимальной будет вторая стратегия, а если игрок 2 применяет вторую стратегию, то для игрока 1 оптимальной будет первая стратегия. Разумно ожидать, что в игре такого типа оба партнера изберут стратегию седловой точки. Однако не все игры двух участников с нулевой суммой является вполне определенными. В общем случае

$$\max_i \min_j \Pi_{ij} \le \min_j \max_i \Pi_{ij} \tag{4.3.11}$$

Игры, в которых выполняется строгое неравенство, называются не полностью определенными играми без седловой точки.

Так примером не вполне определенной игры может служить игра двух лиц с платанной матрицей П

$$\Pi=\begin{pmatrix} 6 & -2 & 3 \\ -4 & 5 & 4 \end{pmatrix} \tag{4.3.12}$$

Если игроки следуют изложенным ранее правилам, то получим, что седловая точка отсутствует:

$$\max\min \Pi_{ij}=-2\langle 4=\max\min \Pi ij \tag{4.3.13}$$

Последовательно выбирая чистые стратегии игроки не придут к определенному консенсусу в отношении выбора цены игры. В общем случае отсутствие седловой точки в игре есть плата за то, что игроки могут использовать только ограниченное число чистых стратегий. Логическим приемом в направлении поиска приемлемого решения для каждого игрока является использование смешанных стратегий.

Лекции 4.3.2 Антагонистические игры двух лиц с нулевой суммой в смешанных стратегиях.

Как было рассмотрено в предшествующей лекции игра двух лиц с нулевой суммой не всегда имеет решение в чистых стратегиях.

Переход к смешанным стратегиям, предполагает, что игроки осуществляют выбор чистых стратегий случайным образом, Математически схему случайного выбора стратегий игроком можно представить как вероятностное распределение на множестве чистых стратегий данного игрока. Тогда **смешанная стратегия** представляет собой вероятностную комбинацию чистых стратегий, т.е. ряд чистых стратегий, взятых в случайном порядке с некоторыми вероятностями. Так m строк платежной матрицы (4.3.3) являются чистыми стратегиями игрока 1. Смешанную стратегию для игрока 1 можно указать с помощью вектора (вектора-строки) вероятностей

$$P^1=\{p_i\}(p_1^1, p_2^1, ...p_m^1), \qquad (4.3.14)$$

где p_i^1- вероятность выбрать i-ю стратегию, (i=1,2,……..,m). Например, вектор $(\frac{1}{4}, \frac{1}{2}, 0, ..., \frac{1}{4})$ соответствует смешанной стратегии, при которой игрок 1 выбирает строку 1 с вероятностью $\frac{1}{4}$ и строку 2 с вероятностью $\frac{1}{4}$, а строку m с вероятностью $\frac{1}{4}$. Так как числа, входящие в вектор p^1, является вероятностями, то они должны быть неотрицательными, а их сумма должна равняться единице, т.е.

$$\sum_{i=1}^{m} p_i^1 = 1, p_i^1 \geq 0, i = 1, 2, ..., m. \qquad (4.3.15)$$

Эти условия можно записать и в векторных обозначениях

$$p^1\mathbf{1}' = 1, p^1 \geq 0, \qquad (4.3.16)$$

где **1** = (1,1,…….,1), т.е. это вектор-строка с элементами, равными единице.

Точно так же игрок 2 выбирает вектор (вектор-столбец) вероятностей

$$P^1=\{p_i\}(p_1^2, p_2^2, ..., p_n^2)', \qquad (4.3.\ 17)$$

где p_j^2- вероятность выбрать j-ю стратегию; j=1,2,…,n, причем можно брать любой вектор p^2, в котором $p_j^2 \geq 0, \sum_j^n p_j^2 = 1$, или

$$\mathbf{1}p^2=1,\ p^2 \geq 0. \qquad (4.3.18)$$

Отметим, что чистые стратегии можно рассматривать как частный случай смешанной стратегии, когда вектор вероятностей представляет собой единичный вектор. Так, вектор-строка (0, 1, 0, ….., 0) означает, что игрок 1 выбирает вторую строку матрицы – вторую чистую стратегию, поскольку вероятность выбрать ее равна единице.

Переход к описанию игры в смешанных стратегиях фактически означает переход от выбора игроком из ограниченного числа чистых стратегий и выбору из бесконечного числа смешанных стратегий.

Основной теоремой в теории игр двух лиц с нулевой суммой является теорема о минимаксе, доказанной Дж. Нейманом, согласно которой любая конечная игра имеет решение, если допускается использование смешанных стратегий[2]. Вполне определенная игра имеет решение в области чистых стратегий, причем это решение может быть не единственным. Не полностью определенная игра имеет решение возможно не единственное, при котором хотя бы один из игроков применят случайное комбинирование (смешивание) стратегий.

Так как при выборе стратегий используется вероятностный подход, то выигрыш игрока 1 (проигрыш игрока 2) уже является не некоторым отдельным элементом платежной матрицы, а взвешенным средним значением элементов, причем в качестве весов этой матрицы берутся вероятности. Так, ожидаемый выигрыш (математическое ожидание выигрыша) игрока 1 в предположении, что он использует вектор вероятностей $p^1=(p_1^1, p_2^1, \ldots\ldots p_m^1)$, а игрок 2 применяет свою j-ю стратегию, равен

$$p_1^1\Pi_{1j} + p_2^1\Pi_{2j} + \ldots + p_m^1\Pi_{mj} = p^1\Pi e'_j, \qquad (4.3.19)$$

$$J=1, 2, \ldots, n,$$

[2] Для бесконечных игр теорема о минимаксе не выполняется. Например, такие игры, как игра, в которой два игрока записывают какие-либо числа и написавший большее число выигрывает некоторую сумму у противника, не имеют решений ни в чистых, ни в смешанных стратегиях.

где $\Pi=(\Pi_{ij})$ – платежная матрица, а e_j есть j-й единичный вектор. Игрок 1, стремящийся достичь наибольшего из гарантированных ожидаемых выигрышей, выбирает вектор вероятностей так, чтобы получить максимум минимальных значений ожидаемых значений ожидаемых выигрышей. Если обозначить

$$\Pi^1(p^1)=\min_j p^1\Pi e'_j, \tag{4.3.20}$$

то игрок 1 действует таким образом, чтобы

$$\max_{p^1}\Pi^1(p^1)=\max_{p^1}\min_j p^1\Pi e'_j. \tag{4.3.21}$$

Ожидаемый выигрыш (математическое ожидание выигрыша) игрока 2 в предложении, что он использует вектор вероятностей $p^2=(p_1^2, p_2^2, \ldots\ldots, p_n^2)'$, а игрок 1 выбрал i-ю стратегию, равен

$$_{i1}\,p_1^2+\Pi_{i2}\,p_2^2+\ldots\ldots+\Pi_{in}\,p_n^2=e_i\,\Pi p^2,\ i=1.\ 2.\ \ldots.,\ m \tag{4.3.22}$$

Цель игрока 2 – достичь минимума максимальных значений ожидаемых платежей

$$\Pi^2(p^2)=\max_i e_i\Pi p^2 \tag{4.3. 23}$$

Следовательно, игрок 2 выбирает p^2 так, чтобы

$$\min_{p^2}\Pi^2(p^2)=\min_{p^2}\max_i e_i\Pi p^2 \tag{4.3.24}$$

Согласно теореме о минимаксе, существуют решения для (4.3.24) и (4.3.22) – p^{1*} и p^{2*}. Если положить, что цена игры равна:

$$V=p^{1*}\Pi p^{2*}=\sum_{i=1}^{m}\sum_{j=1}^{n}p_i^{1*}\Pi_{ij}p_j^{2*}, \tag{4.3.25}$$

то при любых векторах вероятностей p^1 и p^2 выполняются соотношения

$$p^1\Pi p^{2*}\le V\le p^{1*}\Pi p^2 \tag{4.3.26}$$

и

$$\max_{p^1} p^1 \Pi p^{2*} = V = \min_{p^2} p^{1*} \Pi p^2 \qquad (4.3.27)$$

Следовательно, V-цена игры – является одновременно максимальным ожидаемым выигрышем игрока 1 и минимальным ожидаемым проигрышем игрока 2. Таким образом, теорема о минимаксе утверждает, что существует хотя бы одна пара смешанных стратегий (p^1, $p^{2)}$, при которых максимальное и минимаксное значение ожидаемых платежей совпадает, т.е.

$$V = \max_{p^1} \min_{p^2} p^1 \Pi p^2 = \max_{p^2} \min_{p^1} p^1 \Pi p^2, \qquad (4.3.28)$$

Так что любая конечная игра имеет седловую точку в пространстве векторов вероятностей, Цена игры V единственна, а соответствующие ей векторы вероятностей оптимальных смешанных стратегий, согласно (4.3.25), могут быть и не единственными. Однако, если существует более чем одна пара оптимальных смешанных стратегий, то все эти пары стратегий образуют замкнутое выпуклое многогранное множество, причем все входящие в это множество пары доставляют игре одну и ту же цену.

Лекция 4.3.3. Связь теории игр двух лиц с нулевой суммой и линейного программирования.

На связь между линейным программированием и теорией игр впервые указал в 1947г. Дж. Нейман, а несколько позже Дж. Данциг показал, что любой конечной игре с нулевой суммой соответствует некоторая модель линейного программирования. И наоборот, при доказательстве теоремы о минимаксе используется теорема двойственности линейного программирования. Общий метод нахождения смешанных оптимальных стратегий для обоих игроков заключается в построении соответствующей игры с заданной платежной матрицей П = $\{\Pi_{ij}\}$ ($i=1,\overline{m};\ j=1,\overline{n}$) задачи линейного программирования и ее решения. Решение прямой и двойственной к ней задачи дает оптимальные стратегии игроков. Значение целевой функции определяет цену игры.

То, что игрок 1 рассматривает минимальные ожидаемые выигрыши, мы записали ранее с помощью выражения (4.3.20). Это условие можно представить и в форме линейных неравенств:

$$p^1 \Pi e_j' = \sum_{i=1}^{m} p_i^1 \Pi_{ij} \geq \Pi^1(p^1) \qquad (4.3.29)$$

$$j = 1, 2, \ldots, n.$$

Если, как и прежде, обозначить символом **1** вектор-вектор строку из единиц, то эти неравенства можно записать в эквивалентной форме

$$p^1\Pi - \Pi^1(p^1)1 \geq 0\,. \qquad (4.3.30)$$

Тогда задача (4.3.21), стоящая перед игроком 1, может быть сформулирована как следующая задача линейного программирования:

найти $\max\limits_{p^1} \Pi^1(p^1)$

при условии, что $p^1\Pi - \Pi^1(p^1)\,\mathbf{1} \geq 0$ (4.3.31)

$$p^1\mathbf{1}' = 1$$

$$p^1 \geq 0\,.$$

Задачу игрока 2, который минимизирует максимальные платежи, также можно сформулировать как задачу линейного программирования:

найти $\min\limits_{p^2} \Pi^2(p^2)$

при условии, что $\Pi p^2 - 1' \Pi^2(p^2) \le 0$ (4.3.32)

$$1p^2 = 1$$

$$p^2 \ge 0.$$

Эти две задачи являются двойственными задачами линейного программирования, что легко видеть из таблицы 4.3.1.

Табл. 4.3.1.

Двойственные задачи определения оптимальных смешанных стратегий, представленные в форме таблицы для задач линейного программирования.

	p_1^2	p_2^2 …	p_n^2	$-\Pi^2(p^2)$	
p_1^1	Π_{11}	Π_{12} …	Π_{1n}	1	≤ 0
p_2^1 . . .	Π_{21}	Π_{22} …	Π_{2n}	1 . . .	≤ 0
p_m^1	Π_{m1}	Π_{m2} …	Π_{mn}	1	≤ 0
$\Pi^1(p^1)$	1	1 …	1	1	$= 1 - \Pi^2(p^2)$. Найти max [1-$\Pi^2(p^2)$], т.е. найти min $\Pi^2(p^2)$
	≥ 0	≥ 0 …	≥ 0	$= 1 - \Pi^1(p^1)$. Найти min$(1 - \Pi^1(p^1)$, т.е. найти max $\Pi^1(p^1)$	

Представленная таблица является краткой формой записи обеих задач,

если положить, что
$$p_m^1 = 1 - \sum_{i=1}^{m-1} p_i^1$$
$$p_n^2 = 1 - \sum_{j=1}^{n-1} p_j^2 \qquad (4.3.33)$$

для того, чтобы сумма вероятностей равнялась единице. Так как для каждой из задач существует допустимый вектор (например, единичные векторы), то по теореме существования теории линейного программирования эти задачи имеют решения p^{1*} и p^{2*} соответственно. Но по теореме двойственности

$$\Pi^1(p^{1*}) = \max_{p^1} \Pi^1(p^1) = V = \min_{p^2} \Pi^2(p^2) = \Pi^2(p^{2*}), \qquad (4.3.34)$$

где V – цена игры.

Таким образом, теорема о минимаксе теории игр вытекает из теоремы двойственности теории линейного программирования. Кроме того, из теоремы о дополняющей нежесткости следует, что

$$\begin{aligned} &\text{либо} \sum_{i=1}^{m} p_i^{1^*} \Pi_{ij} = V, \text{ либо } p_j^{2^*} = 0,\ j = 1, 2, \dots, n. \\ &\text{либо} \sum_{j=1}^{n} \Pi_{ij} p_j^{2^*} = V, \text{ либо } p_i^{j^*} = 0,\ i = 1, 2, \dots, m. \end{aligned} \qquad (4.3.35)$$

Часто соотношения (4.3.35) называют сильной теоремой о минимаксе. Эта теорема утверждает, в частности, что если при некоторой чистой стратегии игрока 2 ожидаемый выигрыш игрока 1 превосходит цену игры, то игрок 2 избирает эту стратегию с нулевой вероятностью.

В общем случае следует ожидать, что в игре двух участников с нулевой суммой оба игрока применяют свои оптимальные смешанные стратегии. В частном случае вполне определенной игры оптимальной смешанной стратегией будет такая стратегия, в которой чистой стратегии, соответствующей седловой точке, приписана вероятность, равная единице, т.е. векторы оптимальных стратегий единичные. Вообще число нулевых элементов в векторе оптимальной смешанной стратегии не должно превышать минимальное количество чистых стратегий, имеющихся в распоряжении каждого игрока.

Для поиска оптимальных смешанных стратегий игроков как правило используется симплекс-метод, предварительно для упрощения применяя методы уменьшения размерности задачи.

Зачастую для уменьшения размерности игры используют метод доминирования строк и столбцов платежной матрицы $\Pi=\{\Pi_{ij}\}$ $(i=1,\overline{m};\ j=1,\overline{n})$.

Как правило говорят, что k-ая строка платежной матрицы Π доминирует i-ую строку (т.е. одна чистая стратегия игрока 1 доминирует другую), если

$$\Pi_{ij} \le \Pi_{kj} \text{ при всех } j, \qquad (4.3.36)$$
$$\Pi_{ij} < \Pi_{kj} \text{ по крайней мере при одном } j.$$

Аналогично, l-ый столбец доминирует j-ый столбец (т.е. одна чистая стратегия игрока 2 доминирует другую), если

$$\Pi_{il} \le \Pi_{ij} \text{ при всех } I, \qquad (4.3.37)$$
$$\Pi_{il} < \Pi_{ij} \text{ по крайней мере при одном } i.$$

При этом k-ая строка есть доминирующая стратегия, а i-ая строка доминируемая стратегия игрока 1. для второго игрока l-ый столбец есть доминирующая стратегия, а j-ый столбец – доминируемая стратегия.

Смысл этого состоит в том, что доминирующая стратегия никогда не хуже, а в некоторых случаях лучше, чем доминируемая стратегия. Из этого следует, что игрок может исключить из рассмотрения доминируемую стратегию, так как вероятность ее использования будет равна нулю. То есть доминируемые стратегии могут быть исключены из платежной матрицы, что позволит уменьшить ее размерность.

Например, рассмотрим игру со следующей платежной матрицей П:

$$\Pi = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 1 & 3 \\ 3 & 6 & 4 \end{pmatrix}. \qquad (4.3.38)$$

Третья строка доминирует вторую, исключение второй строки приводит к матрице $\overline{\Pi}$:

$$\overline{\Pi} = \begin{pmatrix} 5 & 2 & 1 \\ 3 & 6 & 4 \end{pmatrix}. \qquad (4.3.39)$$

Третий столбец матрицы $\overline{\Pi}$ доминирует второй, исключение которого дает матрицу $\widetilde{\Pi}$:

$$\widetilde{\Pi} = \begin{pmatrix} 5 & 1 \\ 3 & 4 \end{pmatrix}. \qquad (4.3.40)$$

Для решения игры с матрицей $\widetilde{\Pi}$ можно использовать более простые методы решения, например, графические.

Лекция 4.3.4. Графические методы решения игр двух лиц с нулевой суммой

Для игр, в которых хотя бы один игрок имеет только две чистые стратегии, используется графический метод решения для поиска оптимальных смешанных стратегий.

Так для игры, заданной (4.3.12), с платежной матрицей $\Pi = \begin{pmatrix} 6 & -2 & 3 \\ -4 & 5 & 4 \end{pmatrix}$ графическое решение относительно игрока 1 приведено на рис. 4.3.4.

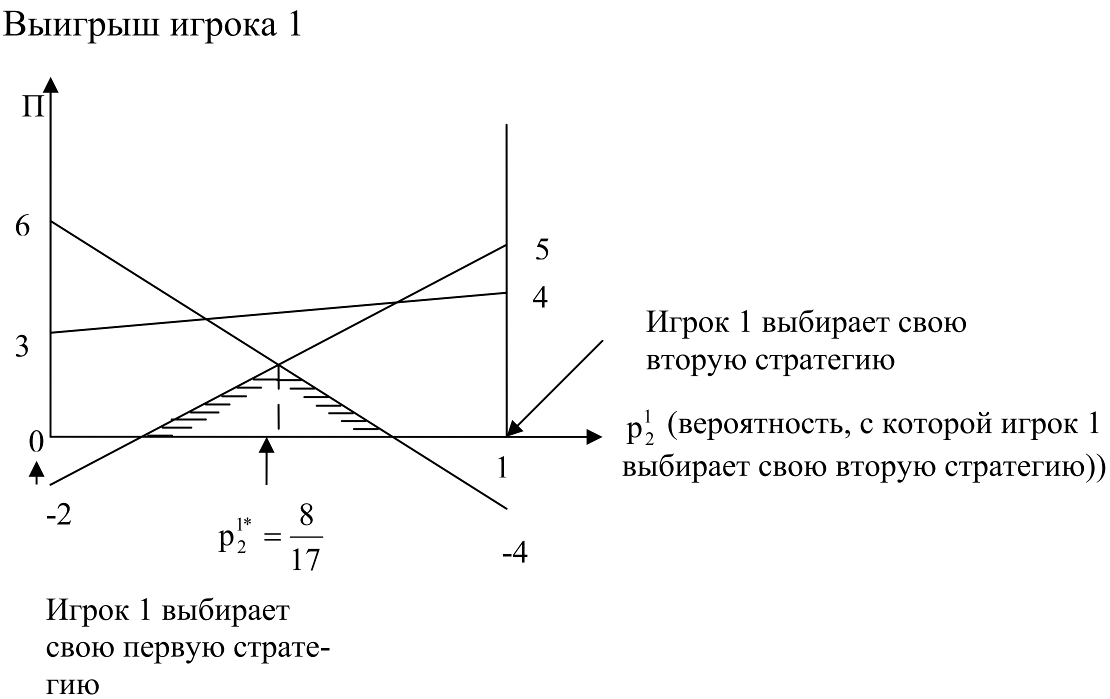


Рис. 4.3.4. Графическое решение игры для игрока 1

Здесь по оси абсцисс откладывается значение p_2^1 - вероятность того, что игрок 1 избирает вторую стратегию (вторую строку матрицы). Так как $p_2^1 = 1 - p_1^1$, то точки 0 и 1 соответствуют двум чистым стратегиям – выбору соответственно стратегии (строк) 1 и 2. По оси ординат откладывается величина выигрыша игрока 1, а каждая из проведенных линий соответствует некоторой чистой стратегии, выбранной игроком 2. Например, если игрок 2 выбирает первый столбец, то выигрыш игрока 1 равен либо 6 при выборе им первой строки ($p_2^1 = 0$), либо – 4, если он избирает вторую строку ($p_2^1 = 1$). Величина выигрыша, равная 6, отложена по вертикальной оси в левой части чертежа, а величина выигрыша, равная – 4, представлена отрезком, отложенным по вертикали в правой части. Выигрыши, которые можно получить при любой из смешанных стратегий, соответствуют точкам прямой, соединяющей концы этих отрезков. Поскольку игрок 1 рассматривает прежде всего наихудшие для себя варианты, то единственным геометрическим местом точек, которое соответствует осуществимым стратегиям, являются два отрезка, сходящиеся в форме перевернутой буквы V. Это геометрическое место точек показано линией со штрихом. Точки этих отрезков отвечают наименьшим значениям ожидаемых выигрышей игрока 1, которые он может получить при той или иной вероятности выбрать вторую строку. Максимальному выигрышу соответствует $p_2^{1*} = \frac{8}{17}$. Это можно показать либо геометрически, либо решая алгебраическое уравнение

$$-2(1-p_2^1)+5p_2^1 = 6(1-p_2^1)-4p_2^1. \tag{4.3.41}$$

Итак, игрок 1 должен выбирать свою первую стратегию с вероятностью $\frac{9}{17}$, а вторую – с вероятностью $\frac{8}{17}$ ($p^{1*}\left(\frac{9}{17},\frac{8}{17}\right)$).Цена игры равна

$$V = -2\left(\frac{9}{17}\right)+5\left(\frac{8}{17}\right)=6\left(\frac{9}{17}\right)-4\left(\frac{8}{17}\right)=\frac{22}{17}. \tag{4.3.42}$$

Графическое решение данной игры относительно игрока 2 дано на рис. 4.3.5.

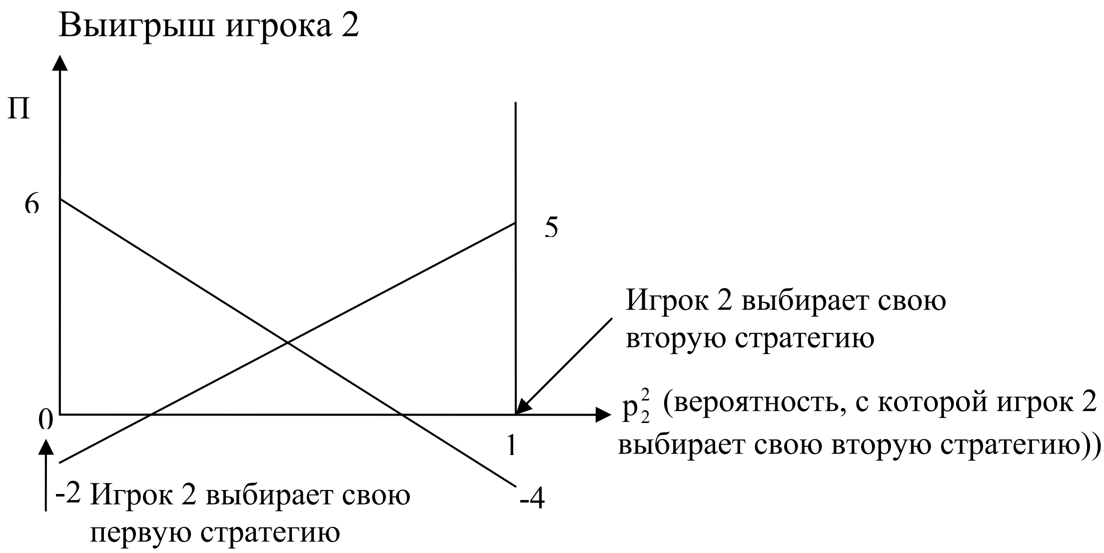


Рис. 4.3.5 Графическое решение игры для игрока 2

Исходя из решения игры относительно игрока 1, мы можем исключить третью стратегию второго игрока, тогда платежная матрица будет иметь

$$\overline{\Pi} = \begin{pmatrix} 6 & -2 \\ -4 & 5 \end{pmatrix}. \qquad (4.3.43)$$

Решая алгебраическое уравнение

$$-2\,p_2^2 + 6\,(1 - p_2^2) = 5p_2^2 - 4\,(1 - p_2^2) \qquad (4.3.44)$$

$$-17p_2^2 = -10$$

$$p_2^{2*} = \frac{10}{17}; p_2^{2*} = \frac{7}{17}$$

Цена игры равна

$$V = 5\frac{10}{17} - 4\frac{7}{17} = \frac{22}{17}, \qquad (4.3.45)$$

таким образом оптимальное решение игры (p^{1*}, p^{2*}):

$$p^{1*} = \left(\frac{9}{17}; \frac{8}{17}\right); p^{2*} = \left(\frac{10}{17}; \frac{7}{17}\right) \text{ при цене игры } \frac{22}{17}. \qquad (4.3.46)$$

В общем случае решение игр с размерностью платежных матриц (2 x n) либо (m x 2) сводится к решению игры размерности (2 x 2) следующим образом:

1) Строится графическое решение игры для игрока с двумя стратегиями. Выделяется нижняя граница игры (в случае размерности игры (2 x n). Либо выделяется верхняя граница для игры (в случае размерности (m x 2)).

2) Устанавливается наибольшая ее ордината, которая равна цене игры V (в случае размерности (2 x n)) либо наименьшая ордината (в случае игры (m x 2)).

3) Находится пара стратегий, пересекающихся в точке с наибольшей ординатой для игрока 2 (размерность (2 x n)), либо с наименьшей ординатой для игрока 1 (размерность игры (m x 2)). Выбирается данная пара стратегий и этим игра сводится к игре с размерностью (2 x 2).

4) Для выбранных стратегий решаются алгебраические уравнения для определения смешанных стратегий игроков.

Проиллюстрируем этот подход на игре (2 x n) (табл. 4.3.4). Предполагаем, что игра не имеет седловой точки в чистых стратегиях.

Таблица 4.3.4

Смешанные стратегии 1-го игрока	Смешанные стратегии 2-го игрока			
	P_1^2	P_2^2	…	P_n^2
P_1^1	$П_{11}$	$П_{12}$	…	$П_{1n}$
$P_2^1 = 1 - x_1$	$П_{21}$	$П_{22}$	…	$П_{2n}$

Введем обозначения: P_1^1 – вероятность применения первым игроком 1-ой стратегии, P_2^1 – вероятность применения первым игроком 2-ой стратегии, причем $P_2^1 = 1 - P_1^1$; P_1^2 – вероятность применения вторым игроком 1-й стратегии, P_2^2 - вероятность применения вторым игроком 2-й стратегии и т.д., P_n^2 – вероятность применения вторым игроком n-й стратегии.

Ожидаемый выигрыш первого игрока при применении вторым игроком 1-й стратегии составит:

$$П_{11} P_1^1 + П_{21} P_2^1 = П_{11} P_1^1 + П_{21}(1 – P_1^1) = П_{11} P_1^1 + П_{21} – П_{21} P_1^1 = (П_{11} – П_{21}) P_1^1 + П_{21}.$$

Аналогично найдем ожидаемые выигрыши игрока 1 при применении вторым игроком 2-й, 3-й, …, n-й стратегии. Полученные данные поместим в табл. 4.3.5.

Таблица 4.3.5

Чистые стратегии второго игрока	Ожидаемые выигрыши первого игрока
1	$(П_{11} – П_{21}) P_1^1 + П_{21}$
2	$(П_{12} – П_{22}) P_1^1 + П_{22}$
…	…
n	$(П_{1n} – П_{2n}) P_1^1 + П_{2n}$

Из табл. 4.3.5 видно, что ожидаемый выигрыш первого игрока линейно зависит от P_1^1.Построим прямые ожидаемых выигрышей игрока 1 (рис. 4.3.6).

Первый игрок должен выбирать такие стратегии, которые позволят ему максимизировать свой минимальный ожидаемый выигрыш. Поэтому оптимальная стратегия первого игрока определяется как точка пересечения прямых, в которой достигается максимум его минимального ожидаемого выиг-рыша.

Пересечение прямых, соответствующих цене игры V, соответствует паре стратегий игрока 2.

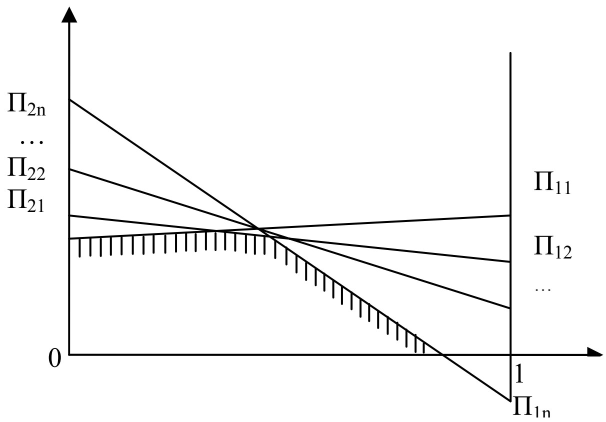


Рис. 4.3.6. Графическое решение игры для игрока 1.

Тема 4.4 Игры двух лиц с ненулевой суммой. Матричные и биматричные игры

Лекция 4.4.1 Бескоалиционные неантагонистические игры

Простейшим примером матричных игр с ненулевой суммой являются игры с постоянной суммой, в которой для двух игроков

$$\Pi^1_{ij} + \Pi^2_{ij} = a \text{, где a – константа.} \quad (4.4.1)$$

Также как и в играх с нулевой суммой две платежные матрицы игроков можно заменить одной матрицей.

Любую игру с постоянной суммой можно представить в виде некоторой матричной игры, поскольку прибавление какой-либо постоянной величины в каждой записи в матрице не влияет на результат. Если пары (Π^1_{ij}, a – Π^1_{ij}) представляют собой элементы платежной матрицы, то, вычитая $\frac{a}{2}$ из каждого числа, входящего в эту пару, получаем

$$\left(\widehat{\Pi}^1_{ij}, \widehat{\Pi}^2_{ij}\right) = \left(\Pi^1_{ij} - \frac{a}{2}, -\Pi^2_{ij} - \frac{a}{2}\right), \quad (4.3.2)$$

где $\widehat{\Pi}^1_{ij} = -\widehat{\Pi}^2_{ij}$.

Другой тип игр – игры с непостоянной ненулевой суммой.

В играх двух лиц с непостоянной ненулевой суммой есть одна существенная отличительная особенность: то, что хорошо для одного игрока, не обязательно плохо для другого. Следовательно, интересы игроков не являются полностью противоположными, и они могут извлечь выгод из сообщения своей стратегии. Будем различать два варианта таких игр: а) некооперативные (бескоалиционные игры, когда запрещены любые соглашения, обмен информацией, побочные платежи и совместный выбор стратегий, и б) кооперативные игры, в которых разрешается любой вид кооперации.

Рассмотрим бескоалиционные (некооперативные) матричные игры.

В некооперативных играх даны две функции выигрыша Π^1 и Π^2, определенные для всех пар стратегий (S^1, S^2) и соответствующие выигрышам игроков 1 и 2. Если множества стратегий конечные, то Π^1 и Π^2 можно представить в виде матриц $\Pi^1 = \left(\Pi^1_{ij}\right)$ и $\Pi^2 = \left(\Pi^2_{ij}\right)$. Игра в такой форме называется биматричной игрой.

В большинстве исследований по биматричным играм основное внимание уделяется вопросу существования равновесных пар стратегий в таких играх. В общем случае равновесных пар чистых стратегий может и не существовать. Однако доказано, что в любой биматричной игре существует по крайней мере одна равновесная пара смешанных стратегий.

В литературе существование равновесных пар стратегий (точки равновесия) называют точками равновесия «по Нэшу»[3].

Для того чтобы дать точное определение понятию точки равновесия, использующую понятие смешанной стратегии, предположим, что игрок 1 выбирает стратегию S^1_i, а игрок 2 – стратегию S^2_j, то, как и в табл. 4.1.2(3), выигрыш первого игрока равен Π^1_{ij}, а выигрыш второго – Π^2_{ij}. Если вероятность того, что игрок 1 выберет i-ю чистую стратегию S^1_i, равна p^1_i (i = 1, 2, …, m), то смешанная стратегия первого игрока выражается вектором

$$P^1 = \left(p^1_1, p^1_2, ..., p^1_m\right), \quad \text{где} \quad p^1_i \geq 0. \qquad (4.4.3)$$

Аналогично, если p^2_j – вероятность выбора j-й чистой стратегии S^2_j игроком 2 (j = 1, 2, …, n), то смешанная стратегия второго игрока выражается вектором

$$P^2 = \left(p^2_1, p^2_2, ..., p^2_n\right)', \text{ где } 1p^2 = 1,\ p^2 \geq 0. \qquad (4.4.4)$$

Точкой равновесия является пара векторов p^{1*}, p^{2*}, определяющих оптимальные смешанные стратегии каждого из игроков, т.е. стратегии, приводящие данного игрока к максимальному ожидаемому выигрышу при усло-

[3] Определение точек равновесия отнюдь не является единственным возможным способом решения некооперативных игр. Другие методы решения: принцип максимина (максимизация своих минимальных выигрышей); принцип максимакса (максимизация своих максимальных выигрышей); принцип максимальной суммы (максимизация суммы выигрышей); принцип максимальной разности (максимизация разности выигрышей) (рассмотрены в лекции 4.2.1).

вии, что противник применяет свою (оптимальную) смешанную стратегию. Следовательно,

$$P^1\Pi^1p^{2*} \le p^{1*}\Pi^1p^{2*} \text{ для всех } p^1, \quad (4.4.5)$$
$$P^{1*}\Pi^2p^2 \le p^{1*}\Pi^2p^{2*} \text{ для всех } p^2.$$

В каждой конечной игре двух лиц существует пара векторов смешанных стратегий, приводящих к точке равновесия. Такая пара векторов может быть не единственной, и может оказаться, что различным парам соответствуют различные значения (ожидаемого выигрыша). Вообще говоря, в каждой игре k лиц с конечным числом стратегии существуют смешанные стратегии, приводящие к равновесию. Равновесие – это набор таких смешанных стратегий, которые невыгодно самостоятельно изменять ни одному из игроков.

К сожалению, для игр с ненулевой суммой равновесные игры не полностью удовлетворяют понятию «решение игры».

Рассмотрим примеры ситуаций, которые могут продемонстрировать соответствующие трудности.

В первой ситуации описывается деятельность двух фирм.

Фирма C_1 может выпускать два сорта кофе: дешевый и дорогой. Вторая фирма C_2 может выпускать два типа кофейников: дорогие и дешевые. Дорогой кофе можно сварить в любом кофейнике, а дешевый кофе можно приготовить только в дорогом кофейнике.

Если одна фирма производит дорогой продукт, а другая – дешевый, то существуют соизмеримые объемы продажных товаров, при которых фирма, выпускающая дорогой продукт, получит большую прибыль, а другая фирма будет иметь некоторую разумную прибыль. Если же обе фирмы начнут выпускать дорогую продукцию, то объемы продажных товаров будут малы и ни одна из фирм не получит никакой прибыли. Если же обе фирмы произведут дешевую продукцию, то они ничего не продадут и не получат прибыли.

Данную игру можно представить с помощью двух платежных матриц:

$П^1 =$

Стратегии фирмы C_1	Стратегии фирмы C_2	
	$S_1^{\,2}$	$S_2^{\,2}$
$S_1^{\,1}$	7	0
$S_2^{\,1}$	0	3

$П^2 =$

Стратегии фирмы C_1	Стратегии фирмы C_2	
	$S_1^{\,2}$	$S_2^{\,2}$
$S_1^{\,1}$	3	0
$S_2^{\,1}$	0	7

Легко видеть что при данной игре равновесными будут пары ($S_1^{\,1}$, $S_1^{\,2}$) и ($S_2^{\,1}$, $S_2^{\,2}$). Тем не менее пары ($S_1^{\,1}$, $S_2^{\,2}$) и ($S_2^{\,1}$, $S_1^{\,2}$) не равновесны, и, более того, выигрыши в равновесных точках различны. Таким образом, предварительное соглашение было бы выгодно в данной игре, но оно запрещено правилами.

Пример второй ситуации известен в литературе как «Дилемма заключенного».

Два преступника ожидают приговора суда за совершенное злодеяние. Адвокат конфиденциально предлагает каждому из преступников облегчить его участь (и даже освободить!), если он сознается и даст показания против сообщника. Если никто не сознается, то обоим угрожает заключение на определенный срок по сфабрикованному обвинению в незначительном преступлении. Каждый заключенный имеет на выбор две стратегии: не сознаваться ($S_1^{\,1}$ и $S_1^{\,2}$) или сознаваться ($S_2^{\,1}$ $S_2^{\,2}$), выдав при этом сообщника. Если разумно назначить полезность, то можно получить следующие платежные матрицы:

$\Pi^1 =$

Стратегии заключенного 1	стратегии заключенного 2	
	S_1^2	S_2^2
S_1^1	5	0
S_2^1	10	1

$\Pi^2 =$

Стратегии заключенного 1	стратегии заключенного 2	
	S_1^2	S_2^2
S_1^1	5	10
S_2^1	0	1

Легко видеть, что (S_2^1 S_2^2) – единственная равновесная пара стратегий. Однако, если оба игрока сыграют не правильно, то их неравновесные стратегии (S_1^1, S_2^2) дадут им больший выигрыш.

Тема 4.5. Игры с непостоянной суммой. Кооперативные игры.

Лекция 4.5.1 Кооперативные игры без перевода и платежей.

Кооперативной игрой называется игра с непостоянной суммой, в которой игрокам разрешается обсуждать перед игрой свои стратегии и договариваться о совместных действиях; иначе говоря, игроки могут образовывать коалиции. Основная задача в кооперативной игре состоит в дележе общего выигрыша между членами коалиции. Следует различать кооперативные игры с побочными платежами, в которых платежи являются переводимыми, и игры без побочных платежей, в которых платежи непереводимы.

Предполагается, что в кооперативной игре без перевода платежей игроками игре достигают некоторого соглашения о согласовании использования своих стратегий. При этом игроки стремятся вступить в сговор с целью получить возможность использовать свои эффективные стратегии, тем самым получить больший выигрыш от их использования. В случае, если не удалось им скоординировать свои действия, то каждый игрок получит некоторый фиксированный платеж Т, называемый платеж при угрозе. Так при индивидуальном участии в игре платеж при угрозе может быть оценен как максиминный.

$$T^1 = \max_i \min_j \Pi_i^1 j, \qquad (4.5.1)$$

где T^1 – платеж при угрозе игрока 1 при известной платежной матрице $\{\Pi ij^1\}$.

Принципы решения кооперативной игры без перевода платежей сформулированы Дж. Нэшем.

На примере случая создания коалиции из 2-х игроков, где каждый игрок максимизирует свой выигрыш, (рис. 4.5.1) они включают:

1) Множество общих выигрышей игроков Π^1 и Π^2 является выпуклым, замкнутым и ограниченным сверху.

2) Игрокам известны значения платежей при угрозе T^1 и T^{2}.

3) Необходимость создания коалиции осознается игроками только на верхней границе допустимо множество и только на множестве Парето оптимальных решений, это множество на котором увеличение выигрыша одного из игроков возможно только за счет изменения выигрыша партнера. Иными словами, игроки не находясь не на верхней границе допустимого множества

(рис.4.5.1), где игроки могут улучшать свое положение индивидуально создавать коалиции не будут. Коалиция может быть образована только при выходе на верхнюю границу (рис. 4.5.1). Переговорное множество о создании коалиции есть платежи из множества Парето оптимальных решений суженных значениями платежей при угрозе T^1 и T^2.

4) Игра симметрична, т.е. решение не зависит от того, какие номера присвоены игрокам.

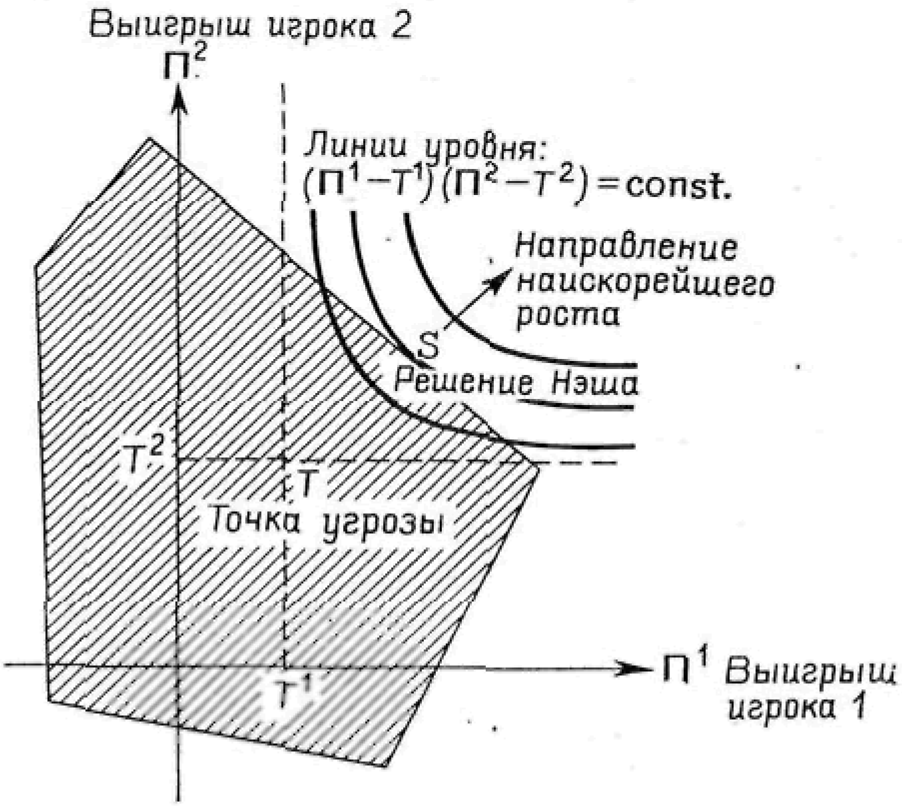
5) Игра инвариантна относительно монотонно линейных преобразований платежей.

6) Решение игры не зависит от заведомо не эффективных стратегий, т.е. решение не изменится если исключить их из рассмотрения.

7) Если условия 1) – 6) выполнить, то единственным равновесным решением по Нэшу является пара платежей $\left(\Pi^{1*};\Pi^{2*}\right)$, которые максимизируют произведение превышения этих платежей над платежами при угрозе

$$\max_{\Pi^1\Pi^2}\left(\Pi^1-T^1\right)\left(\Pi^2-T^2\right) \tag{4.5.2}$$

Графическое решения дано на рис. 4.5.1



Заштрихованная часть плоскости соответствует множеств возможных платежей; это множество выпукло, так как игроки могут применять смешанные стратегии. Прямой линией на границе множества отмечена «передовая линия» платежей, т.е. множество всех пар платежей, которые удовлетворяют допущению оптимальности по Парето. Точкой угрозы является точка T, а решением Нэша является точка S, в которой передовая линия платежей достигает линии наибольшего уровня. Линии уровня в данном случае – это равносторонние гиперболы с центром в T. Решение является единственным, т.е. множеству всех точек на передовой линии платежей, в которых выигрыши игроков больше, чем в точке угрозы.

В общем случае кооперативные игры получаются в тех ситуациях, когда, в игре n игроков разрешается образовывать определенные коалиции. Обозначим через N множество всех игроков, N= {1,2,…, n}, а через K – любое его подмножество. Пусть игроки из K договариваются между собой о совместных действиях и, таким образом, образуют одну коалицию. Очевидно, что число таких коалиций, состоящих из r игроков, равно числу сочетаний из n по к, то есть C_n^r? А число всевозможных коалиций равно

$$\sum_{r=}^{n} \mathrm{C}_{\mathrm{n}}^{\mathrm{r}} = 2^n - 1 \tag{4.5.3}$$

Из этой формулы видно, что всевозможных коалиций значительно растет в зависимости от числа всех игроков в данной игре. Для исследования этих игр необходимо учитывать все возможные коалиции, и поэтому трудности исследований возрастают с ростом n. Образовав коалицию, множество игроков K действуют как один игрок против остальных игроков, и выигрыш этой коалиции зависит от применяемых стратегий каждым из n игроков.

Функция υ, ставящая в соответствие каждой коалиции K наибольший, уверенно получаемый его выигрыш υ (K), называется **характеристической функцией игры.** Так, например, для бескоалиционной игры n игроков υ (K) может получиться, когда игроки из множества K оптимально действуют как один игрок против остальных N| K игроков, образующих другую коалицию (второй игрок).

Характеристическая функция υ называется **простой** или **нормализованной,** если она принимает только два значения: 0 и 1. Если характеристическая функция υ простая, то коалиция К, для которых υ (К) = 1, называются **выигрывающими**, а коалиции К, для которых υ (К) = 0, **проигрывающими**.

Если в простой характеристической функции υ выигрывающими являются те и только те коалиции, которые содержат фиксированную непустую коалицию R, то характеристическая функция υ, обозначаемая в этом случае через υ_R, называется **простейшей**.

Содержательно простые характеристические функции возникают, например, в условиях голосования, когда коалиция является выигрывающей, если она собирает более половины голосов (простое большинство) или не менее двух третей голосов (квалифицированное большинство).

Обозначим через υ_G характеристическую функцию бескоалиционной игры. Эта функция обладает следующими свойствами:

1)персональность

$$\upsilon_G\ (\varnothing) = 0, \tag{4.5.4}$$

т.е. коалиция, не содержащая ни одного игрока, ничего не выигрывает;

2) супераддитивность

$$\upsilon_G(\mathrm{K} \cup \mathrm{L}) \geq \upsilon_G(\mathrm{K}) + \upsilon_G(\mathrm{L}),\ \text{ если } \mathrm{K}, \mathrm{L} \subset \mathrm{N}, \mathrm{K} \cap \mathrm{L} \neq \varnothing \tag{4.5.5}$$

т.е. общий выигрыш коалиции не меньше суммарного выигрыша всех участников коалиции;

3) дополнительность

$$\upsilon_G(K) + \upsilon(N \setminus K) = \upsilon(N) \tag{4.5.6}$$

т.е. для бескоалиционной игры с постоянной суммой сумма выигрышей коалиции и остальных игроков должна равняться общей сумме выигрышей всех игроков.

Лекция 4.5.2 Кооперативные игры с дележом (с переводом платежей).

Кооперативными играми с дележом (с переводом платежей, с побочными платежами) называются игры, в которых допускается заключение взаимнообязывающих соглашений о стратегиях, а платежи могут распределяться между игроками. Поскольку разрешены побочные платежи, то следует рассматривать только общие выигрыши любых возможных коалиций. Игры такого типа будем анализировать, пользуясь характеристической функцией игры, с помощью которой описываются все возможные коалиции, а именно указывается, какой максимальный общий выигрыш может гарантировать себе каждая из коалиций.

Рассмотри пример, значения характеристической функции для игры трех лиц представлены в табл. 4.5.2.

Табл. 4.5.2

Характеристическая функция для игры трех участников

Варианты коалиций	Значение выигрышей
Ноль игроков	υ (φ) = 0
Один игрок	υ (1) = 0 υ (2) = 0 υ (3) = 0
Два игрока	υ (1,2) = 0,1 υ (1, 3) = 0,2 υ (2,3) 0,2
Три игрока	υ (1, 2, 3) = υ (N) = 1

Каждая из четырех строк соответствует значениям характеристической функции для коалиций, число игроков в которых равно соответственно 0, 1, 2, 3. Первая строка таблицы отражает предположение, что максимальный выигрыш для пустого множества равен нулю. Вторая строка таблицы показывает, что выигрыш любого игрока, действующего в одиночку равен нулю. В третьей строке таблицы указаны выигрыши трех коалиций, которые могут быть составлены из трех игроков. Если игроки 1 и 2 действуют совместно, они могут гарантировать себе выигрыш в размере 0,1; коалициям игроков 1 и 3 или 2 и 3 гарантирован выигрыш, равный 0,2. Наконец, последняя строка показывает, что если все игроки объединяются в «большую коалицию», то выигрыш равен 1. Рассматриваемая игра представлена в простой нормализованной форме 0 – 1:

$$\upsilon(i) = 0 \text{ при любом } i \in N, \qquad (4.5.7)$$
$$\upsilon(N) = 1, \text{ где } N = \{1, \ldots, n\},$$

т.е. выигрыш самостоятельно действующего игрока равен нулю, а выигрыш большой коалиции, включающий всех игроков, равен единице. Если для всех непересекающихся подмножеств K и L выполняется неравенство

$$\upsilon(K \cup L) \geq \upsilon(K) + \upsilon(L), \qquad (4.5.8)$$

то характеристическая функция является супераддитивной. Это значить, что если нет ни одного игрока, которых входил бы в обе коалиции K и L, то коалиция, составленная как объединение этих двух подмножеств, будет иметь выигрыш, не меньший, чем сумма выигрышей K и L. Предположение о супераддитивности характеристической функции вполне приемлемо, поскольку создание коалиции было бы бессмысленным, если бы величина выигрыша уменьшилась с увеличением числа участников коалиции.

Вектор П n-мерного евклидова пространства, компонентами которого являются суммарные выигрыши каждого отдельного игрока, называется «дележом»:

$$П = (П^1, П^2, \ldots, П^n), \qquad (4.5.9)$$

где $П^i$ – выигрыш i-го игрока (i = 1, 2, …, n). Примером дележа для игры в таблице 4.5.2., является вектор (0,3; 0,3; 0,5), т.е. игрок 1 получает 0,3, игрок 2 получает 0,2, а игрок 3 получает 0,5.

Распределение выигрышей (дележ) игроков должно удовлетворять следующим естественным условиям: во-первых, должно удовлетворяться условие **индивидуальной рациональности**

$$П^i \geq \upsilon(i), \text{ для } i \in N \qquad (4.5.10)$$

т.е. любой игрок должен получить выигрыш в коалиции не меньше, чем которое получил бы, не участвуя в ней (в противном случае он не будет участвовать в коалиции); во-вторых, должно удовлетворяться условие **коллективной рациональности**

$$\sum_{i \in N} \Pi_i = \upsilon(N) \qquad (4.5.11)$$

т.е. сумма выигрышей игроков должна соответствовать возможностям (если сумма выигрышей всех игроков меньше, чем υ(N), то игрокам незачем вступать в коалицию; если же потребовать, чтобы сумма выигрышей была больше, чем υ(N), то это значит, что игроки должны делить между собой сумму большую, чем у них есть).

Таким образом, вектор $\Pi = (\Pi_1, \ldots, п_n)$, удовлетворяющий условиям индивидуальной и коллективной рациональности, называется **дележом** в условиях характеристической функции υ.

Система {N, υ}, состоящая из множества игроков, характеристической функции над этим множеством и множеством дележей, удовлетворяющих соотношениям (4.5.10) и (4.5.11) в условиях характеристической функции, называется **классической кооперативной игрой с переводом платежей**.

Из этих определений непосредственно вытекает следующее утверждение: чтобы вектор $\Pi = (\Pi_1, \ldots, \Pi_n)$ был дележом в классической кооперативной игре {N, υ}, необходимо и достаточно чтобы

$$\Pi_i = \upsilon\,(i) + \alpha_i,\ (i \in N)$$

причем

$$\alpha_i \geq 0\ (i \in N) \qquad (4.5.12)$$

$$\sum_{i \in N} \alpha_i = \upsilon(N) - \sum_{i \in N} \upsilon(i).$$

В бескоалиционных играх исход формируется в результате действий тех самых игроков, которые в этой ситуации получают свои выигрыши. Исходом в кооперативной игре является дележ, возникающий не как следствие действия игроков, а как результат их соглашений. Поэтому в кооперативных играх сравниваются не ситуации, как это имеет место в бескоалиционных играх, а дележи, и сравнение это носит более сложный характер.

Кооперативные игры считаются **существенными**, если для любых коалиций K и L выполняется неравенство

$$\upsilon(K) + \upsilon(L) < \upsilon(K \cup L), \qquad (4.5.13)$$

т.е. в условии супераддитивности выполняется строгое неравенство. Если же в условии супераддитивности выполняется равенство

$$\upsilon(K) + \upsilon(L) = (K \cup L), \qquad (4.5.14)$$

т.е. выполняется свойство аддитивности, то такие игры называются **несущественными**.

Справедливы следующие свойства:

1) Для того чтобы характеристическая функция была аддитивной (кооперативная игра – несущественной), необходимо и достаточно выполнение следующего равенства:

$$\sum_{i \in N} \upsilon(i) = \upsilon(N) \tag{4.5.15}$$

2) В несущественной игре имеется только один дележ

$$\{\upsilon(1), \upsilon(2), \ldots, \upsilon(n)\}; \tag{4.5.16}$$

3) в существенной игре с более чем одним игроком множество дележей бесконечно

$$(\upsilon(1) + \alpha_1, \upsilon(2) + \alpha_2, \ldots, \upsilon(n) + \alpha_n) \tag{4.5.17}$$

где $\alpha_i \geq 0$ $(i \in N)$, $\upsilon(N) - \sum_{i \in N} \upsilon(i) > 0$.

Кооперативная игра с множеством игроков N и характеристической функцией υ называется **стратегически эквивалентной** игрой с тем же множеством игроков и характеристической функцией υ^1, если найдутся такие $k > 0$ и произвольные вещественные C_i $(i \in N)$, что для любой коалиции $K \subset N$ имеет место равенство:

$$\upsilon^1(K) = k\upsilon(K)\sum_{i \in K} C_i \tag{4.5.18}$$

Смысл определения стратегической эквивалентности кооперативных игр (СЭКИ) состоит в том, что характеристические функции СЭКИ отличаются только масштабом измерения выигрышей k и начальным капиталом C_i. Стратегическая эквивалентность кооперативных игр с характеристическими функциями υ и υ^1 обозначается так $\upsilon \sim \upsilon^1$. Часто вместо стратегической эквивалентности кооперативных игр говорят о стратегической эквивалентности их характеристических функций.

Справедливы следующие свойства для стратегически эквивалентных игр:

1) **Рефлективность**, т.е. каждая характеристическая функция эквивалентна себе $\upsilon \sim \upsilon$.

2) **Симметрия**, т.е. если $\upsilon \sim \upsilon^1$, то $\upsilon^1 \sim \upsilon$.

3) **Транзитивность**, т.е. если $\upsilon \sim \upsilon^1$ и $\upsilon^1 \sim \upsilon^2$, то $\upsilon \sim \upsilon^2$.

Из свойств рефлективности, симметрии и транзитивности вытекает, что множество всех характеристических функций единственным образом распадается на попарно непересекающиеся классы, которые называются **классами стратегической эквивалентности**.

Отношение стратегической эквивалентности игр и их характеристических функций переносится на отдельные дележи:

Пусть $\upsilon \sim \upsilon^1$, тогда $\Pi = (\Pi_1, \ldots, \Pi_n)$ – дележи в условиях характеристической функции υ; рассмотрим вектор $\Pi^1 = (\Pi_1{}^1, \ldots, \Pi_n{}^1)$, где $\Pi^1 = k\Pi_i + C_i$; для него выполняется

$$\Pi_i{}^1 = k\Pi_i + C_i \geq k\upsilon(i) + C_i = \upsilon^1(i); \qquad (4.5.19)$$

т.е. выполняется условие индивидуальной рациональности, и

$$\sum_{i\in N} \Pi_i^1 = \sum_{i\in N} \left(k\Pi_i + C_i\right) = k\sum_{i\in N} \Pi_i + \sum_{i\in N} C_i = k\upsilon(N) + \sum_{i\in N} C_i = \upsilon^1\left(N\right) \qquad (4.5.20)$$

т.е. выполняется условие коллективной рациональности. Поэтому вектор Π^1 является дележом в условиях υ^1. говорят, что дележ Π^1 соответствует дележу Π при стратегической эквивалентности $\upsilon \sim \upsilon^1$.

Кооперативная игра называется **нулевой**, если все значения ее характеристической функции равны нулю. Содержательное значение нулевой игры состоит в том, что в ней игроки не имеют никакой заинтересованности.

Всякая несущественная игра стратегически эквивалентна нулевой.

В теории игр показано, что каждая существенная кооперативная игра стратегически эквивалентна одной и только одной игре в нормализованной форме, где

$$\upsilon(i) = 0 \ (i \in N),$$
$$\upsilon(N) = 1.$$

Это означает, что мы можем выбрать игру в нормализованной форме для представления любого класса эквивалентности игр. Удобство этого выбора состоит в том, что в такой форме значение $\upsilon(K)$ непосредственно демонстрирует нам силу коалиции S (т.е. ту дополнительную прибыль, которую получают члены коалиции, образовав ее), а все дележи являются вероятностными векторами.

В игре в нормализованной форме дележом является любой вектор $\Pi = \Pi_1, \ldots, \Pi_n)$, для которого

$$\Pi_i \geq 0 \ (i \in N) \ \sum_{i \in N} \Pi_i = 1$$

Множество таких дележей является чрезвычайно большим.

Поэтому, для того чтобы еще дополнительно сузить это множество, приходится вводить какой либо критерий допустимости или доминирования на множестве платежей.

Одним из слабых критериев доминирования на множестве дележей является критерий, называемый решение по Нейману – Моргенштерну. Множество игроков называется эффективным при данном дележе, если их общий выигрыш после объединения в коалицию будет, по меньшей мере, равен сумме выигрышей, получаемых каждым игроком в отдельности. Следовательно, коалиция S является эффективной при дележе $\Pi = (\Pi^1, \ldots, \Pi^n)$, если

$$\upsilon\,(S) \geq \sum_{i \in S} \Pi^i \ . \qquad (4.5.21)$$

Дележ $\Pi_1 = (\Pi_1^{\,1}, \Pi_1^{\,2}, \ldots, \Pi_1^{\,n})$ доминирует над дележом $\Pi_2 = (\Pi_2^{\,1}, \Pi_2^{\,2}, \ldots, \Pi_2^{\,n})$, если существует такая коалиция игроков, эффективная при Π_1, что каждый из игроков, вступивших в коалицию, получает при Π_1 больше, чем при Π_2. Иначе говоря, имеется некоторая коалиция S, такая, что

$$\upsilon\,(S) \geq \sum_{i \in S} \Pi_1^i \ , \qquad (4.5.22)$$

причем каждый член этой коалиции получает Π_1 больше, чем при Π_2.

$$\Pi_1^{\,i} > \Pi_2^{\,i} \qquad (4.5.23)$$

при всех $i \in S$.

Решением игры по Нейману и Моргенштерну называется множество дележей со следующими свойствами: ни один из дележей этого множества не доминирует над другим дележом из того же множества; для любого дележа, не входящего в множество, существует доминирующий дележ, принадлежащий данному множеству. Этот слабый критерий доминирования обычно несколько сужает множество дележей, однако он не приводит, как правило, к множеству, состоящему из одного дележа. В действительности, решение Неймана – Моргенштерна часто содержит бесконечное число дележей.

Более сильный критерий доминирования на множестве дележей можно задать с помощью понятия «ядра» игры. Ядро – это некоторое подмножество каждого решения Неймана – Моргенштерна (если такие решения существуют). Число дележей, входящих в ядро, значительно меньше, чем в решении Неймана – Моргенштерна, поскольку дележи, входящие в ядро, должны удовлетворять следующему условию: каждая из коалиций при данном дележе получает, по меньшей мере, столько, сколько могли бы получить в сумме входящие в нее игроки, действуя самостоятельно. Ядрами называется множество всех недоминируемых дележей, т.е. таких дележей $\Pi = (\Pi^1, \ldots, \Pi^n)$, которые удовлетворяют условию

$$\sum_{i \in S} \Pi^i \geq \nu(S) \text{ для любого подмножества S из N.} \qquad (4.5.24)$$

Следовательно, множество дележей, входящих в ядро, удовлетворяет условию «коалиционной рациональности». Это условие включает более частные условия «индивидуальной рациональности» (когда рассматриваются подмножества, состоящие из отдельных игроков), «групповой рациональности » (когда подмножеством является большая коалиция, включающая всех игроков) и условие рациональности любой коалиции промежуточного размера.

Представляется вполне разумным следующее предположение: если игра имеет ядро, то все выбираемые дележи должны принадлежать ядру. Это означает, что игроки учитывают все возможные коалиции. Однако, к сожалению, многие игры не имеют ядра (ядро является пустым множеством), т.е. не существует дележей, удовлетворяющих условию коалиционной рациональности для какой бы то ни было коалиции.

Число дележей, входящих в ядро, как правило, либо равно нулю (т.е. ядро пустое), либо их много. Ядра, состоящие из единственного дележа, встречаются очень нечасто. Однако в играх с ценой Шепли ядро всегда состоит из одного дележа. Ценой Шепли называется дележ, величина платежей в котором зависит от «силы» каждого игрока. Последняя учитывается, исходя из значения дополнительного выигрыша, который может получить коалиция, если данный игрок войдет в нее.

Предположим, что каждый игрок получает выигрыш, равный средней величине своих вкладов во все те коалиции, куда он мог бы вступить. Выигрыш i-го игрока равен средней взвешенной из [υ (S $\cup$ {i}) – υ (S)], где S – это любое подмножество игроков, не содержащее игрока i, а [S $\cup$ {i}] тоже самое подмножество, включающее игрока i. Средняя взвешенная равна платежу

$$\Pi^{i} = \sum_{\text{все } S \subset N} \gamma_n(S)[\nu(S \cup \{i\}) - \nu(S)], \tag{4.5.25}$$

где взвешивающие множители γ_n (S) равны

$$\gamma_n(S) = \frac{s!(n-s-1)!}{n!}, \tag{4.5.26}$$

где s – это число игроков в S. Выбор именно таких взвешивающих множителей обусловлен следующими обстоятельствами: коалиция из n участников может быть образована n! различными способами; существует s!различных способов организации для s игроков, входящих в коалицию S до того, как к ней присоединяется игрок i; игроки, не входящие в расширенную коалицию, число которых равно n – s – 1, могут быть организованы (n – s – 1)! различными способами. Следовательно, если предположить, что все n! способов формирования коалиций, состоящих из n игроков, равновероятны, то γ_n (S) представляет собой не что иное, как вероятность присоединения игрока i к коалиции S.

Одним из ярких примеров игр с дележом Шепли является игра с N – экономическими агентами, где одним из игроков является государство. При этом сила игрока – государство такова, что он усиливает любую коалицию с позиции увеличения дележа. Поэтому игры с ценой Шепли могут быть использованы для оценки стоимости инструментов государственного регулирования экономических процессов.

Основная литература:

1. *О.О. Замков, А.В. Толстопятенко, Ю.Н. Черемных Математические методы в экономике. Москва, изд-во ДИС, 1998, 365 с.*
2. *М.С. Красс, Б.П. Чупрынов. Математика в экономике. Математические методы и модели. М., Финансы и статистика, 2007г., 342 с.*
3. *Н.Н. Воробьев, Теория игр – М., Знание, 1976.*

Дополнительная литература:

1. Х. Таха. Введение в исследование операций. Т. 2. М., Мир, 1995, 496 с.

2. Исследование операций: Методологические основы и математические методы. Т. 1. М., Мир, 1981 г., 712 с.

3. М. Интрилигатор. Математические методы оптимизации и экономическая теории. М., Прогресс, 1975 г., 606 с.