ПРЕДИСЛОВИЕ

Настоящее учебное пособие написано на основании курса лекций по методам оптимизации и математической статистике, читаемых авторами в течение ряда лет студентам Белорусского государственного технологического университета. Оно содержит теоретический материал, полностью охватывающий основные разделы указанного курса в рамках действующей типовой программы.

Пособие состоит из двух разделов. В первом рассматриваются методы статистической обработки результатов измерений, вводятся основные понятия математической статистики, анализируются основные типы эмпирических зависимостей, приводятся точечные и интервальные оценки параметров генеральной совокупности и их классификация, излагаются элементы регрессионного и корреляционного анализа, критерии проверки статистических гипотез. Во втором разделе пособия дается представление о классическом вариационном исчислении и его современной ветви – математической теории оптимального управления, рассматриваются основные задачи и методы математического программирования и их классификация. Основное внимание уделяется задачам линейного программирования, приводятся примеры экономических моделей, сводящихся к таким задачам, даются правила построения двойственных задач. Рассматриваются задачи на условный экстремум функций многих переменных, даются необходимые, достаточные условия существования экстремумов при наличии связей. Для простейшей задачи вариационного исчисления доказываются необходимые условия экстремума, излагаются случаи интегрируемости уравнения Эйлера. Исследуются функционалы, зависящие от производных высших порядков, и формулируются необходимые условия их экстремума. Приводится основополагающий метод современной теории оптимального управления – принцип максимума Понтрягина как необходимое, а в линейном случае и как достаточное условие оптимальности. В заключение (как необходимые условия оптимальности) рассматриваются вопросы управляемости линейных стационарных динамических систем.

Все основные понятия и теоремы, изложенные в пособии, иллюстрируются примерами. Для лучшего усвоения изложенного материала формулируются задачи, имеющие сквозную нумерацию. Пособие можно использовать для самостоятельной работы студентов как очной, так и заочной форм обучения.

Авторы считают своим приятным долгом выразить глубокую благодарность рецензентам: кафедре математического моделирования и анализа данных Белорусского государственного университета, особенно ее заведующему члену-корреспонденту НАН Беларуси профессору Ю. С. Харину и доценту В. И. Лобачу, а также заведующему кафедрой высшей математики Белорусского государственного экономического университета профессору М. П. Дымкову за высказанные замечания, способствующие улучшению структуры и содержания учебного пособия.

СПИСОК СОКРАЩЕНИЙ

БП – базисные переменные
ЗВИ – задача вариационного исчисления
ЗДЛП – задача дробно-линейного программирования
ЗЛП – задача линейного программирования
ЗМП – задача математического программирования
ЗОБ – задача оптимального быстродействия
ЗОУ – задача оптимального управления
ЗПП – задача параметрического программирования
ЗТУ – задача терминального управления
ЗЦП – задача целочисленного программирования
ЛП – линейное программирование
МНК – метод наименьших квадратов
МП – математическое программирование
МС – математическая статистика
ПЗВИ – простейшая задача вариационного исчисления
СВ – случайная величина
СП – свободные переменные
СЭ – случайный эксперимент
ТВ – теория вероятностей
ТОУ – теория оптимального управления

Раздел I. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

1. ЗАДАЧИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. ГЕНЕРАЛЬНАЯ И ВЫБОРОЧНАЯ СОВОКУПНОСТИ. ВАРИАЦИОННЫЙ РЯД И ЕГО ОСНОВНЫЕ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Теория вероятностей (ТВ) – математическая наука, изучающая закономерности случайных явлений. Под случайными явлениями понимаются явления с неоднозначным исходом, происходящие при неоднократном воспроизведении определенного комплекса условий [3, 10, 15].

Математическая статистика (МС) – раздел математики, посвященный математическим методам систематизации, обработки и использования экспериментальных данных для научных и практических выводов [1, 9, 15]. При этом *статистическими данными* называются сведения о числе объектов в какой-либо более или менее обширной совокупности, обладающих теми или иными признаками. МС опирается на ТВ. Если ТВ изучает закономерности случайных явлений на основе абстрактного определения действительности (теоретической вероятностной модели), то МС оперирует непосредственно результатами наблюдений над случайным явлением, представляющими *выборку* из некоторой конечной или бесконечной гипотетической совокупности однородных объектов. *Выборкой объема* n называется множество $x_1, x_2, \ldots, x_n$ наблюдаемых значений изучаемой случайной величины (СВ) X, которые соответствуют n независимым испытаниям (опытам). МС, используя методы ТВ, позволяет не только оценить значения искомых характеристик, но и выявить степень точности выводов, получаемых при обработке данных. Другими словами, ТВ позволяет находить вероятности «сложных» событий через вероятности «простых» событий (связанных с ними определенным образом), а МС по наблюдаемым значениям (выборке) оценивает вероятности этих событий и/или осуществляет проверку предположений (гипотез) относительно этих вероятностей.

Из задач МС можно отметить следующие: 1) указать способы сбора и группировки статистических сведений, полученных в результате наблюдений или в результате специально поставленных экспериментов; 2) разработать методы анализа статистических данных в зависимости от целей исследования; 3) использовать статистические данные для научных и практических выводов. В частности, в МС осуществляется: а) оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости случайной величины от одной или нескольких случайных величин; б) проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения, вид которого известен.

Пусть рассматривается совокупность однородных объектов и изучается ее определенный *качественный* или *количественный признак X*, характеризующий эти объекты. Например, в партии изделий *качественным признаком* может служить стандартность деталей, а *количественным* – контролируемый размер деталей. Если совокупность содержит достаточно большое число объектов, то провести их сплошное обследование практически невозможно. В таком случае из всей совокупности объектов случайным образом отбирают некоторое их число и подвергают изучению. Вся подлежащая изучению совокупность предметов называется ***генеральной совокупностью***. Как уже отмечалось, исследовать данный признак у всех предметов этой совокупности зачастую не представляется возможным (либо этих предметов очень много, либо изучаемый признак в процессе исследования исчезает, либо по другим причинам). В этом случае используют выборочный *метод*, согласно которому из генеральной совокупности *случайным образом* (наудачу) выбирают n элементов. Та часть объектов, которая отобрана для непосредственного изучения из генеральной совокупности, называется ***выборочной совокупностью*** или ***выборкой*** (случайной выборкой).

Под ***генеральной совокупностью*** зачастую понимают также *множество значений* изучаемого признака этой совокупности. В этом случае под ***выборкой*** понимается множество значений $x_1, x_2, \ldots, x_n$ признака для объектов совокупности, отобранных для непосредственного изучения. Тогда ***размахом*** w ***выборки*** называют разность между максимальным $x_{\max}$ и минимальным $x_{\min}$ значениями элементов выборки: $w = x_{\max} - x_{\min}$.

Объемом совокупности (выборочной или генеральной) называют *число* (n или N) *объектов* этой совокупности. Например, если из 1000 деталей отобрано для обследования 85, то объем генеральной совокупности $N = 1000$, объем выборки $n = 85$.

Для того чтобы результаты обследования выборки отражали основные черты изучаемого признака, необходимо, чтобы объем выборки не был слишком мал. Выборка называется ***репрезентативной*** (*представительной*), если она достаточно хорошо отражает количественные соотношения генеральной совокупности. Например, о распределении жителей города по росту нельзя судить по результатам обследования жильцов одной квартиры. Ясно, что данные, относящиеся к одному высотному дому или группе домов, более показательны (репрезентативны). Считается, что выборка *репрезентативна*, если она получена в результате *случайного эксперимента* (СЭ), т. е. взята случайным образом, и ее объем достаточно велик.

Предположим, что проводится СЭ, например измеряется некоторый признак X генеральной совокупности. На измерения могут влиять как систематические ошибки (погрешность прибора и т. п.), так и случайные ошибки, получаемые в результате воздействия различных (случайных) факторов. Тогда величину X, можно интерпретировать как СВ. Ее функция распределения $F(x)$ называется ***теоретической функцией распределения*** – *функцией распределения генеральной совокупности*. На практике, как правило, она неизвестна. В этом случае можно произвести измерение (СЭ) X, в результате чего получить значение x_1 СВ. Однако, как правило, судить о значениях СВ X по одному измеренному значению x_1 не представляется убедительным. Поэтому на практике производят серию из *n независимых* СЭ – измерений (испытаний) данной СВ X. В результате получают n реализовавшихся значений

$$x_1,\ x_2,\ \ldots,\ x_n, \tag{1}$$

которые называют *наблюдаемыми* (*выборочными*) *значениями* (данной СВ X). Совокупность (1) можно интерпретировать как *выборку объема n*. Выборочные значения (1) называют *вариантами*. Последовательность вариант, записанных в порядке неубывания, называется ***вариационным рядом***.

Основное предположение МС: варианты (1) считаются независимыми в совокупности одинаково распределенными СВ, имеющими одну и ту же функцию распределения – функцию распределения $F(x)$ генеральной совокупности.

Всякий раз при необходимости подчеркнуть случайный характер выборочных значений вместо $x_1, x_2, \ldots, x_n$ будем писать $X_1, X_2, \ldots, X_n$.

Пусть из генеральной совокупности извлечена выборка объема n, причем значение x_1^* признака X наблюдалось n_1 раз, значение x_2^* – n_2 раза, …, значение x_k^* – n_k раз. Тогда $\sum_{i=1}^{k} n_i = n$. Число наблюдений n_i значения x_i^* называют ***частотой***, а отношение частоты к объему выборки $w_i = n_i / n$ – ***относительной частотой*** w_i: $i = 1, 2, \ldots, k$ (в сокращенной записи: $i = \overline{1,\ k}$), при этом $\sum_{i=1}^{k} w_i = \sum_{i=1}^{k} \frac{n_i}{n} = \frac{1}{n} \sum_{i=1}^{k} n_i = 1.$

В ТВ в случае дискретной СВ под законом распределения понимают [3, 9, 15] соответствие между возможными значениями x_i, $i = \overline{1,\ k}$, случайной величины X и их вероятностями $p_i = P\ (X = x_i)$:

Значения x_i	x_1	x_2	…	x_k
Вероятности p_i	p_1	p_2	…	p_k

$$\sum_{i=1}^{k} p_i = 1.$$

В МС под ***статистическим распределением*** понимают [9, 10, 15] соответствие между наблюдаемыми вариантами x_i^* и их частотами n_i (или относительными частотами w_i).

Под ***статистическим*** (группированным) ***рядом*** будем понимать совокупность пар $\left(x_i^*, w_i\right)$, $i = \overline{1,\ k}$, где x_1^*, x_2^*, …, x_k^* – различные элементы выборки, а w_1, w_2, …, w_k – соответствующие относительные частоты $\left(\sum_{i=1}^{k} w_i = 1\right)$. Статистическое распределение выборки записывается в следующем виде:

Варианты x_i^*	x_1^*	x_2^*	…	x_k^*
Частоты n_i	n_1	n_2	…	n_k

$$\sum_{i=1}^{k} n_i = n,$$

или

Варианты x_i^*	x_1^*	x_2^*	…	x_k^*
Относительные частоты w_i	w_1	w_2	…	w_k

$$\sum_{i=1}^{k} w_i = 1.$$

Последнее распределение можно интерпретировать как ряд распределения фиктивной СВ ξ. Для наглядности статистическое распределение (*дискретной* СВ) иллюстрируют ***полигоном*** распределения и/или ***гистограммой*** (для *непрерывной* СВ).

Полигоном частот (относительных частот) статистического ряда называется ломаная с вершинами в точках $\left(x_i^*; n_i\right)$, $i=\overline{1,k}$, $\left(\left(x_i^*; w_i\right), i=\overline{1,k}\right)$; по оси абсцисс откладываются выборочные значения x_i^*, по оси ординат – соответствующие частоты n_i (рис. 1) или относительные частоты w_i.

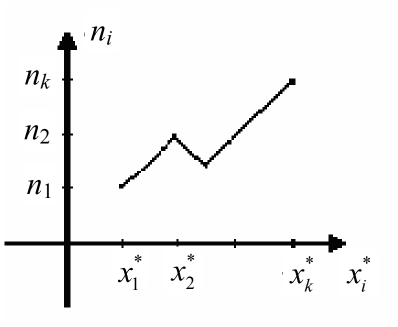


Рис. 1. Полигон частот

При большом объеме выборки или при изучении непрерывной СВ ее элементы объединяют в группы (разряды, интервалы), представляя результаты опытов в виде *интервального статистического ряда*. Для этого весь диапазон значений случайной величины X (от $x_{\min}$ до $x_{\max}$) разбивают на k интервалов (обычно) одинаковой длины $h=\dfrac{x_{\max}-x_{\min}}{k}$. Число интервалов k во избежание громоздкости статистического ряда следует брать не очень большим, но, чтобы не потерять особенности распределения признака, и не очень малым. Обычно $k\in[5;\,20]$. Для определения количества интервалов k иногда применяется *формула Стерджеса* [15]: $k = 1 + 3{,}322\lg n$. Затем подсчитывают частоты n_i (или относительные частоты w_i) значений выборки, попавших в соответствующие интервалы. Величина n_i / h называется *плотностью частоты*, а w_i / h – *плотностью относительной частоты*.

Пусть далее x_i^* – середина i-го интервала, n_i – число элементов выборки, попавших в i-й интервал (элемент, совпавший с верхней границей интервала, относится к последующему интервалу). Таким образом, получаем *интервальный* (*группированный*) *статистический ряд* [9]:

Интервал $[x_i; x_{i+1})$	$[x_{\min}; x_2)$	$[x_2; x_3)$	…	$[x_k; x_{\max}]$
Середина интервала x_i^*	x_1^*	x_2^*	…	x_k^*
Частота n_i	n_1	n_2	…	n_k
Относительная частота $w_i=\dfrac{n_i}{n}$	w_1	w_2	…	w_k

Гистограммой частот (относительных частот) интервального статистического ряда называется [9, 10, 15] ступенчатая

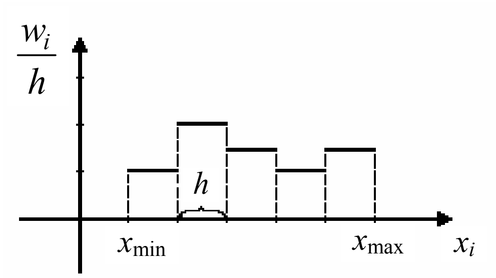


Рис. 2. Гистограмма относительных частот

фигура, составленная из прямоугольников, построенных на интервалах группировки длины h и высоты n_i / h (w_i / h).

Для построения гистограммы относительных частот (рис. 2) на оси абсцисс откладывают частичные интервалы длины h, над ними параллельно оси ординат проводят отрезки длиной w_i / h.

Площадь i-го прямоугольника равна w_i – относительной частоте вариант, попавших в i-й интервал. Следовательно, *площадь гистограммы относительных частот равна сумме всех относительных частот (т. е. равна 1), а площадь гистограммы частот равна объему выборки n.*

2. ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

Пусть известно статистическое распределение количественного признака X; n_x – число наблюдений, при которых наблюдалось значение признака, меньшее x ($X < x$); n – общее число наблюдений – объем выборки. Тогда относительная частота события $X < x$ есть n_x / n. При изменении x меняется и n_x / n, т. е. относительная частота n_x / n зависит от x. Так как эта зависимость находится эмпирическим (опытным) путем, то ее называют ***эмпирической.***

Эмпирической функцией распределения (функцией распределения выборки) называется [9, 10, 15] функция

$$F^*(x) = \frac{n_x}{n}, \tag{2}$$

определяющая для каждого значения x относительную частоту события $X < x$. В (2) n_x – число вариант, меньших x, $n_x = \sum\limits_{x_i < x} n_i$, x_i – варианты. Различие между эмпирической $F^*(x)$ и теоретической $F(x)$ функциями распределения в том, что $F(x)$ определяет вероятность события $X < x$, а $F^*(x)$ – относительную частоту того же события. Функция $F^*(x)$ обладает свойствами, аналогичными свойствам $F(x)$, и является ее статистическим аналогом.

Свойства эмпирической функции распределения $F^*(x)$:

1) значения $F^*(x)$ принадлежат отрезку [0; 1], т. е. $F^*(x) \in [0; 1]$;

2) $F^*(x)$ – неубывающая функция;

3) если x_1 – наименьшая варианта, а x_k – наибольшая варианта, то $F^*(x) = 0$ для $x \le x_1$, $F^*(x) = 1$ для $x > x_k$;

4) $F^*(x)$ – непрерывная слева функция.

Эмпирическая функция распределения выборки** $F^*(x)$ **служит для оценки теоретической функции распределения** $F(x)$ **генеральной совокупности.

Пример 1. Построить эмпирическую функцию распределения, полигон частот и относительных частот по данному распределению выборки:

Варианты x_i	2	6	10
Частоты n_i	15	27	18

$$n = \sum_{i=1}^{3} n_i = 60.$$

Здесь $x_1 = 2$ – наименьшая варианта, следовательно, $F^*(x) = 0$ для $x \le 2$; $x_3 = 10$ – наибольшая варианта, тогда $F^*(x) = 1$ при $x > 10$. Для $2 < x \le 6$ имеем $F^*(x) = n_x / n = \sum_{x_i < x} n_i / n = 15/60 = 0{,}25$, а для $6 < x \le 10$ следует $F^*(x) = (15 + 27) / 60 = 0{,}7$.

Таким образом, эмпирическая функция распределения $F^*(x)$ может быть записана в виде

$$F^*(x) = \begin{cases} 0, & x \le 2, \\ 0,25, & 2 < x \le 6, \\ 0,7, & 6 < x \le 10, \\ 1, & x > 10. \end{cases}$$

Ее график представлен на рис. 3, а полигоны частот и относительных частот – на рис. 4, 5.

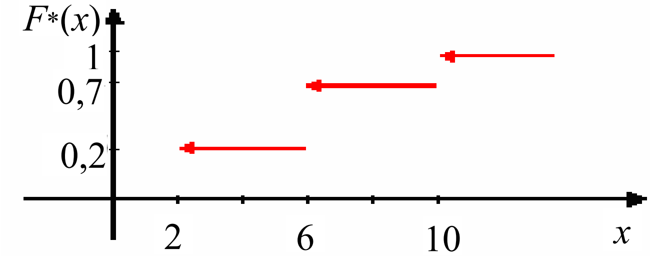


Рис. 3. Эмпирическая функция распределения (пример 1)

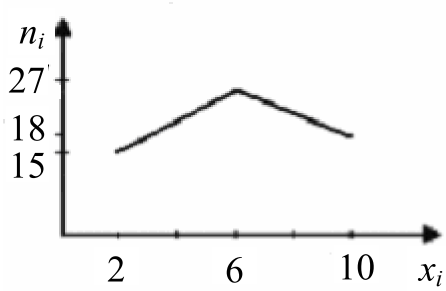


Рис. 4. Полигон частот (пример 1)

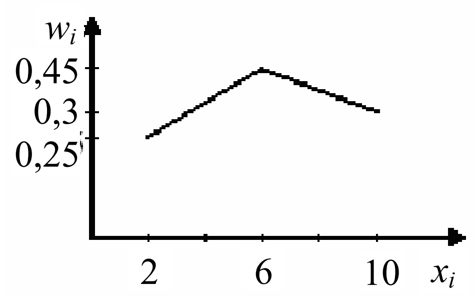


Рис. 5. Полигон относительных частот (пример 1)

Координаты w_i вершин (x_i; w_i) полигона относительных частот определяются по формулам

$$w_1 = \frac{15}{60} = 0,25;\ w_2 = \frac{27}{60} = 0,45;\ w_3 = \frac{18}{60} = 0,3;\ \sum_{i=1}^{3} w_i = 1. \quad \blacktriangle$$

3. ТОЧЕЧНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Анализ полигона, гистограммы, эмпирической функции распределения дает возможность сделать предположение о законе распределения СВ. По виду полученной гистограммы можно строить гипотезы об истинном характере распределения СВ X. Например, из гистограммы вида, представленного на рис. 6, естественно предположить, что распределение СВ X является *нормальным* [9, 10, 15].

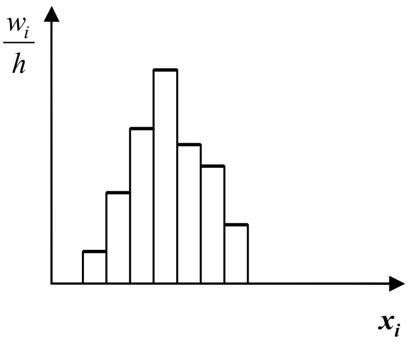


Рис. 6. Гистограмма относительных частот

Получив гистограмму, изображенную на рис. 7, можно предположить, что СВ X на отрезке [0,5; 2,5] распределена *равномерно*. На практике, однако, редко встречаются такие ситуации, когда изучаемый закон распределения СВ X полностью неизвестен. Чаще всего из каких-либо соображений вид закона распределения ясен заранее и требуется найти только некоторые *параметры*, от которых он зависит. Например, если известно, что закон распределения СВ X нормальный, т. е. $X \in N(a; \sigma)$ (с плотностью распределения $f_X(x) = \dfrac{1}{\sigma\sqrt{2p}} e^{-\frac{(x-a)^2}{2y^2}}$), то задача сводится к

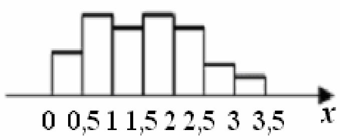


Рис. 7. Гистограмма

нахождению значений двух параметров: a и σ. В некоторых задачах знание закона распределения не требуется, а востребованы только его числовые характеристики. В подобных случаях, как правило, исследовать все элементы генеральной совокупности не представляется возможным. Часто о неизвестном параметре судят на основании сравнительно небольшого числа наблюдений – порядка одного или нескольких десятков.

При изучении *числовых характеристик* СВ X мы рассматриваем математическое ожидание $M(X)$, дисперсию $D(X)$, среднее квадратическое отклонение $\sigma(X)$. Числовые (т. е. *точечные*) характеристики играют большую роль в ТВ, и в МС существуют их статистические аналоги.

Пусть закон распределения СВ X содержит некоторый неизвестный параметр θ. В распоряжении исследователя имеются значения x_1, x_2, …, x_n количественного признака X, полученные в результате n независимых наблюдений. Используя эти значения (которые можно рассматривать как независимые в совокупности СВ X_1, X_2, …, X_n, имеющие тот же закон распределения, что и СВ X), требуется оценить (статистически) неизвестный параметр θ. В связи с этим возникает следующая з**адача:** исходя из набора наблюдаемых значений (выборки) $x_1, x_2, \ldots, x_n$ СВ X, полученного в результате n независимых испытаний, найти статистическую оценку θ^* неизвестного параметра θ теоретического распределения как функцию наблюдаемых выборочных значений $x_1, x_2, \ldots, x_n$, дающую приближенное (в некотором смысле) значение оцениваемого параметра θ.

Статистической оценкой θ^* (*оценкой, статистикой*) неизвестного параметра θ теоретического распределения называют всякую функцию $\theta^*(x_1, x_2, \ldots, x_n)$ наблюдаемых (выборочных) значений $x_1, x_2, \ldots, x_n$ СВ X, с помощью которой судят о значении параметра θ.

Так как X_1, X_2, …, X_n – СВ, то и оценка θ^* является СВ, закон распределения которой зависит от вида распределения СВ X и объема выборки n. В отличие от оценки θ^* оцениваемый параметр θ – величина детерминированная (неслучайная).

Вообще говоря, под статистикой в МС понимают термин, употребляемый для названия функций от результатов наблюдений (иногда от этой функции требуется свойство статистической устойчивости, т. е. ее значения в сериях независимых в совокупности испытаний «мало» различаются между собой). Статистическая устойчивость в каждой конкретной ситуации есть эмпирический физический закон, который может быть проверен только опытным путем.

Методы теории статистических оценок служат основой современной теории ошибок. Обычно в качестве неизвестных параметров выступают измеряемые физические константы, а в качестве СВ – результаты непосредственных измерений, подверженные случайным ошибкам. Статистическая оценка как функция от СВ чаще всего задается теми или иными формулами, выбор которых определяется требованиями практики. Оценки параметров подразделяются на *точечные* и *интервальные*.

Статистическую оценку параметра θ называют (при данной выборке $x_1, x_2, \ldots, x_n$) ***точечной***, если она определяется одним числом

$$\theta^* = \theta^*(x_1, x_2, \ldots, x_n), \tag{3}$$

и ***интервальной***, если она задается двумя числами $\alpha(x_1, x_2, \ldots, x_n)$ и $\beta(x_1, x_2, \ldots, x_n)$ – концами интервала, содержащего параметр θ.

К точечным оценкам θ^* предъявляется ряд *требований*:

1. Желательно, чтобы при использовании величины θ^* из (3) вместо θ не совершались *систематические ошибки* ни в сторону занижения ($M(\theta^*) < \theta$) оцениваемого параметра θ, ни в сторону его завышения ($M(\theta^*) > \theta$), т. е. чтобы выполнялось равенство

$$M(\theta^*) = \theta. \tag{4}$$

Оценка, удовлетворяющая условию (4), называется *несмещенной*.

Определение 1. **Несмещенной** называют точечную статистическую оценку θ^*, математическое ожидание которой равно оцениваемому параметру θ при любом объеме выборки.

Определение 2. **Смещенной** называют статистическую оценку θ^*, математическое ожидание которой не равно оцениваемому параметру θ.

2. Желательно, чтобы с увеличением числа n опытов ($n \to +\infty$) значение случайной величины $\theta^* = \theta_n^*$ стремилось по вероятности к оцениваемому значению θ: чтобы «рассеяние» θ^* вокруг θ удовлетворяло условию

$$D\left(\theta_n^*\right) \xrightarrow{P} 0 \quad \text{при} \quad n \to +\infty,$$

т. е. для любого сколь угодно малого $\varepsilon > 0$ выполнялось

$$P\left(\left|\theta_n^* - \theta\right| > \varepsilon\right) \to 0 \text{ при } n \to +\infty. \tag{5}$$

Оценку, обладающую свойством (5), называют ***состоятельной***. Примером состоятельной статистической оценки может служить любая точечная оценка, дисперсия которой при $n \to +\infty$ стремится к нулю.

Требование несмещенности оценки θ^* особенно важно при малом числе испытаний. Если равенство (4) не выполняется, то оценка θ^*, полученная по разным выборкам, в среднем будет либо завышать значение оцениваемого параметра θ (при $M(\theta^*) > \theta$), либо занижать его ($M(\theta^*) < \theta$). *Требование несмещенности оценки* θ^* *гарантирует отсутствие систематических ошибок при оценивании неизвестного параметра* θ.

Пусть θ_1^* и θ_2^* – две несмещенные оценки параметра θ. Тогда оценка θ_1^* является *эффективней* оценки θ_2^*, если «рассеяние» вокруг оцениваемого параметра θ для оценки θ_1^* меньше, чем для оценки θ_2^*, т. е.

$$D\left(\theta_1^*\right) < D\left(\theta_2^*\right). \tag{6}$$

Определение 3. ***Эффективной*** называют точечную статистическую оценку $\theta_n^* = \theta^*(x_1, x_2, \dots, x_n)$, которая при заданном объеме выборки n имеет минимально возможную дисперсию (т. е. ее эффективность – отношение дисперсии к наименьшей дисперсии – равна 1).

Пусть для изучения признака X генеральной совокупности извлечена выборка объема n. Наиболее распространенными оценками в математической статистике являются *выборочное среднее* $\overline{x}_\text{в}$ – оценка математического ожидания $M(X)$, *выборочная дисперсия* $D_\text{в}(X)$ (или *исправленная выборочная дисперсия* s^2) – оценка дисперсии $D(X)$, *выборочное среднеквадратическое отклонение* s – оценка среднеквадратического отклонения $\sigma(X)$.

Пусть статистическое распределение выборки задано через x_i^* и n_i, как на с. 8. Тогда *выборочным средним* $\overline{x}_\text{в}$ называется величина

$$\overline{x}_\text{в} = \frac{\sum_{j=1}^{k} n_j x_j^*}{n}, \tag{7}$$

т. е. *выборочное среднее есть среднее взвешенное значений признака с весами, равными соответствующим частотам.*

Определение 4. ***Генеральным средним*** $\overline{x}_\text{г}$ называют математическое ожидание исследуемого признака генеральной совокупности.

Основные свойства выборочного $\overline{x}_\text{в}$ и генерального $\overline{x}_\text{г}$ среднего аналогичны свойствам математического ожидания СВ X.

Таким образом, в *качестве точечной оценки математического ожидания $M(X) = m$ СВ X может служить выборочное среднее* $\overline{x}_\text{в}$: СВ x_1, x_2, …, x_n (точнее, X_1, X_2, …, X_n) имеют один и тот же закон распределения, совпадающий с законом распределения СВ X. Тогда

$$M\left(\overline{x}_{\text{в}}\right)=M\left(\frac{\sum\limits_{i=1}^{n}x_i}{n}\right)=\frac{1}{n}M\left(\sum\limits_{i=1}^{n}x_i\right)=\frac{1}{n}\sum\limits_{i=1}^{n}M\left(x_i\right)=\frac{1}{n}nM(X)=m.$$

Следовательно, согласно определению 1, оценка (7) математического ожидания СВ X является несмещенной.

Выясним, является ли оценка $\overline{x}_{\text{в}}$, представленная формулой (7), состоятельной? Для этого рассмотрим дисперсию $D\left(\overline{x}_{\text{в}}\right)$:

$$D\left(\overline{x}_{\text{в}}\right)=D\left(\frac{\sum\limits_{i=1}^{n}x_i}{n}\right)=\frac{1}{n^2}D\left(\sum\limits_{i=1}^{n}x_i\right)=\frac{1}{n^2}\sum\limits_{i=1}^{n}D\left(x_i\right)=\frac{1}{n^2}nD(X)=\frac{D(X)}{n},\quad (8)$$

где $D(X)$ – дисперсия СВ X. Так как $D\left(\overline{x}_{\text{в}}\right)\to 0$ при $n\to\infty$, то оценка (7) является состоятельной. Таким образом, ***выборочное среднее*** $\overline{x}_{\text{в}}$ ***является состоятельной несмещенной оценкой математического ожидания генеральной совокупности.*** Поэтому статистическую оценку $\overline{x}_{\text{в}}$ целесообразно использовать в качестве оценки математического ожидания генеральной совокупности. Отметим, что выборочное среднее $\overline{x}_{\text{в}}=M\xi$ является математическим ожиданием фиктивной СВ ξ (см. с. 8).

Рассмотрим точечную оценку дисперсии $D(X)$ СВ X. По определению $D(X)=M[X-m]^2$, где $m=M(X)$, т. е. $D(X)$ есть математическое ожидание СВ $[X-m]^2$. Поэтому для выборки $x_1, x_2, \ldots, x_n$ со средним значением $\overline{x}_{\text{в}}$ естественной точечной *оценкой* $D(X)$ в силу (7) представляется выражение

$$D_{\text{в}}\left(X\right)=\frac{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}_{\text{в}}\right)^2}{n}=\frac{\sum\limits_{j=1}^{k}n_j\left(x_j^*-\overline{x}_{\text{в}}\right)^2}{n}=\frac{1}{n}\sum\limits_{j=1}^{k}n_j\left(x_j^*\right)^2-\left(\overline{x}_{\text{в}}\right)^2,\quad (9)$$

называемое *выборочной дисперсией*. Отметим, что выборочная дисперсия $D_{\text{в}}(X)=D\xi$ является дисперсией фиктивной СВ ξ.

Выборочная дисперсия $D_{\text{в}}(X)$ есть *среднее арифметическое квадратов отклонений наблюдаемых значений признака X от их выборочного среднего* $\overline{x}_{\text{в}}$. Она характеризует меру рассеяния выборочных значений вокруг выборочного среднего $\overline{x}_{\text{в}}$.

Итак, математическое ожидание $M\xi$ дискретной СВ ξ оценивается как

$$M\xi\cong\overline{x}_{\text{в}}=\frac{1}{n}\sum\limits_{i=1}^{n}x_i=\frac{1}{n}\sum\limits_{j=1}^{k}n_jx_j^*,\quad \sum\limits_{j=1}^{k}n_j=n,\quad (10)$$

а дисперсия $D\xi$ –

$$D\xi \cong D_{\text{в}}\left(X\right)=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}_{\text{в}}\right)^2=\frac{1}{n}\sum_{j=1}^{k}n_j\cdot\left(x_j^*-\overline{x}_{\text{в}}\right)^2. \qquad (11)$$

Чтобы выяснить вопрос о том, является ли оценка $D_{\text{в}}(X)$ генеральной дисперсии $D(X)$ несмещенной, рассмотрим ее представление в виде $D_{\text{в}}\left(X\right)=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}_{\text{в}}\right)^2$ и выполним следующие преобразования:

$$D_{\text{в}}\left(X\right)=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}_{\text{в}}\right)^2=\frac{1}{n}\sum_{i=1}^{n}\left[\left(x_i-m\right)-\left(\overline{x}_{\text{в}}-m\right)\right]^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-m\right)^2-2\cdot\frac{1}{n}\left(\overline{x}_{\text{в}}-m\right)\sum_{i=1}^{n}\left(x_i-m\right)+\frac{1}{n}n\left(\overline{x}_{\text{в}}-m\right)^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-m\right)^2-2\left(\overline{x}_{\text{в}}-m\right)^2+\left(\overline{x}_{\text{в}}-m\right)^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-m\right)^2-\left(\overline{x}_{\text{в}}-m\right)^2. \qquad (12)$$

Найдем математическое ожидание оценки $D_{\text{в}}(X)$:

$$M\left(D_{\text{в}}(X)\right)=\frac{1}{n}\sum_{i=1}^{n}M\left(\left(x_i-m\right)^2\right)-M\left(\left(\overline{x}_{\text{в}}-m\right)^2\right)=\frac{1}{n}n\cdot D\left(X\right)-D\left(\overline{x}_{\text{в}}\right)\overset{(8)}{=}$$

$$=D\left(X\right)-\frac{D\left(X\right)}{n}=\frac{n-1}{n}D\left(X\right). \qquad (13)$$

Соотношение (13) показывает, что точечная оценка (9) дисперсии $D(X)$ является *смещенной*, так как $M\left(D_{\text{в}}\right)\neq D(X)$, а именно:

$$M\left(D_{\text{в}}\left(X\right)\right)=\frac{n-1}{n}D\left(X\right). \qquad (14)$$

Таким образом, математическое ожидание выборочной дисперсии не равно оцениваемой генеральной дисперсии.

• **Задача 1.** Докажите формулу $\sum_{i=1}^{n}\left(x_i-m\right)=\left(\overline{x}_{\text{в}}-m\right)n$, использованную при выводе соотношения (12).

Нетрудно «исправить» оценку $D_{\text{в}}(X)$ так, чтобы математическое ожидание исправленной оценки стало равно генеральной дисперсии $D(X)$. Для этого достаточно определить *исправленную выборочную дисперсию* s_X^2 следующим соотношением:

$$s_X^2=\frac{n}{n-1}D_{\text{в}}\left(X\right)\overset{(9)}{=}\frac{n}{n-1}\frac{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}_{\text{в}}\right)^2}{n}=\frac{\sum\limits_{i=1}^{n}(x_i-\overline{x}_{\text{в}})^2}{n-1}=\frac{\sum\limits_{j=1}^{k}n_j(x_j-\overline{x}_{\text{в}})^2}{n-1}.$$

Полученная точечная оценка s_X^2 генеральной дисперсии является несмещенной, так как в этом случае

$$M(s_X^2)=M\left(\frac{n}{n-1}D_{\text{в}}(X)\right)=$$

$$=\frac{n}{n-1}M\left(D_{\text{в}}\left(X\right)\right)\overset{(14)}{=}\frac{n}{n-1}\frac{n-1}{n}D\left(X\right)=D\left(X\right).$$

Следовательно, ***несмещенной оценкой генеральной дисперсии*** $D(X)$ ***служит исправленная выборочная дисперсия***

$$D\left(X\right)\cong s_X^2=\frac{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}_{\text{в}}\right)^2}{n-1}=\frac{\sum\limits_{j=1}^{k}n_j\left(x_j^*-\overline{x}_{\text{в}}\right)^2}{n-1}. \qquad (15)$$

Замечание 1. Из формулы $s_X^2=\dfrac{n}{n-1}D_{\text{в}}$ следует, что $s_X^2>D_{\text{в}}$.

Исправленная выборочная дисперсия (15) может быть представлена в более удобном для вычислений виде

$$s_X^2=\frac{1}{n-1}\sum_{i=1}^{n}x_i^2-\frac{n}{n-1}\overline{x}_{\text{в}}^2=\frac{1}{n-1}\sum_{j=1}^{k}n_j\left(x_j^*\right)^2-\frac{n}{n-1}\overline{x}_{\text{в}}^2, \qquad (16)$$

что справедливо и для интервального статистического ряда, где $x_1^*,\ x_2^*,\ \ldots,\ x_k^*$ – середины соответствующих интервалов.

Для оценки среднего квадратического отклонения генеральной совокупности используют *исправленное среднее квадратическое отклонение*

$$\sigma\left(X\right)\cong s_X=\sqrt{\frac{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}_{\text{в}}\right)^2}{n-1}}.$$

• **Задача 2.** Докажите справедливость представления исправленной выборочной дисперсии s_X^2 формулой (16).

• **Задача 3*.** Докажите, что исправленное среднее квадратическое отклонение s_X не является в общем случае несмещенной оценкой среднего квадратического отклонения $\sigma(X)$ генеральной совокупности.

Из сопоставления оценок дисперсии (15) и (11) видно, что они отличаются лишь знаменателями. Очевидно, что при достаточно большом объеме выборки n выборочная дисперсия (11) и исправленная выборочная дисперсия (15) различаются незначительно, так как множитель $\dfrac{n}{n-1}$ близок к 1, и в этом случае $s_X^2 \approx D_{\text{в}}(X)$. Исправленная выборочная дисперсия (15) используется на практике при небольшом объеме выборки $(n < 30)$.

Для вычислений $D_{\text{в}}(X)$ зачастую более удобной, чем (9), является формула

$$D_{\text{в}}(X) = \overline{x}_{\text{в}}^2 - (\overline{x}_{\text{в}})^2 = \frac{\sum\limits_{i=1}^{n} x_i^2}{n} - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n^2} = \frac{\sum\limits_{j=1}^{k} n_j \cdot (x_j^*)^2}{n} - \frac{\left(\sum\limits_{j=1}^{k} n_j x_j^*\right)^2}{n^2}.$$

Пример 2. Найти выборочную дисперсию и исправленную выборочную дисперсию по данному распределению выборки:

x_i	1	2	3	4
n_i	20	15	10	5

$n = 50$.

Решение. Вычислим сначала $\overline{x}_{\text{в}} = \dfrac{20\cdot 1 + 15\cdot 2 + 10\cdot 3 + 5\cdot 4}{50} = 2$ и $\overline{x}_{\text{в}}^2 = \dfrac{20\cdot 1 + 15\cdot 4 + 10\cdot 9 + 5\cdot 16}{50} = 5.$

Тогда $D_{\text{в}}(X) = \overline{x}_{\text{в}}^2 - (\overline{x}_{\text{в}})^2 = 1$. Чтобы найти s_X^2, воспользуемся формулой (16) при $n = 50$, $k = 4$: $s_X^2 = \dfrac{1}{49}\sum\limits_{j=1}^{4} n_j x_j^2 - \dfrac{50}{49}\overline{x}_{\text{в}}^2$. Следовательно,

$$s_X^2 = \frac{1}{49}\left(20\cdot 1^2 + 15\cdot 2^2 + 10\cdot 3^2 + 5\cdot 4^2\right) - \frac{50}{49}\cdot 2^2 = \frac{50}{49}. \quad \blacktriangle$$

Замечание 2. Если первоначальные варианты x_i – большие числа, то целесообразно перейти к *условным вариантам* $u_i = x_i - C$, т. е. вычесть из всех вариант одно и то же число C, равное $\overline{x}_{\text{в}}$ или близкое к нему. При таком преобразовании $M(U) = M(X) - M(C) = \overline{x}_{\text{в}} - C$, а дисперсия не изменится ($D_{\text{в}}(U) = D_{\text{в}}(X)$ в силу свойства $D(C) = 0$). Тогда

$$\overline{x}_{\text{в}} = C + \frac{\sum\limits_{j=1}^{k} n_j u_j^*}{n} = C + \overline{u}_{\text{в}};$$

$$D_{\text{в}}(X) = D_{\text{в}}(U) = \overline{u}^2 - (\overline{u})^2 = \frac{\sum\limits_{j=1}^{k} n_j (u_j^*)^2}{n} - \frac{\left(\sum\limits_{j=1}^{k} n_j u_j^*\right)^2}{n^2} = \frac{1}{n}\sum_{j=1}^{k} n_j (u_j^*)^2 - \overline{u}_{\text{в}}^2,$$

а исправленная выборочная дисперсия s_X^2 через условные варианты u_i выражается в виде

$$s_X^2 = s_U^2 = \frac{1}{n-1}\sum_{i=1}^{n} u_i^2 - \frac{n}{n-1}\overline{u}_{\text{в}}^2 = \frac{1}{n-1}\sum_{j=1}^{k} n_j (u_j^*)^2 - \frac{n}{n-1}\overline{u}_{\text{в}}^2 .$$

Пример 3. По данному распределению выборки

x_i^*	1250	1270	1280
n_i	2	5	3

$n = 10$

найти выборочное среднее и исправленную выборочную дисперсию.

Решение. Поскольку первоначальные варианты – большие числа, то перейдем к условным вариантам. Выберем $C = 1270$, тогда условные варианты u_i вычисляются по формулам $u_i = x_i - 1270$. Распределение выборки в условных вариантах принимает вид

u_i^*	–20	0	10
n_i	2	5	3

Тогда $\overline{x}_{\text{в}} = C + \overline{u}_{\text{в}} = 1270 + \dfrac{2\cdot(-20) + 5\cdot 0 + 3\cdot 10}{10} = 1269$, $\overline{u}_{\text{в}} = -1$, а

несмещенная оценка дисперсии равна $s_X^2 = s_U^2 = \dfrac{1}{9}\sum\limits_{j=1}^{3} n_j (u_j^*)^2 - \dfrac{10}{9}\overline{u}_{\text{в}}^{\,2} =$

$$= \frac{1}{9}\left(2\cdot(-20)^2 + 5\cdot 0^2 + 3\cdot 10^2\right) - \frac{10}{9}\cdot(-1)^2 = \frac{1090}{9}. \blacktriangle$$

4. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ (ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ)

В подразд. 3 был рассмотрен вопрос об оценке неизвестного параметра θ одним числом θ^*, т. е. о *точечной* оценке. В ряде задач требуется не только найти для параметра θ подходящее численное значение, но и оценить его *точность* и *надежность*. Требуется знать, к каким ошибкам может привести замена θ его точечной оценкой θ^* и с какой степенью уверенности можно ожидать, что эта оценка не выйдет за известные пределы. Ведь точечная оценка θ^* является лишь приближенным значением неизвестного параметра θ даже в том случае, когда она является несмещенной (т. е. в среднем совпадает с θ), состоятельной (с ростом *n* объема выборки стремится по вероятности к θ) и эффективной (обладает наименьшим рассеиванием случайных отклонений от θ). Такого рода задачи особенно актуальны при малом числе наблюдений, когда точечная оценка θ^* в значительной мере случайна и замена θ на θ^* может привести к существенным ошибкам. Для определения *точности* и *надежности* оценки θ^* в МС вводят понятия *доверительного интервала* и *доверительной вероятности*. В приложениях часто предполагается, что СВ X распределена по нормальному закону с плотностью вероятности

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2y^2}}; \; a = M(X); \quad \sigma = \sqrt{D(X)}.$$

Возникает задача оценки параметров a и σ или одного из них, если известны наблюдаемые (выборочные) значения $x_1, x_2, \ldots, x_n$ СВ X. При этом возможны следующие случаи:

1) σ известно, требуется оценить параметр a;
2) σ неизвестно, оценивается параметр a;
3) a известно, требуется оценить параметр σ;
4) a и σ неизвестны, оцениваются параметры a и σ.

Пусть для параметра θ из опыта получена несмещенная оценка θ^*. Оценим возможную при этом ошибку. Назначим некоторую достаточно большую вероятность γ (γ = 0,9; 0,95; 0,99), такую, что событие с вероятностью γ можно считать практически достоверным. Найдем такое значение ε, ε > 0, для которого вероятность отклонения оценки на величину, не превышающую ε, равна γ:

$$P\left(\left|\theta^* - \theta\right| < \varepsilon\right) = \gamma. \tag{17}$$

Тогда диапазон практически возможных значений ошибки, возникающей при замене θ на θ^*, будет равен ±ε, при этом превосходящие ε по абсолютной величине ошибки будут появляться с малой вероятностью $\alpha = 1 - \gamma$, называемой *уровнем значимости*. На практике [9, 15, 19] обычно используют следующие уровни значимости: α = 0,1; 0,05; 0,01.

Перепишем уравнение (17) в виде

$$P(\theta^* - \varepsilon < \theta < \theta^* + \varepsilon) = \gamma\ . \tag{18}$$

Равенство (18) означает, что с вероятностью γ неизвестное значение параметра θ находится в интервале I_γ, равном

$$I_\gamma = (\theta^* - \varepsilon; \theta^* + \varepsilon). \tag{19}$$

Вероятность γ принято называть *доверительной вероятностью* (*надежностью*), а интервал I_γ – *доверительным интервалом*.

Интервал I_γ является случайным, так как случайным является центр θ^* – оценка параметра θ – интервала I_γ . Случайной является и его длина, равная 2ε, поскольку ε, как правило, вычисляется на основании опытных данных. Поэтому в (18) величину γ лучше толковать не как вероятность γ попадания точки θ в интервал I_γ, а как вероятность того, что случайный интервал I_γ «накроет» точку θ (рис. 8).

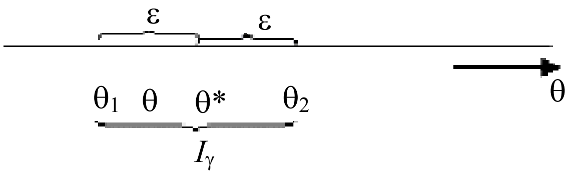


Рис. 8. Доверительный интервал I_γ

На рисунке θ^* – центр доверительного интервала, $\theta_1 = \theta^* - \varepsilon$, $\theta_2 = \theta^* + \varepsilon$ – его границы. Границы θ_1, θ_2 интервала I_γ и его величина находятся по выборочным данным и потому являются случайными величинами в отличие от оцениваемого параметра θ – величины неслучайной. Поэтому и целесообразно говорить, что доверительный интервал $I_\gamma = (\theta_1; \theta_2)$ «накрывает» истинное значение параметра θ с надежностью γ, а не «содержит» его.

Определение 5. Интервал $(\theta_1; \theta_2)$ называют ***доверительным интервалом*** для оценки параметра θ при заданной *доверительной вероятности* γ (или при заданном *уровне значимости* $\alpha = 1 - \gamma$), если он с вероятностью γ «накрывает» оцениваемый параметр θ, т. е.

$$P(\theta \in (\theta_1;\ \theta_2)) = P(\theta_1 < \theta < \theta_2) = \gamma .$$

Границы интервала θ_1 и θ_2 называют *доверительными границами*. Доверительный интервал можно рассматривать как интервал значений параметра θ, совместимых с опытными данными и не противоречащих им. Метод доверительных интервалов был разработан американским математиком Ю. Нейманом, который использовал идеи английского статистика Р. Фишера.

Рассмотрим вопрос о нахождении доверительных границ θ_1 и θ_2. Пусть для параметра θ имеется несмещенная оценка θ^*. Если бы был известен закон распределения величины θ^*, задача нахождения доверительного интервала была бы простой: достаточно было бы найти такое значение ε, для которого выполнено равенство (17). Сложность, однако, состоит в том, что закон распределения оценки θ^* зависит от закона распределения СВ X следовательно, от его неизвестных параметров, в частности от параметра θ.

Нахождение доверительных интервалов излагается ниже для параметров нормального распределения, наиболее широко применяемого на практике. С другими методами построения как интервальных, так и точечных оценок можно ознакомиться в [9, 15].

4.1. Доверительные интервалы для оценки неизвестного математического ожидания нормально распределенной СВ при известном среднеквадратическом отклонении

Пусть количественный признак X генеральной совокупности распределен по нормальному закону с параметрами a, σ, т. е. $X \in N(a,\sigma)$, причем среднее квадратическое отклонение $\sigma(X)$ известно. Оценим неизвестное математическое ожидание a теоретического распределения по выборочному среднему $\bar{x}_{\text{в}}$, т. е. по $\bar{x}_{\text{в}}$ найдем *интервал*, накрывающий параметр a с *надежностью* γ. Будем рассматривать $\bar{x}_{\text{в}}$ как СВ $\bar{X}_{\text{в}}$, $\bar{X}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n} X_i$, где $M(X_i) = a$, $\sigma(X_i) = \sigma$, $i = \overline{1,\ n}$.

Известно [9, 15], что если СВ X распределена нормально, то выборочное среднее $\bar{X}_{\text{в}} = \bar{x}_{\text{в}}$, найденное по независимым в совокупности наблюдениям, также распределено нормально и распределение $\bar{X}_{\text{в}}$ имеет следующие числовые характеристики:

$$M(\bar{X}_{\text{в}}) = a,\ D(\bar{X}_{\text{в}}) = \frac{\text{у}^2}{n},\ \text{у}(\bar{X}_{\text{в}}) = \sqrt{D(\bar{X}_{\text{в}})} = \frac{\text{у}}{\sqrt{n}}.$$

Для нахождения доверительного интервала для оценки неизвестного математического ожидания a потребуем выполнения соотношения $P\left(\left|\bar{X}_{\text{в}} - a\right| < \delta\right) = \gamma$, где γ – заданная надежность. Поскольку для нормально распределенной СВ X вероятность ее попадания в интервал выражается формулой $P(\left|X - a\right| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$, где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ – *функция Лапласа*, то, сделав замену $X \to \bar{X}_{\text{в}}$, $\sigma \to \sigma(\bar{X}_{\text{в}}) = \frac{\sigma}{\sqrt{n}}$, получим $P(\mid \bar{X}_{\text{в}} - a \mid\ <\ \text{д}) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) = 2\Phi(t_{\text{г}})$, где $t_{\text{г}} = \frac{\delta\sqrt{n}}{\sigma}$. Следовательно, $\delta = \frac{t_\gamma \sigma}{\sqrt{n}}$ – *точность оценки*, и можно написать, что $P\left(\left|\overline{X}_{\text{в}} - a\right| < \dfrac{t_{\text{г}}\text{у}}{\sqrt{n}}\right) = 2\Phi(t_{\text{г}})$. Так как вероятность P априори задана и равна γ, то, заменив $\overline{X}_{\text{в}}$ на $\overline{x}_{\text{в}}$, получим

$$P\left(\overline{x}_{\text{в}} - \frac{t_{\text{г}}\text{у}}{\sqrt{n}} < a < \overline{x}_{\text{в}} + \frac{t_{\text{г}}\text{у}}{\sqrt{n}}\right) = 2\Phi\left(t_{\text{г}}\right) = \gamma. \qquad (20)$$

Таким образом, можно утверждать, что *интервал* $\left(\overline{x}_{\text{в}} - \dfrac{t_{\text{г}}\text{у}}{\sqrt{n}};\ \overline{x}_{\text{в}} + \dfrac{t_{\text{г}}\text{у}}{\sqrt{n}}\right)$ *с точностью* $\text{д} = \dfrac{t_{\text{г}}\text{у}}{\sqrt{n}}$ *с надежностью* (*доверительной вероятностью*) γ *накрывает неизвестное математическое ожидание а нормально распределенной СВ с известным средним квадратическим отклонением* σ. Число t_γ, входящее в определение доверительного интервала, находится из соотношения $\Phi(t_\gamma) = \dfrac{\gamma}{2}$, при этом t_γ называется *квантилем* нормального распределения, соответствующим γ. Квантиль t_γ находят по таблице значений функции Лапласа [9, 10, 15, 19] по заданной надежности γ.

Замечание 3. Из формулы $\text{д} = \dfrac{t_{\text{Г}}\text{у}}{\sqrt{n}}$ точности оценки математического ожидания $\left|\overline{x}_{\text{в}} - a\right| < \delta$ (такую оценку называют *классической*) следует, что *при возрастании объема выборки n* величина $\delta \to 0$ и, следовательно, *точность оценки увеличивается.*

Замечание 4. Из формулы (20) надежности оценки $\gamma = 2\Phi(t_\gamma)$ следует, что поскольку $\Phi(t_\gamma)$ – возрастающая функция, то увеличение надежности γ влечет возрастание t_γ, а следовательно, и возрастание числа δ. Таким образом, *увеличение надежности классической оценки влечет уменьшение ее точности.*

Пример 4. СВ X распределена нормально с $\sigma = 3$. Найти доверительный интервал для оценки неизвестного математического ожидания a по выборочному среднему $\overline{x}_{\text{в}}$, если объем выборки $n = 49$ и задана надежность оценки $\gamma = 0{,}95$.

Решение. По заданному $\gamma = 0{,}95$ вычислим $\Phi\left(t_{\text{г}}\right) = \dfrac{\gamma}{2} = \dfrac{0{,}95}{2} = 0{,}475.$

По таблице значений функции Лапласа [9, 10, 15, 19] находим $t_\gamma = 1{,}96$; следовательно, точность оценки $\delta = \dfrac{t_{\text{г}}\text{у}}{\sqrt{n}} = \dfrac{1{,}96 \cdot 3}{7} = 0{,}94$. Тогда доверительный интервал $\left(\overline{x}_{\text{в}} - \delta; \overline{x}_{\text{в}} + \delta\right)$ имеет вид

$(\overline{x}_{\text{в}} - 0{,}94;\ \overline{x}_{\text{в}} + 0{,}94)$. Таким образом, *a* с вероятностью 0,95 попадет в интервал $(\overline{x}_{\text{в}} - 0{,}94;\ \overline{x}_{\text{в}} + 0{,}94)$. ▲

Смысл заданной надежности: надежность $\gamma = 0{,}95$ означает, что при достаточно больших объемах выборок в среднем в 95% из них доверительные интервалы накрывают неизвестный параметр *a*, в среднем в 5% случаев параметр *a* может выйти за границы доверительных интервалов.

Замечание 5. Если необходимо оценить неизвестное математическое ожидание *a* не только с надежностью γ, но и с наперед заданной точностью δ, то минимальный объем *n* выборки, обеспечивающий эту точность, можно найти из равенства $\text{д} = \frac{t_\gamma \text{у}}{\sqrt{n}}$: $n = \frac{t_\gamma^2 \sigma^2}{\delta^2}$.

• **Задача 4.** Найдите минимальный объем выборки, на основании которой можно было бы оценить математическое ожидание СВ *X* с ошибкой, не превышающей 0,2, и надежностью 0,98, если известно, что *X* имеет нормальное распределение с σ = 4.

• **Задача 5.** В течение продолжительного срока при анализе данного материала на содержание железа установлено стандартное отклонение 0,12%. Найдите с надежностью $\gamma = 0{,}95$ доверительный интервал для истинного содержания железа в образце, если по результатам шести анализов среднее содержание железа составило 32,56%. Объясните смысл полученного результата.

4.2. Доверительные интервалы для оценки неизвестного математического ожидания нормально распределенной СВ при неизвестном среднеквадратическом отклонении

Наряду с нормальным распределением, которое занимает центральное место в теории и практике вероятностно-статистических исследований, в МС рассматриваются и другие важные распределения, которые используются для построения статистических критериев, оценок, доверительных интервалов и в других целях. К таким распределениям относятся, в частности, χ^2-распределение («хи квадрат»-распределение) и *t-распределение* (или *распределение Стьюдента*).

4.2.1. Распределение χ^2. СВ χ^2 используется при интервальном оценивании параметров распределений, при статистической проверке гипотез и т. д.

Нормальный закон *N*(*a*; σ) распределения СВ *X* с параметрами *a* = 0, σ = 1 (т. е. $X \in N(0;1)$) называется *стандартным* (или *нормированным*), а соответствующая нормальная кривая – *стандартной* (или *нормированной*) [9, 15].

Пусть X_i, $i=\overline{1,k}$ – независимые в совокупности СВ, распределенные по *стандартному нормальному закону*.

Определение 6. Распределением χ^2 с *k степенями свободы* называется распределение суммы квадратов k независимых в совокупности случайных величин $X_i \in N(0;1), i=\overline{1,k}$ распределенных по стандартному нормальному закону:

$$\chi^2 = \sum_{i=1}^{k} X_i^2. \qquad (21)$$

Если СВ $X_i, i=\overline{1,k}$, связаны некоторым линейным соотношением, например $\sum_{i=1}^{n} X_i = n\overline{x}_в$, то число степеней свободы $k = n - 1$. Плотность вероятности распределения χ^2 равна [9, 15]

$$f_{\chi^2}(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{1}{2^{k/2}\Gamma\left(\dfrac{k}{2}\right)} e^{-x/2} \cdot x^{k/2-1}, & x > 0, \end{cases}$$

где $\Gamma(x) = \int\limits_0^{+\infty} t^{x-1} e^t dt$ – *гамма-функция Эйлера*. В частности, можно показать, что $\Gamma(n+1) = n!$ для целых положительных значений n. Распределение χ^2 полностью определяется величиной k, называемой *числом степеней свободы.* Кривые χ^2-распределения для различных значений числа степеней свободы k асимметричны и расположены, в силу определения $f_{\chi^2}(x)$, в правой полуплоскости. Таким образом, распределение χ^2 определяется одним параметром – числом степеней свободы k. С увеличением числа степеней свободы k график распределения χ^2 становится похожим на график нормального распределения, а при $k > 30$ распределение СВ $X = 2\chi^2 - \sqrt{2k}$ *близко* к стандартному нормальному закону $N(0;\ 1)$ (график χ^2-распределения СВ $X = 2\chi^2 - \sqrt{2k}$ имеет вид нормального закона распределения). Таким образом, с увеличением числа степеней свободы k распределение χ^2 *медленно* приближается к нормальному распределению.

Важным свойством распределения χ^2 является его *воспроизводимость* по параметру. Это означает, что сумма k независимых СВ, распределенных по закону χ^2, также распределена по этому закону с числом степеней свободы, равным сумме степеней свободы слагаемых.

4.2.2. Распределение Стьюдента. Распределение Стьюдента широко используется в МС, в частности при построении доверительного интервала для математического ожидания, проверке гипотезы о равенстве математических ожиданий двух нормальных СВ и др. Оно впервые описано в 1908 г. английским статистиком В. Госсетом, опубликовавшим свою научную работу под псевдонимом Стьюдент (Student).

Пусть $X \in N(0;1)$ – СВ, распределенная по стандартному нормальному закону, v – независимая от X СВ, распределенная по закону χ^2 с k степенями свободы. Тогда СВ

$$T = \frac{X}{\sqrt{\nu/k}} \tag{22}$$

имеет распределение, называемое ***t*-*распределением*** (или ***распределением Стьюдента***) с k степенями свободы. Таким образом, *отношение нормированной нормально распределенной СВ X к квадратному корню от независимой СВ ν, распределенной по закону χ^2 с k степенями свободы, деленной на k, распределено по закону Стьюдента с k степенями свободы.*

Плотность $S(t, n-1)$ распределения Стьюдента с $k = n - 1$ степенями свободы имеет вид

$$S(t, n-1) = B_n \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}, \ t \in R, \tag{23}$$

где n – объем выборки; коэффициент B_n выражается через *гамма-функцию Эйлера* Γ: $B_n = \dfrac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)}$. В силу второго замечательного предела в (23) имеем $\left(1 + \dfrac{t^2}{n-1}\right)^{-\frac{n}{2}} \xrightarrow[n\to+\infty]{} e^{-\frac{t^2}{2}}$. Отсюда с учетом четности функции $S(t, n-1)$ по переменной t следует, что кривая распределения Стьюдента симметрична относительно оси ординат и при больших n мало отличается от кривой стандартного нормального распределения. Однако по сравнению с ней кривая распределения Стьюдента является более пологой [9, 15]. Можно считать, что при $n > 30$ распределение Стьюдента является *приближенно нормальным*. Таким образом, *на практике при объеме выборки $n > 30$ вместо распределения Стьюдента иногда пользуются нормальным распределением.*

Из (23) также видно, что распределение Стьюдента определяется параметром n – объемом выборки (или, что то же самое, числом степеней свободы $k = n - 1$) и не зависит от неизвестных параметров a и σ. Такая особенность распределения Стьюдента является его большим достоинством. С возрастанием числа степеней свободы k (а следовательно, и n) распределение Стьюдента *быстро* приближается к нормальному. Для малых выборок ($n \le 30$) замена распределения Стьюдента нормальным распределением нежелательна, поскольку может привести к неоправданному сужению доверительных интервалов.

4.2.3. Доверительные интервалы. В качестве применения распределения Стьюдента рассмотрим задачу построения доверительных интервалов для оценки неизвестного математического ожидания a нормально распределенной СВ $X \in N(a, \sigma)$ при неизвестном среднеквадратическом отклонении σ.

Воспользоваться результатами п. 4.1 не представляется возможным, так как параметр σ в данном случае неизвестен. Для решения задачи из генеральной совокупности извлечем выборку $x_1, x_2, \ldots, x_n$ объема n и по ней найдем выборочное среднее $\bar{x}_{\text{в}}$ и исправленное среднее квадратическое отклонение s. Рассмотрим статистику

$$T = \frac{(\bar{x}_{\text{в}} - a)\sqrt{n}}{s} = \frac{(\bar{x}_{\text{в}} - a)\sqrt{n-1}}{\sqrt{D_{\text{в}}}}, \tag{24}$$

где n – объем выборки; $D_{\text{в}}$ – выборочная дисперсия. В МС доказывается [10, 15], что СВ (24) имеет распределение Стьюдента с $k = n - 1$ степенями свободы.

В силу четности функции $S(t,\ n - 1)$ плотности распределения Стьюдента из (23) по переменной t следует, что вероятность осуществления неравенства $\left|\frac{(\bar{x}_{\text{в}} - a)\sqrt{n}}{s}\right| < t_{\text{г}}$ определяется по формуле

$$P\left(\left|\frac{(\bar{x}_{\text{в}} - a)\sqrt{n}}{s}\right| < t_{\text{г}}\right) = 2\int_0^{t_{\text{г}}} S(t,\ n-1)\,dt = \gamma,$$

откуда следует

$$P\left(\bar{x}_{\text{в}} - \frac{t_{\text{г}} s}{\sqrt{n}} < a < \bar{x}_{\text{в}} + \frac{t_{\text{г}} s}{\sqrt{n}}\right) = \gamma.$$

Таким образом, с помощью распределения Стьюдента найден *доверительный интервал* $\left(\bar{x}_{\text{в}} - \frac{t_{\text{г}} s}{\sqrt{n}};\ \bar{x}_{\text{в}} + \frac{t_{\text{г}} s}{\sqrt{n}}\right)$, накрывающий неизвестный параметр a с надежностью γ. По табл. 3 приложения [10] по заданным n и γ находят квантиль t_γ ($t_\gamma = t(\gamma, n)$), а следовательно, и искомый доверительный интервал.

Пример 5. Количественный признак X генеральной совокупности распределен нормально. По выборке объема $n = 16$ найдены $\overline{x}_в = 20{,}2$ и исправленное среднее квадратическое отклонение $s = 0{,}8$. Оценить неизвестное математическое ожидание a при помощи доверительного интервала с надежностью $\gamma = 0{,}95$.

Решение. Для нахождения доверительного интервала следует по табл. 3 приложения [10] по заданному объему выборки $n = 16$ и доверительной вероятности $\gamma = 0{,}95$ найти $t_\gamma = 2{,}13$. Тогда границы доверительного интервала имеют вид:

$$\overline{x}_в - t_\gamma \frac{s}{\sqrt{n}} = 20{,}2 - \frac{2{,}13 \cdot 0{,}8}{4} = 19{,}774;$$

$$\overline{x}_в + t_\gamma \frac{s}{\sqrt{n}} = 20{,}2 + \frac{2{,}13 \cdot 0{,}8}{4} = 20{,}626.$$

Таким образом, доверительный интервал (19,774; 20,626) с надежностью $\gamma = 0{,}95$ накрывает неизвестный параметр a генеральной совокупности. ▲

● **Задача 6.** Известны результаты измерения роста (см) случайно отобранных 100 студентов:

Рост	[154; 158)	[158; 162)	[162; 166)	[166; 170)	[170; 174)	[174; 178)	[178; 182)
Число студентов	8	12	26	28	12	10	4

Найдите доверительные интервалы с надежностью 0,95 для оценки среднего роста студентов.

4.3. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения по исправленному выборочному среднему квадратическому отклонению

Пусть количественный признак X генеральной совокупности распределен нормально. Требуется оценить неизвестное среднее квадратическое отклонение σ по исправленному выборочному среднему квадратическому отклонению s.

Задача. Найти доверительный интервал, накрывающий параметр σ с заданной надежностью γ.

Для решения поставленной задачи потребуем, чтобы вероятность отклонения среднего квадратического отклонения σ от исправленного выборочного среднего квадратического отклонения s была равна заданной доверительной вероятности γ: $P\left(\left|\sigma - s\right| < \delta\right) = \gamma$, или, раскрывая модуль, чтобы $P\left(s - \delta < \sigma < s + \delta\right) = \gamma$. Преобразуем двойное неравенство $s - \delta < y < s + \delta$ к виду $s\left(1 - \frac{\delta}{s}\right) < \sigma < s\left(1 + \frac{\delta}{s}\right)$. Положив $\frac{\delta}{s} = q$, получим

$$s\left(1 - q\right) < \sigma < s\left(1 + q\right). \tag{25}$$

Таким образом, для нахождения доверительных интервалов, накрывающих параметр σ с заданной надежностью γ, необходимо, в силу (25), найти q. С этой целью введем в рассмотрение СВ χ: $\chi = \frac{s\sqrt{n-1}}{\sigma}$, где n – объем выборки. Известно [9, 10, 15], что СВ $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, где s^2 – исправленная выборочная дисперсия, распределена по закону χ^2 с $k = n - 1$ степенями свободы. Так как плотность распределения СВ χ имеет вид [9, 10, 15]

$$\rho(\chi, n) = \frac{ч^{n-2} e^{-\frac{ч^2}{2}}}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-1}{2}\right)},$$

то $P(\chi_1 < \chi < \chi_2) = \gamma$ (т. е. вероятность того, что неравенство $\chi_1 < \chi < \chi_2$ имеет заданную надежность γ) определяется формулой $\int_{ч_1}^{ч_2} \rho(\chi, n) d\chi = \gamma$. При $q < 1$ неравенство (25) можно переписать следующим образом: $\frac{1}{s(1+q)} < \frac{1}{у} < \frac{1}{s(1-q)}$. Умножив все члены неравенства на положительное число $s\sqrt{n-1}$ и приняв во внимание, что $\frac{s\sqrt{n-1}}{у} = \chi$, получим $\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}$. Вероятность того, что будет осуществлено это неравенство, а следовательно, и равносильное ему неравенство (25), выражается формулой

$$\int\limits_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} \rho(\chi, n) d\chi = \gamma.$$

Таким образом, для нахождения q получено интегральное уравнение, из которого по заданным n и γ можно найти q. На практике для отыскания значений $q = q(\gamma, n)$ пользуются таблицами приложений (см. табл. 4 [10]). Найдя по таблице q и вычислив по выборке s, получим доверительный интервал (25), накрывающий среднее квадратическое отклонение σ с заданной надежностью γ.

Замечание 6. При нахождении доверительного интервала (25) предполагалось, что $q < 1$. Если $q > 1$, то $s(1 - q) < 0$, $s(1 + q) > 0$, и так как $\sigma > 0$, то из неравенства $s(1 - q) < \sigma < s(1 + q)$ следует $0 < \sigma < s(1 + q)$. Проведя для этого неравенства преобразования, аналогичные случаю $q < 1$, получим $\frac{\sqrt{n-1}}{1+q} < \chi < \infty$. Таким образом, значения $q > 1$ можно найти из уравнения $\int\limits_{\frac{\sqrt{n-1}}{1+q}}^{+\infty} \rho(\chi, n) d\chi = \gamma$ и затем в силу $0 < \sigma < s(1 + q)$ найти доверительный интервал. На практике для отыскания $q > 1$ так же, как и в случае $q < 1$, пользуются таблицами приложений (табл. 4 [10]).

Итак, ***интервальной оценкой (с надежностью γ) среднего квадратического отклонения σ нормально распределенного количественного признака X по исправленному выборочному среднему квадратическому отклонению s служат доверительные интервалы:*** а) $s(1 - q) < \sigma < s(1 + q)$ при $q < 1$; б) $0 < \sigma < s(1 + q)$ при $q > 1$, где $q = q(\gamma, n)$ находят по таблице (см., например, табл. 4 приложения [10]) по заданному объему выборки n и надежности γ.

Пример 6. По данным выборки объема $n = 16$ из генеральной совокупности найдено исправленное среднее квадратическое отклонение $s = 1$ нормально распределенного количественного признака. Найти доверительный интервал, накрывающий генеральное среднее квадратическое отклонение σ с надежностью $\gamma = 0,95$.

Решение. Для $n = 16$, $\gamma = 0,95$ по табл. 4 приложения [10] находим $q = 0,44$. Поскольку $q < 1$, то из формулы (25) имеем $1(1 - 0,44) < \sigma < 1(1 + 0,44)$, или $0,56 < \sigma < 1,44$. ▲

Пример 7. Количественный признак X генеральной совокупности распределен нормально. По выборке объема $n = 15$ найдено исправленное среднее квадратическое отклонение $s = 0,25$. Найти доверительный интервал, накрывающий генеральное среднее квадратическое отклонение у с надежностью $\gamma = 0{,}999$.

Решение. Для нахождения доверительного интервала по табл. 4 приложения [10] по данному $\gamma = 0{,}999$ и объему выборки $n = 15$ найдем $q = 1{,}15$. Так как $q > 1$, то доверительный интервал имеет вид $0 < \sigma < s(1 + q)$, т. е. $0 < \sigma < 0{,}25(1 + 1{,}15)$. Следовательно, $0 < \sigma < 0{,}5625$.▲

5. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

С теорией статистического оценивания параметров тесно связана проверка статистических гипотез. Она используется, когда необходим обоснованный вывод о преимуществах того или иного технологического процесса, того или иного способа вложения инвестиций, об уровне доходности ценных бумаг, об эффективности вносимых удобрений, о значимости построенной математической модели и т. д.

При изучении многих статистических данных необходимо знать закон распределения генеральной совокупности. Если закон распределения неизвестен, но есть основания предположить, что он имеет определенный вид (например, *A*), то выдвигают гипотезу: генеральная совокупность распределена по закону *A*. В данной гипотезе речь идет о *виде* предполагаемого закона распределения.

Возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр θ равен определенному значению θ_0, то выдвигают гипотезу: $\theta = \theta_0$. Здесь речь идет о *предполагаемой величине параметра* известного распределения. Возможны гипотезы о равенстве параметров двух или нескольких распределений, о независимости выборок и др.

Все выводы, которые делаются в МС, вообще говоря, являются *гипотетическими*, т. е. предположениями о неизвестных параметрах известных распределений, об общем виде неизвестного теоретического распределения или функции распределения изучаемой СВ. Такие гипотезы называют *статистическими гипотезами*.

Статистической называется гипотеза о виде неизвестного распределения или о параметрах известных распределений. Например, статистическими являются гипотезы: 1) генеральная совокупность распределена по равномерному закону; 2) дисперсии двух совокупностей, распределенных по нормальному закону, равны между собой. В гипотезе 1 сделано предположение о *виде* неизвестного распределения, в гипотезе 2 – *о параметрах* двух известных распределений.

Гипотеза «на Марсе есть жизнь» не является статистической, поскольку в ней не идет речь ни о виде, ни о параметрах распределения.

Различают *простые* и *сложные, параметрические* и *непараметрические* статистические гипотезы.

Статистическая гипотеза называется *простой*, если она содержит только одно предположение. *Сложной* называют гипотезу, состоящую из конечного или бесконечного числа простых гипотез. Например, гипотезы «вероятность появления события A в схеме Бернулли равна $1/3$», «закон распределения СВ – нормальный с параметрами $a = 0$, σ известно» являются *простыми* в отличие от *сложных* гипотез: «вероятность появления события A в схеме Бернулли заключена между $1/3$ и $1/2$», «закон распределения СВ не является нормальным». Гипотеза называется *параметрической*, если в ней содержится некоторое условие о значении параметра известного распределения. Гипотезу, в которой сформулированы предположения относительно вида распределения, называют *непараметрической*.

Если исследовать всю генеральную совокупность, то, естественно, можно наиболее точно установить справедливость выдвигаемой гипотезы. Однако такое исследование не всегда возможно, и суждение об истинности статистических гипотез проверяется на основании выборки.

Выдвигаемую (проверяемую) гипотезу называют *основной* или *нулевой* гипотезой H_0. Если, например, по полигону или гистограмме частот, построенным по некоторой выборке, можно выдвинуть гипотезу, что СВ распределена по нормальному закону, то может быть выдвинута гипотеза H_0: $a = a_0$, $\sigma = \sigma_0$. Одновременно с гипотезой H_0 выдвигается *альтернативная* (*конкурирующая*, *противоречащая* H_0) гипотеза H_1. Если гипотеза H_0 будет отвергнута, то принимается альтернативная гипотеза H_1. *Нулевая* H_0 и *альтернативная* H_1 гипотезы представляют собой две возможности выбора, осуществляемого в задачах проверки статистических гипотез. Например, если H_0: $\theta = \theta_0$, то альтернативная гипотеза может быть в виде H_1: $\theta \neq \theta_0$, H_1: $\theta > \theta_0$ или H_1: $\theta < \theta_0$.

• **Задача 7.** Генеральная совокупность X распределена нормально. Выдвинута нулевая гипотеза $H_0 : \begin{cases} a = a_0 \\ \sigma = \sigma_0 \end{cases}$. Постройте альтернативную гипотезу H_1.

Выдвинутая гипотеза может быть верной или неверной, в связи с чем возникает необходимость ее проверки. Поскольку проверку гипотезы осуществляют статистическими методами, такую проверку называют *статистической.* В результате статистической проверки гипотезы могут быть допущены ошибки двух родов: 1) на основании результатов опыта отвергается верная гипотеза – в этом случае допускается *ошибка первого рода*; 2) принимается ложная гипотеза, и тогда *допускается ошибка второго рода* (табл. 1).

Таблица 1

H_0	H_0 принимается	H_0 отвергается
Верна	Правильное решение	Ошибка I рода
Ошибочна	Ошибка II рода	Правильное решение

Очевидно, последствия этих ошибок могут оказаться различными. Отметим, что правильное решение может быть принято также в двух случаях: 1) гипотеза принимается, и она в действительности является правильной; 2) гипотеза отвергается, и она в действительности не верна.

Для проверки справедливости нулевой гипотезы H_0 используют специально подобранную СВ K – *статистическую оценку* (статистику), распределение которой известно.

Статистическим критерием (статистикой) называется статистический показатель (СВ) K, который служит для проверки нулевой гипотезы H_0 и характеризует количественную меру расхождения между теоретическим и эмпирическим распределением (при условии, что гипотеза H_0 верна).

Статистический критерий СВ – *статистику* – следует отличать от *критерия* проверки статистической гипотезы, т. е. от правила, согласно которому выдвинутая гипотеза принимается (нет оснований ее отвергнуть) или отвергается (противоречит данным СЭ).

Определение 7. Вероятность α *совершить ошибку I рода*, т. е. отвергнуть правильную гипотезу H_0, называется *уровнем значимости* критерия.

Обычно принимают α = 0,1; 0,05; …; 0,01; 0,001. Смысл α: при α = 0,05 в 5 случаях из 100 имеется риск допустить ошибку I рода, т. е. отвергнуть правильную гипотезу H_0. *Вероятность допустить ошибку II рода*, т. е. принять гипотезу H_0, когда она ошибочна, обозначают β. Используя терминологию статистического контроля качества продукции можно сказать, что вероятность α представляет «*риск поставщика»* (или «*риск производителя»*), связанный с вероятностью признать негодной по результатам выборочного контроля всю партию годных изделий, удовлетворяющих стандарту, а вероятность β – «*риск потребителя»*, связанный с вероятностью принять по анализу выборки негодную партию, не удовлетворяющую стандарту.

Определение 8. Вероятность $1 - \beta$ не допустить ошибку II рода, т. е. отвергнуть гипотезу H_0, когда она ошибочна, называется *мощностью критерия.*

Другими словами, мощность критерия есть вероятность того, что H_0 будет отклонена, если верна альтернативная гипотеза H_1.

Для проверки статистической гипотезы по данным выборок вычисляют частные значения входящих в критерий K величин и получают частное (*наблюдаемое*) значение критерия $K_{\text{набл}}$.

Наблюдаемым значением $K_{\text{набл}}$ называют значение критерия K, вычисленное по выборке.

После выбора определенного статистического критерия для решения вопроса о принятии или непринятии гипотезы множество его возможных значений разбивают на два непересекающихся подмножества, одно из которых называется *областью принятия гипотезы* (или *областью допустимых значений критерия*), а второе – *критической областью*.

Критической областью называется совокупность значений статистического критерия K, при которых нулевую гипотезу H_0 отвергают.

Областью принятия гипотезы (*областью допустимых значений критерия*) называется совокупность значений статистического критерия K, при которых нулевую гипотезу H_0 принимают.

Основной принцип проверки статистических гипотез. *Если наблюдаемое значение* $K_{\text{набл}}$ *статистического критерия K принадлежит критической области, то нулевая гипотеза* H_0 *отвергается в пользу альтернативной* H_1*; если оно принадлежит области принятия гипотезы, то гипотезу* H_0 *принимают (нет оснований отвергнуть).*

Критическими точками (границами) $k_{\text{кр}}$ называют точки, отделяющие критическую область от области принятия гипотезы.

В отличие от рассмотренного в подразд. 4 интервального оценивания параметров, в котором имелась лишь одна возможность совершения ошибки – получение доверительного интервала, не накрывающего оцениваемый параметр, – при проверке статистических гипотез возможна ошибка двух родов. Вероятности α и β ошибок I и II родов однозначно определяются выбором критической области. Естественным является желание сделать α и β как можно меньше. Однако эти требования являются противоречивыми, ибо при фиксированном объеме выборки уменьшение α сопряжено с увеличением β и наоборот. *Одновременное уменьшение вероятностей* α и β возможно лишь при увеличении объема выборки.

Поскольку одновременное уменьшение вероятности ошибок I и II рода невозможно, то при нахождении критических областей для данной статистики уровень значимости задают, стараясь подобрать такой критерий, чтобы вероятность ошибки II рода была наименьшей.

Различают *одностороннюю* (правостороннюю и левостороннюю) и *двустороннюю* критические области.

Правосторонней называют критическую область, определяемую неравенством $K > k_{кр}$.

Левосторонней называют критическую область, определяемую неравенством $K < k_{кр}$.

Двусторонней называют критическую область, определяемую условием $K < k_1$ либо $K > k_2$, где k_1, k_2 – левая и правая критические точки, $k_1 > k_2$.

Если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется условием $K < -k_{кр}$ либо $K > k_{кр}$, где $k_{кр} > 0$, или, что равносильно, $|K| > k_{кр}$.

Так как мощность критерия K равна $1 - \beta$ (β – вероятность ошибки второго рода), то это означает, что при выбранном уровне значимости α критическую область нужно строить так, чтобы мощность критерия была максимальной.

Как найти критическую область? Пусть $K = K(x_1, x_2, \ldots, x_n)$ – статистика, выбранная для проверки нулевой гипотезы H_0, k_0 – некоторое число, $k_0 \in R$. Найдем правостороннюю критическую область, определяемую неравенством $K > k_{кр}$. Для ее отыскания достаточно найти критическую точку $k_{кр}$. Рассмотрим вероятность $P(K > k_0)$ в предположении, что гипотеза H_0 верна. Очевидно, что с ростом k_0 вероятность $P(K > k_0)$ уменьшается. Тогда k_0 можно выбрать настолько большим, что вероятность $P(K > k_0)$ станет ничтожно малой. Другими словами, при заданном уровне значимости α можно определить критическое значение $k_{кр}$ из неравенства $P(K > k_{кр}) = \alpha$.

Критическую точку $k_{кр}$ ищут из следующего требования: при условии справедливости нулевой гипотезы H_0 вероятность того, что критерий K примет значение, большее $k_{кр}$, должна быть равна принятому уровню значимости α: $P(K > k_{кр}) = \alpha$. Для каждого из известных статистических критериев (Стьюдента, Пирсона χ^2, Колмогорова, Фишера – Снедекора, Кочрена и др.) имеются соответствующие таблицы, по которым находят $k_{кр}$, удовлетворяющее этим требованиям. После нахождения $k_{кр}$ по данным выборок вычисляют реализовавшееся (наблюдаемое) значение $K_{набл}$ критерия K. Если окажется, что

$K_{набл} > k_{кр}$, (т. е. реализовалось маловероятное событие), то нулевая гипотеза H_0 отвергается. Следовательно, принимается конкурирующая гипотеза H_1. Если же $K_{набл} \leq k_{кр}$, то в этом случае нет оснований отвергнуть выдвинутую гипотезу H_0. Следовательно, гипотеза H_0 принимается. Другими словами, *выдвинутая статистическая гипотеза согласуется с результатами эксперимента* (выборочными данными).

Левосторонняя критическая область определяется неравенством $K < k_{кр}$. *Критическую точку $k_{кр}$ находят из следующего требования: при условии справедливости нулевой гипотезы H_0 вероятность того, что критерий K примет значение, меньшее $k_{кр}$, должна быть равна принятому уровню значимости* α: $P(K < k_{кр}) = \alpha$.

Двусторонняя критическая область определяется условием $K < k_1$ или $K > k_2$, где $k_2 > k_1$. *Критические точки k_1, k_2 находят из следующего требовани: при условии справедливости нулевой гипотезы H_0 сумма вероятностей того, что критерий K примет значение, меньшее k_1 или большее k_2, должна быть равна принятому уровню значимости* α:

$$P(K < k_1) + P(K > k_2) = \alpha.$$

Если плотность распределения статистики K симметрична относительно нуля и (для увеличения его мощности) выбраны симметричные относительно нуля точки $-k_{кр}$ и $k_{кр}$, $k_{кр} > 0$, то $P(K < -k_{кр}) =$ $= P(K > k_{кр})$, и из $P(K < k_1) + P(K > k_2) = \alpha$ следует $P(K > k_{кр}) = \alpha / 2$.

Это соотношение и определяет нахождение критических точек двусторонней критической области. Односторонний критерий более «жестко» отвергает нулевую гипотезу H_0.

Отметим, что *принцип проверки статистической гипотезы не дает логического доказательства ее истинности или ложности. Принятие гипотезы H_0 следует расценивать не как установленный, абсолютно верный факт, а лишь как достаточно правдоподобное, не противоречащее СЭ утверждение.*

Если проверка статистических гипотез основана на предположении о неизвестных параметрах известного закона распределения генеральной совокупности, то критерии проверки таких гипотез называют *параметрическими*. Если закон распределения генеральной совокупности неизвестен, то соответствующие критерии называются *непараметрическими*. Понятно, что непараметрические критерии обладают значительно меньшей мощностью, чем параметрические. Отсюда следует, что для сохранения той же мощности при использовании непараметрического критерия по сравнению с параметрическим необходимо иметь значительно больший объем наблюдений.

Наиболее распространенным критерием проверки статистических гипотез о виде распределения генеральной совокупности (т. е. непараметрическим критерием) является *критерий Пирсона* χ^2.

6. ПРОВЕРКА ГИПОТЕЗ О СРЕДНЕМ ЗНАЧЕНИИ НОРМАЛЬНО РАСПРЕДЕЛЕННОЙ СВ ПРИ ИЗВЕСТНОЙ И НЕИЗВЕСТНОЙ ДИСПЕРСИИ

Пусть имеется генеральная совокупность X, распределенная по нормальному закону с *известной дисперсией* $D(X) = \sigma^2$ (т. е. σ известно). Генеральное среднее a неизвестно, но есть основания предполагать, что оно равно гипотетическому значению a_0. Например, если X – совокупность размеров x_i партии деталей, изготавливаемых станком-автоматом, то можно выдвинуть гипотезу H_0; генеральное среднее a этих размеров равна проектному размеру a_0. Для проверки этой гипотезы делают выборку, находят $\bar{x}_в$ и устанавливают, *значимо* (т. е. неслучайно) или *незначимо* (т. е. случайно) различаются $\bar{x}_в$ и a_0. Если различие окажется незначимым, то станок в среднем обеспечивает проектный размер; если же различие значимое, то станок требует дополнительной наладки.

Из нормальной генеральной совокупности извлечем выборку $x_1, \ldots, x_n$ объема n, по которой найдем $\bar{x}_в$. При этом дисперсия σ^2 известна. Будем рассматривать $x_1, \ldots, x_n$ как независимые, одинаково распределенные СВ $X_1, \ldots, X_n$. Следовательно, они имеют одинаковые нормальные распределения и одинаковые числовые характеристики (математическое ожидание a, дисперсию σ^2, среднее квадратическое отклонение σ).

Задача. По известному $\bar{x}_в$ при заданном уровне значимости α проверить гипотезу H_0: $a = a_0$ о равенстве генеральной средней a гипотетическому значению a_0.

Решение. Поскольку $\bar{x}_в$ является несмещенной оценкой генеральной средней (см. подразд. 3), т. е. $M(\bar{X}_в) = a$, то гипотезу H_0: $a = a_0$ можно записать в виде $H_0: M(\bar{X}_в) = a_0$. Таким образом, требуется проверить, что математическое ожидание выборочной средней $\bar{X}_в$ равно гипотетической генеральной средней a_0, т. е. проверить, *значимо* или *незначимо* различаются выборочная $\bar{X}_в$ и генеральная a_0 средние.

В качестве статистики, используемой в *критерии проверки* гипотезы H_0, примем СВ $U = \dfrac{\bar{X}_в - a_0}{\sigma(\bar{X}_в)}$. В силу свойства $\sigma(\bar{X}_в) = \dfrac{\sigma}{\sqrt{n}}$ одинаково распределенных независимых в совокупности СВ (см. п. 4.1) критерий проверки гипотезы H_0 принимает вид $U = \frac{\overline{X}_{\text{в}} - a_0}{\sigma}\sqrt{n}$. Случайная величина U распределена по стандартному нормальному закону $N(0;\ 1)$. Критическая область строится в зависимости от вида конкурирующей гипотезы H_1. Сформулируем правила проверки гипотезы H_0, обозначив через $U_{\text{набл}}$ значение критерия U, вычисленное по данным наблюдений.

Правило 1. Для того чтобы при заданном уровне значимости α проверить гипотезу H_0: $a = a_0$ о равенстве неизвестной генеральной средней a нормальной совокупности с известной дисперсией σ^2 гипотетическому значению a_0 при конкурирующей гипотезе H_1: $a \neq a_0$, необходимо вычислить $U_{\text{набл}} = \frac{\overline{x}_{\text{в}} - a_0}{\sigma}\sqrt{n}$ и по таблице значений функции Лапласа [9, 10, 15] найти критическую точку $u_{\text{кр}}$ *двусторонней критической области* из равенства $\Phi\left(u_{\text{кр}}\right) = \frac{1-\alpha}{2}$. Если $\left|U_{\text{набл}}\right| \le u_{\text{кр}}$ – нет оснований отвергнуть гипотезу H_0; если $\left|U_{\text{набл}}\right| > u_{\text{кр}}$ – гипотезу H_0 отвергают.

Правило 2. При конкурирующей гипотезе H_1: $a > a_0$ критическую точку $u_{\text{кр}}$ *правосторонней критической области* находят из равенства $\Phi\left(u_{\text{кр}}\right) = \frac{1-2\alpha}{2}$. Если $U_{\text{набл}} \le u_{\text{кр}}$ – нет оснований отвергнуть гипотезу H_0; если $U_{\text{набл}} > u_{\text{кр}}$ – гипотезу H_0 отвергают.

Правило 3. При конкурирующей гипотезе H_1: $a < a_0$ критическую точку $u_{\text{кр}}$ находят по правилу 2, а затем полагают границу *левосторонней критической области* $u'_{\text{кр}} = -u_{\text{кр}}$. Если $U_{\text{набл}} \ge -u_{\text{кр}}$ – нет оснований отвергнуть гипотезу H_0; если $U_{\text{набл}} < -u_{\text{кр}}$ – гипотезу H_0 отвергают.

Замечание 7. Из правила 1 следует, что если область принятия гипотезы H_0 есть интервал $-u_{\text{кр}} < U_{\text{набл}} < u_{\text{кр}}$, то область ее отклонения – $U \in (-\infty;\ -u_{\text{кр}}) \cup$ $\cup (u_{\text{кр}};\ +\infty)$.

Пример 8. Из нормальной генеральной совокупности с известным $\sigma = 0{,}49$ извлечена выборка объема $n = 49$ и по ней найдено выборочное среднее $\overline{x}_{\text{в}} = 21{,}7$. При уровне значимости $\alpha = 0{,}05$ проверить гипотезу H_1: $a = a_0 = 21$ при конкурирующей гипотезе H_1: $a > 21$.

Решение. По данным задачи найдем $U_{\text{набл}} = \frac{\overline{x}_{\text{в}} - a_0}{\sigma}\sqrt{n} = \frac{21{,}7 - 21}{0{,}49} \times$ $\times\sqrt{49} = 10$. Поскольку конкурирующая гипотеза H_1 имеет вид H_1:

$a > 21$, то критическая область – правосторонняя. По правилу 2 критическую точку $u_{кр}$ находим из равенства $\Phi\left(u_{кр}\right) = \frac{1-2\alpha}{2} = \frac{1-2\cdot 0{,}05}{2} = 0{,}45$. По таблице значений функции Лапласа находим $u_{кр} = 1{,}65$. Так как $U_{набл} = 10 > 1{,}65$, то гипотезу H_0 отвергаем. Таким образом, различие между выборочной и гипотетической генеральной средней *значимое*. ▲

● **Задача 8.** На станке изготавливаются детали с номинальным контролируемым размером 14 мм. Известно, что распределение контролируемого размера является нормальным с параметрами $\alpha = 14$ мм, $\sigma = 0{,}5$ мм. В течение суток было отобрано 30 деталей и сделаны замеры контролируемого параметра, средний размер которого оказался $\bar{x}_в = 14{,}3$ мм. Установите, можно ли с уровнем значимости 0,05 считать, что станок изготавливает детали увеличенного размера?

● **Задача 9.** Техническая норма предусматривает на выполнение некоторой операции на конвейере 30 с. Поступила жалоба от рабочих, что они затрачивают больше времени на эту операцию. ОТК произвел хронометраж у 16 рабочих и получил следующие результаты:

x_i	28	29	30	31	32	33	34
n_i	1	2	3	4	3	2	1

$\sum_{i=1}^{7} n_i = 16.$

Установите, можно ли при уровне значимости 0,02 по имеющимся хронометрическим данным отклонить предположение, что действительное среднее время выполнения операции соответствует норме?

Рассмотрим случай, когда *дисперсия* $D(X) = \sigma^2$ генеральной совокупности, распределенной по нормальному закону, *неизвестна* (т. е. σ неизвестно). Такая ситуация может возникнуть, например, в случае малых выборок. В качестве статистики для проверки гипотезы H_0 принимают (см. формулу (24) в п. 4.2.2) СВ $T = \frac{\bar{X}_в - a_0}{s}\sqrt{n}$, где s – исправленное среднее квадратическое отклонение. Случайная величина T имеет распределение Стьюдента с $k = n - 1$ степенями свободы [9, 10, 15]. Критическая область, как и в рассмотренном выше случае с известной дисперсией $D(X) = \sigma^2$, строится в зависимости от вида конкурирующей гипотезы.

Правило 1. Для того чтобы при заданном уровне значимости α проверить гипотезу H_0: $a = a_0$ о равенстве неизвестной генеральной средней a нормальной совокупности с неизвестной дисперсией σ^2 гипотетическому значению a_0 при конкурирующей гипотезе H_0: $a \neq a_0$, необходимо вычислить $T_{\text{набл}} = \dfrac{\overline{x}_{\text{в}} - a_0}{s}\sqrt{n}$ и по таблице критических точек распределения Стьюдента [9, 10, 15] по заданному уровню значимости α, помещенному в верхней строке таблицы, и числу степеней свободы $k = n - 1$ найти критическую точку $t_{\text{двуст.кр}}(\alpha;\ k)$ *двусторонней критической области*. Если $\left|T_{\text{набл}}\right| \le t_{\text{двуст.кр}}$ – нет оснований отвергнуть гипотезу H_0; если $\left|T_{\text{набл}}\right| > t_{\text{двуст.кр}}$ – гипотезу H_0 отвергают.

Правило 2. При конкурирующей гипотезе H_1: $a > a_0$ по заданному уровню значимости α, помещенному в нижней строке таблицы критических точек распределения Стьюдента [9, 10, 15], и числу степеней свободы $k = n - 1$ найти критическую точку $t_{\text{правост.кр}}(\alpha;\ k)$ *правосторонней критической области*. Если $T_{\text{набл}} \le t_{\text{правост.кр}}$ – нет оснований отвергнуть гипотезу H_0; если $T_{\text{набл}} > t_{\text{правост.кр}}$ – гипотезу H_0 отвергают.

Правило 3. При конкурирующей гипотезе H_1: $a < a_0$ сначала по правилу 2 находят «вспомогательную» критическую точку $t_{\text{правост.кр}}(\alpha;\ k)$, а затем полагают границу *левосторонней критической области* $t_{\text{левост.кр}}(\alpha;\ k) = -t_{\text{правост.кр}}(\alpha;\ k)$. Если $T_{\text{набл}} > -t_{\text{правост.кр}}$ – нет оснований отвергнуть гипотезу H_0; если $T_{\text{набл}} < -t_{\text{правост.кр}}$ – гипотезу H_0 отвергают.

Пример 9. По выборке объема $n = 20$, извлеченной из нормальной генеральной совокупности, найдены выборочное среднее значение $\overline{x}_{\text{в}} = 18$ и исправленное среднее квадратическое отклонение $s = 4{,}5$. При уровне значимости 0,05 проверить гипотезу H_0: $a = a_0 = 17$ при конкурирующей гипотезе H_1: $a \ne 17$.

Решение. Вычислим наблюдаемое значение критерия:

$$T_{\text{набл}} = \frac{\overline{x}_{\text{в}} - a_0}{s}\sqrt{n} = \frac{18-17}{4{,}5}\sqrt{20} = 0{,}99\,.$$

Поскольку конкурирующая гипотеза H_1: $a \ne 17$ – двусторонняя, то по таблице критических точек распределения Стьюдента [9, 10, 15] по уровню значимости $\alpha = 0{,}05$, помещенному в верхней строке таблицы, и по числу степеней свободы $k = 20 - 1 = 19$, согласно правилу 1, находим критическую точку $t_{\text{двуст.кр}}(\alpha;\ k) =$ $= t_{\text{двуст.кр}}(0{,}05;\ 19) = 2{,}09$. Так как $\left|T_{\text{набл}}\right| = 0{,}99 < t_{\text{двуст.кр}} = 2{,}09$, то нет оснований отвергнуть нулевую гипотезу H_0: $a = a_0 = 17$. Следовательно, выборочное среднее *незначимо* отличается от гипотетической генеральной средней. ▲

● **Задача 10.** В ОТК поступают подшипники, средний диаметр которых 12 см. В течение суток отобрано 90 подшипников и сделаны замеры их диаметров. Среднее значение диаметра оказалось $\overline{x}_{\text{в}} = 12{,}075$ см. При уровне значимости $\alpha = 0{,}05$ выясните, случайно ли это отклонение, если размеры подшипников распределены по нормальному закону с дисперсией $\sigma^2 = 0{,}069$.

7. НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

В подразд. 6 закон распределения генеральной совокупности предполагался известным. Если же он неизвестен, но имеются основания предположить, что этот закон имеет определенный вид (например, A), то проверяют нулевую гипотезу H_0: генеральная совокупность распределена по закону A. Проверка гипотезы о предполагаемом законе распределения так же, как и проверка гипотезы о неизвестных параметрах известного закона распределения, производится при помощи специально подобранной СВ – *статистического критерия* (статистики). Как бы хорошо ни был подобран теоретический закон распределения, между эмпирическим и теоретическим распределениями возможны расхождения. Возникает вопрос: обусловлены ли эти расхождения только случайными обстоятельствами, связанными, например, с ограниченным числом наблюдений, или они являются существенными и связаны с тем, что теоретический закон распределения подобран неудачно. Для ответа на этот вопрос и служат критерии согласия.

Критерием согласия называют критерий проверки статистической гипотезы о предполагаемом законе неизвестного распределения.

Одним из основных критериев согласия является *критерий* χ^2 (или *критерий Пирсона*).

7.1. Критерий согласия Пирсона

Критерий согласия Пирсона позволяет, в частности, проверить гипотезу о нормальном распределении генеральной совокупности. Для проверки этой гипотезы будем сравнивать *эмпирические* (наблюдаемые) *частоты* n_i и *теоретические* (т. е. вычисленные в предположении нормального закона распределения) *частоты* n_i'. Эти частоты, как правило, различаются. Случайно или значимо это расхождение? Критерий согласия Пирсона дает ответ на поставленный вопрос.

В качестве статистики для проверки гипотезы H_0 примем СВ

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i'\right)^2}{n_i'}, \qquad n_i' = np_i. \tag{26}$$

Ясно, что значение χ^2 в (26) тем меньше, чем меньше различаются эмпирические n_i и теоретические n_i' частоты. Таким образом, статистический критерий (26) характеризует близость эмпирического и теоретического распределения.

Как доказал Пирсон, при $n \to \infty$ закон распределения СВ (26) стремится к закону распределения χ^2 с k степенями свободы. Поэтому СВ в (26) обозначается через χ^2, а сам критерий называют *критерием согласия* χ^2. Число степеней свободы k находят из равенства $k = l - r - 1$, где l – число групп (частичных интервалов), r – число параметров предполагаемого распределения, которые оценены по данным выборки (для нормального закона распределения $r = 2$, поэтому $k = l - 3$).

Пусть по выборке объема n нормально распределенной генеральной совокупности X с равноотстоящими вариантами x_i: $x_{i+1} - x_i = h$, $i = \overline{1, l-1}$, получено эмпирическое распределение

x_i	x_1	x_2	…	x_l
n_i	n_1	n_2	…	n_l

$\sum_{i=1}^{l} n_i = n$.

Задача. При уровне значимости α проверить справедливость гипотезы H_0: генеральная совокупность распределена по нормальному закону.

Правило 1. Для проверки гипотезы H_0 необходимо:

1) вычислить $\overline{x}_в$ и $\sigma_в$;

2) вычислить теоретические частоты $n_i' = \dfrac{nh}{\sigma_в}\varphi(u_i)$, где n – объем выборки, $u_i = \dfrac{x_i - \overline{x}_в}{\sigma_в}$, $\varphi(u) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$;

3) сравнить эмпирические n_i и теоретические n_i' частоты с помощью критерия χ^2 из (26). Для этого:

а) составить таблицу 2 для нахождения наблюдаемого значения критерия $\chi^2_{набл} = \sum_{i=1}^{l} \dfrac{\left(n - n_i'\right)^2}{n_i'}$;

б) из табл. 5 приложения [10] критических точек распределения χ^2 при заданном уровне значимости α и числе степеней свободы $k = s - 3$ (s – число различных значений вариант) найти критические точки $\chi^2_{кр}(\alpha;\ k)$ правосторонней критической области.

Таблица 2

Интервал	n_i	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\dfrac{(n_i - np_i)^2}{np_i}$	$\dfrac{n_i^2}{np_i}$
$(-\infty; x_2)$	n_1	p_1	np_1	$n_1 - np_1$	$(n_1 - np_1)^2$	$\dfrac{(n_1 - np_1)^2}{np_1}$	$\dfrac{n_1^2}{np_1}$
$[x_2; x_3)$	n_2	p_2	np_2	$n_2 - np_2$	$(n_2 - np_2)^2$	$\dfrac{(n_2 - np_2)^2}{np_2}$	$\dfrac{n_2^2}{np_2}$
…	…	…	…	…	…	…	…
$[x_{l-1}; +\infty)$	n_l	p_l	np_l	$n_l - np_l$	$(n_l - np_l)^2$	$\dfrac{(n_l - np_l)^2}{np_l}$	$\dfrac{n_l^2}{np_l}$
	$\sum_{i=1}^{l} n_i = n$	$\sum_{i=1}^{l} p_i = 1$				$\chi^l_{\text{набл}} = \sum_{i=1}^{l}\dfrac{(n_i - np_i)^2}{np_i}$	$\sum_{i=1}^{l}\dfrac{n_i^2}{np_i}$

Построим *правостороннюю критическую область* (так как односторонний критерий более «жестко» отвергает гипотезу H_0), исходя из требования, чтобы, в предположении справедливости гипотезы H_0, вероятность попадания критерия в эту область была равна принятому уровню значимости α: $P\,[\chi^2 > \chi^2_{\text{кр}}(\alpha; k)] = \alpha$. Следовательно, правосторонняя критическая область определяется неравенством $\chi^2 > \chi^2_{\text{кр}}(\alpha; k)$, а область принятия гипотезы H_0 – неравенством $\chi^2 \le \chi^2_{\text{кр}}(\alpha; k)$.

Если $\chi^2_{\text{набл}} \le \chi^2_{\text{кр}}(\alpha;\, k)$ – нет оснований отвергнуть гипотезу H_0 о нормальном распределении генеральной совокупности. Другими словами, эмпирические n_i и теоретические n_i' частоты различаются незначимо. Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}(\alpha;\, k)$ – гипотезу H_0 отвергают. Следовательно, различие эмпирических и теоретических частот значимо.

Заметим, что последний столбец табл. 2 служит для контроля сделанных вычислений: $\sum_{i=1}^{l}\dfrac{n_i^2}{np_i} - n = \chi^2_{\text{набл}}$, где $\chi^2_{\text{набл}}$ – значение критерия (26), вычисленное по данным наблюдений.

• **Задача 11.** Покажите, что для контроля вычислений наблюдаемого критерия $\chi^2_{\text{набл}}$ можно использовать равенство $\sum_{i=1}^{k}\dfrac{n_i^2}{n_i'} - n = \chi^2_{\text{набл}}$.

● **Задача 12.** При уровне значимости 0,05 проверить гипотезу H_0: генеральная совокупность распределена по нормальному закону, если известны эмпирические и теоретические частоты:

Эмпирические частоты n_i	5	6	14	32	43	39	30	20	6	5
Теоретические частоты np_i	4	7	12	29	48	35	34	18	7	6

Пусть x_i, $i=\overline{1,\ k}$, – выборка объема n нормально распределенной генеральной совокупности с изучаемым признаком X:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

$$\sum_{i=1}^{k} n_i = n.$$

Для того чтобы проверить гипотезу H_0: генеральная совокупность распределена нормально, используем критерий Пирсона и проведем вычисления согласно правилу 2.

Правило 2. Для проверки гипотезы H_0 при заданном уровне значимости α необходимо:

1. Весь интервал наблюдаемых значений СВ X выборки объема n от $x_{\min}$ до $x_{\max}$ разделить на l частичных интервалов $(x_i;\ x_{i+1})$ одинаковой длины $x_{i+1} - x_i = h$, найти середины $x_i^* = \dfrac{x_i + x_{i+1}}{2}$, $i=\overline{1,\ k}$, частичных интервалов. В качестве частоты n_i варианты x_i^* принять число вариант, попавших в i-й интервал. В результате получим последовательность равноотстоящих вариант и соответствующих им частот:

x_i	x_1^*	x_2^*	…	x_l^*
n_i	n_1	n_2	…	n_l

$$\sum_{i=1}^{l} n_i = n.$$

2. Вычислить $\overline{x}_{\text{в}}^*$ и выборочное среднее квадратическое отклонение $\sigma_{\text{в}}^*$.

3. Пронормировать СВ X, т. е. перейти к величине $Z = \dfrac{(X - \overline{x}_{\text{в}}^*)}{\sigma_{\text{в}}^*}$ и вычислить концы интервалов $(z_i;\ z_{i+1})$: $z_i = \dfrac{x_i - \overline{x}_{\text{в}}^*}{\sigma_{\text{в}}^*}$, $z_{i+1} = \dfrac{x_{i+1} - \overline{x}_{\text{в}}^*}{\sigma_{\text{в}}^*}$, при этом принять наименьшее значение $z_1 = -\infty$, а наибольшее $z_l = +\infty$.

4. Вычислить теоретические вероятности p_i попадания СВ X в интервалы $(x_i;\ x_{i+1})$ по формуле $p_i = \Phi(z_{i+1}) - \Phi(z_i)$, где ($\Phi(z)$ – функция Лапласа [1, 3, 9, 10, 15, 19]). Найти теоретические частоты $n_i' = np_i$.

5. Вычислить наблюдаемое значение критерия Пирсона $\chi^2_{\text{набл}} = \sum_{i=1}^{l} \frac{(n_i - n_i')^2}{n_i'}$. По таблице критических точек распределения χ^2 по заданному уровню значимости α и числу степеней свободы $k = l - 3$ найти критическую точку $\chi^2_{\text{кр}}(\alpha;\ k)$. Если $\chi^2_{\text{набл}} \le \chi^2_{\text{кр}}(\alpha;\ k)$ – нет оснований отвергнуть нулевую гипотезу H_0. Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}(\alpha; k)$ – *нулевую гипотезу* H_0 *отвергают.*

7.2. Критерий согласия Колмогорова

На практике наряду с критерием χ^2 часто используют *критерий Колмогорова* (*критерий согласия* λ), который применяется для проверки гипотез о законах распределения непрерывных СВ. Отличие от критерия согласия Пирсона χ^2 состоит в том, что сравниваются эмпирическая $F^*(x)$ и теоретическая $F(x)$ функции распределения при известных параметрах распределения. В качестве меры расхождения между теоретическим и эмпирическим распределениями рассматривают максимальное значение абсолютной величины разности $D^* = \sup_x |F^*(x) - F(x)|$. СВ D^* называется *статистикой Колмогорова*. Для *непрерывных СВ* с помощью этого критерия можно проверить гипотезу о виде функции распределения, а также:

1) согласие эмпирического распределения с гипотетическим (теоретическим);

2) гипотезу о том, что две выборки взяты из одной и той же генеральной совокупности, т. е. определяются одним и тем же теоретическим распределением.

Теорема Колмогорова. *Пусть* $F(x)$ *– теоретическая функция распределения СВ* X; $x_1, \ldots, x_n$ *– выборка объема* n *и* $F^*(x)$ *– эмпирическая функция распределения. Тогда*

$$P(D^*\sqrt{n} < \lambda) \xrightarrow[n\to\infty]{} K(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2},\ \lambda > 0,$$

где $D^* = \sup_x |F^*(x) - F(x)|$.

При использовании этого критерия следует учитывать, что функция распределения $F(x)$ не должна зависеть от параметров выборки.

Статистика Колмогорова используется для: 1) проверки гипотезы H_0: $F^*(x) = F(x)$ ($F^*(x)$ – эмпирическая функция распределения

СВ X, вычисленная по выборке $x_1, x_2, \ldots, x_n$; $F(x)$ – гипотетическая функция распределения СВ X); 2) построения доверительных границ для $F(x)$.

При использовании статистики Колмогорова D^* следует иметь в виду таблицу приложений критических значений λ_α статистики Колмогорова [15, 19], в которой для типичных уровней значимости α приведены значения *критической точки* λ_α, удовлетворяющей уравнению

$$P(D^* = \sup_x |F^*(x) - F(x)| \ > \lambda_\alpha) = \alpha.$$

Так как распределение статистики Колмогорова D^* не зависит от оцениваемой функции $F(x)$ и в качестве меры расстояния между $F^*(x)$ и $F(x)$ используют максимальное отклонение, то величину D^* можно применять для построения доверительных границ непрерывной функции распределения $F(x)$.

Для любой неизвестной непрерывной функции $F(x)$ при произвольном x имеем

$$P(F^*(x) - \lambda_\alpha \le F(x) \le F^*(x) + \lambda_\alpha) = 1 - \alpha.$$

Значит, доверительная область – полоса шириной $2\lambda_\alpha$, в центре которой находится эмпирическая функция распределения $F^*(x)$. С вероятностью $1 - \alpha$ значения истинной функции распределения $F(x)$ целиком лежат внутри этой полосы. Сформулированный результат дает основание для оценивания максимального объема выборки, необходимого для аппроксимации неизвестной функции распределения $F(x)$ с заданной точностью. Например, при $\alpha = 0{,}01$, $n = 100$ значения эмпирической функции распределения $F^*(x)$ повсюду отстоят от значений истинной функции $F(x)$ не более чем на 0,161 [15, 19].

Проверку гипотезы H_0 о согласии теоретического закона распределения с данными выборки с помощью критерия Колмогорова проводят в следующем порядке.

Схема применения критерия Колмогорова:

1) располагают результаты наблюдений x_i по возрастанию их значений;

2) находят эмпирическую функцию распределения $F^*(x) = \dfrac{n_x}{n}$;

3) вычисляют, пользуясь предполагаемой теоретической функцией распределения $F(x)$, ее значения, соответствующие наблюдаемым значениям x_i СВ X;

4) находят для каждого x_i модуль разности между эмпирической и теоретической функциями распределения D^*: $D^* = \sup_x |F^*(x) - F(x)|$;

5) вычисляют величину $\lambda_{набл} = D^* \sqrt{n}$;

6) по таблице значений функции Колмогорова находят $\lambda_{кр} = \lambda(\alpha)$. Если $\lambda_{набл} > \lambda_{кр}$, то гипотеза H_0 о том, что функция распределения генеральной совокупности X равна $F(x)$, отвергается. Если $\lambda_{набл} \le \lambda_{кр}$, то считают, что гипотеза H_0 не противоречит опытным данным.

8. ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ. ЛИНЕЙНАЯ РЕГРЕССИЯ

Одной из основных задач МС является нахождение зависимости между двумя или несколькими СВ. В естественных науках часто речь идет о ***функциональной зависимости*** между величинами X и Y, когда *каждому значению одной переменной соответствует вполне определенное значение другой* (например, скорость свободного падения тела в вакууме зависит от времени падения). Однако строгая функциональная зависимость реализуется редко, так как обе СВ или одна из них подвержены действию случайных факторов. В этом случае возникает ***статистическая зависимость***.

Статистической (*вероятностной, стохастической, корреляционной*) называют зависимость, при которой *изменение одной из величин влечет изменение распределения другой*. Примером статистической зависимости может служить зависимость всхожести семян некоторых культур от количества микроэлементов при их обработке, зависимость производительности труда на предприятии от его энерговооруженности и т. д. Таким образом, статистическая зависимость между двумя СВ Y и X неоднозначна. Статистическая зависимость, в частности, проявляется в том, что при изменении одной из величин изменяется *среднее значение* другой. Для исследователя представляет интерес *усредненная* схема зависимости, т. е. закономерность в изменении *условного математического ожидания* $M_x(Y)$ в зависимости от x. Такую статистическую зависимость называют ***регрессионной*** [9, 10, 15].

Условным математическим ожиданием $M_x(Y) = M(Y|X = x) = \overline{Y}_x$ (условной средней) называется математическое ожидание СВ Y, вычисленное в предположении, что СВ X приняла значение x.

Регрессионной зависимостью между двумя переменными величинами называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой.

Отметим, что в корреляционном анализе изучается в основном сила (теснота) корреляционной зависимости (т. е. зависимости, не имеющей, вообще говоря, строго функционального характера), а в регрессионном анализе ее форма.

Регрессионная зависимость, например, имеется: а) между ростом и весом человека: с увеличением роста *средний вес* также возрастает; б) между надежностью автомобиля и его возрастом: чем автомобиль старше, тем *в среднем* меньше его надежность.

Рассмотрим пример СВ Y, которая связана с другой СВ X не функционально, а корреляционно.

Пример 10. Пусть Y – успеваемость студентов, X – посещаемость ими учебных занятий. У одинаковых студенческих групп (по количеству студентов и количеству часов лекционных и практических занятий) по результатам экзаменационной сессии успеваемость разная, т. е. Y не является функцией от X – посещаемости учебных занятий: $Y \neq f(X)$. Однако, как показывает опыт, результаты экзаменационной сессии лучше у тех студентов, которые систематически посещали учебные занятия. Это означает, что Y связано с X корреляционной зависимостью. ▲

В качестве *оценок* условных математических ожиданий принимают *условные выборочные средние* [9, 10, 15], которые находят по данным наблюдений, т. е. по выборке. Так, например, если при $x_1 = 2$ СВ Y приняла значения $y_1 = 3, y_2 = 7, y_3 = 5$, то *условное выборочное среднее* $\overline{y}_{x_1}$ равно: $\overline{y}_{x_1} = \dfrac{y_1 + y_2 + y_3}{3} = \dfrac{3+7+5}{3} = 5$. Условное математическое ожидание $M_x(Y)$ СВ Y есть функция от x: $M_x(Y) = f(x)$, которую называют (теоретической) *функцией регрессии Y на X*. Поскольку каждому значению x соответствует одно значение условного среднего, т. е. $\vec{Y}_x = f(x)$ является функцией от x, то можно сказать, что СВ Y зависит от СВ X *корреляционно*.

Уравнение $\vec{Y}_x = f(x)$ (или $y = f(x)$) называется (теоретическим) *уравнением регрессии Y на X*. Функция $f(x)$ называется (теоретической) *регрессией Y на X* а ее график – *линией регрессии СВ Y на СВ X*.

Аналогично для СВ X определяются *условное математическое ожидание* $M_y(X)$ (или в других обозначениях – $M(X|Y = y) = \overline{X}_y$), *корреляционная зависимость* СВ X от СВ Y, *функция регрессии* СВ X на СВ Y: $\overline{X}_y = \varphi(y)$.

Основными задачами теории корреляции являются:

1. Установление формы корреляционной связи, т. е. установление *вида* функции регрессии (линейная, квадратичная, показательная и т. д.).

2. Оценка *тесноты* корреляционной связи Y и X, которая оценивается величиной рассеяния значений Y около $\vec{Y}_x$. Большое рассеяние означает слабую зависимость Y от X либо вообще отсутствие тако-

вой. Малое рассеяние указывает на существование достаточно сильной зависимости Y от X.

Важной с точки зрения приложений является ситуация, когда обе функции регрессии $f(x)$, $\varphi(y)$ являются линейными. Тогда говорят, что СВ X и Y связаны линейной корреляционной зависимостью (имеет место *линейная корреляция*). Такая ситуация возникает, например, если система СВ (X, Y) имеет совместное нормальное распределение. Тогда модельные (теоретические) уравнения регрессии являются линейными, а их графики – линии регрессии – прямыми.

Рассмотрим методы нахождения линейной регрессии, представляющей наибольший интерес. Пусть даны результаты n измерений двух СВ X и Y: (x_1, y_1), (x_2, y_2), …, (x_n, y_n). Предварительное представление о характере зависимости между СВ X и Y можно получить, если элементы выборки (x_i, y_i), $i=\overline{1, n}$, изобразить графически точками координатной плоскости в выбранной системе координат. В результате получим *точечную диаграмму* статистической зависимости, которая называется ***корреляционным полем***. По его виду можно составить предварительное мнение о степени и типе зависимости двух СВ.

Как известно [9, 10, 15], для описания системы двух СВ (X, Y) вводят математические ожидания $M(X)$, $M(Y)$, дисперсии $D(X)=\sigma_X^2$, $D(Y)=\sigma_Y^2$ составляющих СВ, *корреляционный момент* (*ковариацию*) $K(X, Y) = M\{[X - M(X)][Y - M(Y)]\}$ и *коэффициент корреляции* $r=\dfrac{K(X, Y)}{\sigma(X)\sigma(Y)}$.

Корреляционный момент $K(X, Y)$ служит для характеристики связи между СВ X и Y: если X и Y независимы, то $K(X, Y) = 0$, следовательно, и $r = 0$.

Две СВ X и Y называются *коррелированными*, если их корреляционный момент (или, что то же, коэффициент корреляции) отличен от нуля. В противном случае СВ X и Y называются *некоррелированными*.

Рассмотрим вопрос о силе связи между признаками X и Y. Из определения *теоретического коэффициента корреляции* $r=\dfrac{K(X, Y)}{\sigma(X)\sigma(Y)}$ и оценок параметров теоретического распределения через выборочные данные *выборочный коэффициент корреляции* $r_в$ может быть представлен в виде

$$r_в=\frac{K_в(X, Y)}{\sigma_в(X)\sigma_в(Y)}, \tag{27}$$

где $K_B(X, Y)$ – выборочный корреляционный момент,

$$K_в(X,\ Y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_в)(y_i-\overline{y}_в)=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\overline{x}_в\overline{y}_в=\overline{xy}_в-\overline{x}_в\overline{y}_в,$$

$\sigma_в(X)=\sqrt{D_в(X)}=\sqrt{\overline{x}_в^2-(\overline{x}_в)^2}$, $\sigma_в(Y)=\sqrt{D_в(Y)}=\sqrt{\overline{y}_в^2-(\overline{y}_в)^2}$ – выборочные среднеквадратические отклонения признаков X и Y. Следовательно, из (27) имеем

$$r_в=\frac{\overline{xy}_в-\overline{x}_в\overline{y}_в}{\sqrt{\overline{x}_в^2-(\overline{x}_в)^2}\sqrt{\overline{y}_в^2-(\overline{y}_в)^2}}. \tag{28}$$

● **Задача 13.** Докажите, что $K_в(X,\ Y)\ =\ \overline{xy}_в-\overline{x}_в\overline{y}_в$.

Формула (28) симметрична относительно двух переменных (т. е. x и y можно менять местами). Выборочный коэффициент корреляции $r_в$ обладает теми же свойствами, что и теоретический коэффициент корреляции r.

***Основные свойства выборочного коэффициента корреляции*:**

1. $r_в\in[-1;\ 1]$, т. е. $-1\le r_в\le 1$.

2. Если все значения переменных увеличить (уменьшить) на одно и то же число или в одно и то же число раз, то величина выборочного коэффициента корреляции не изменится.

3. При $r_в=\pm1$ корреляционная связь представляет *линейную зависимость*. При этом линии регрессии Y на X и X на Y совпадают, все наблюдаемые значения располагаются на общей прямой.

4. Если с ростом значений одной СВ значения второй в среднем возрастают, то $r_в>0$, если убывают, то $r_в<0$.

5. При $r_в=0$ *линейная* корреляционная связь отсутствует, групповые средние переменных совпадают с их общими средними, а линии регрессии Y на X и X на Y параллельны соответствующим осям координат.

Замечание 8. Равенство $r_в=0$ говорит лишь об отсутствии *линейной* корреляционной зависимости (т. е. о *некоррелированности* СВ X и Y), но не вообще об отсутствии корреляционной (статистической) зависимости.

Выборочный коэффициент корреляции $r_в$ является *оценкой генерального коэффициента корреляции* $r_г=\dfrac{K(X,\ Y)}{\sigma(X)\sigma(Y)}$, характеризующего тесноту связи между СВ X и Y генеральной совокупности. На практике о тесноте корреляционной зависимости между рассматриваемыми СВ судят не по величине $r_г$, который, как правило, неизвестен, а по величине его выборочного аналога $r_в$. Так как $r_в$ вычисляется по значениям переменных, случайно попавших в выборку из генеральной совокупности, то в отличие от параметра $r_г$ параметр $r_в$ – *величина случайная*.

8.1. Проверка значимости выборочного коэффициента корреляции в случае выборки из нормального двумерного распределения

Пусть двумерная генеральная совокупность (X, Y) распределена нормально. Из этой совокупности извлечена выборка объема n, по которой найден выборочный коэффициент корреляции $r_в$, $r_в \neq 0$. Поскольку выборка произведена случайно, нельзя утверждать, что $r_г \neq 0$. Требуется проверить, значимо ли отличие выборочного коэффициента корреляции $r_в$ от нуля либо это вызвано только случайными изменениями выборочных значений.

Задача. При заданном уровне значимости α проверить справедливость гипотезы H_0: $r_г = 0$ о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе H_1: $r_г \neq 0$.

Для проверки гипотезы H_0 при заданном уровне значимости α и объеме выборки n вычисляют статистику $T_{набл} = \dfrac{|r_в|}{\sqrt{1-r_в^2}}\sqrt{n-2}$. По таблице критических точек распределения Стьюдента [9, 10 15] по заданному уровню значимости α и числу степеней свободы $k = n - 2$ находят $t_{кр} = t_{\alpha;n-2}$ для двусторонней критической области. Если $T_{набл} \le t_{кр}$ – нет оснований отвергнуть гипотезу H_0. Если $T_{набл} > t_{кр}$, то гипотезу H_0 о равенстве коэффициента корреляции нулю отвергают. Другими словами, $r_в$ значимо отличается от нуля, т. е. СВ X и Y коррелированны.

Пример 11. По выборке объема $n = 102$, извлеченной из нормальной двумерной совокупности, найден выборочный коэффициент корреляции $r_в = 0,3$. При уровне значимости 0,05 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции $r_г$ при конкурирующей гипотезе H_1: $r_г \neq 0$.

Решение. Найдем наблюдаемое значение критерия $T_{набл} = \dfrac{|r_в|}{\sqrt{1-r_в^2}}\sqrt{n-2} = \dfrac{0,3}{\sqrt{1-0,3^2}}\sqrt{102-2} = \dfrac{3}{0,954} = 3,14$. По условию конкурирующая гипотеза имеет вид $r_г \neq 0$, следовательно, критическая область – двусторонняя. По уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = n - 2 = 100$ по таблице критических точек распределения Стьюдента для двусторонней критической области [9, 10, 15] находим критическую точку $t_{кр} = t_{\alpha;n-2} = t_{кр}(0{,}05;\ 100) = 1{,}98$. Так как $T_{набл} = 3{,}14 > t_{кр} = 1{,}98$, то нулевую гипотезу H_0 отвергаем, т. е. выборочный коэффициент корреляции значимо отличается от нуля. Таким образом, СВ X и Y коррелированны. ▲

• **Задача 14.** На химическом производстве по результатам 17 наблюдений величины Y – выхода продуктов (кг/ч) и величины X – температуры реакции (°С) рассчитан выборочный коэффициент корреляции $r_в = 0{,}992$. Проверьте, можно ли с вероятностью ошибки 0,05 принять гипотезу о наличии линейной корреляционной зависимости выхода продуктов Y от температуры X реакции, если совместное распределение этих величин является нормальным.

8.2. Определение коэффициентов уравнения линейной регрессии

Рассмотрим систему двух СВ (X, Y). Если обе функции $f(x)$, $\varphi(y)$ регрессии Y на X ($\bar{Y}_x = f(x)$) и X на Y ($\bar{X}_y = \varphi(y)$) линейны, то говорят, что СВ X и Y связаны *линейной корреляционной зависимостью.* Графиками линейных функций регрессии являются прямые линии. Пусть, например, приближенное представление СВ Y задано в виде линейной функции СВ X: $Y \approx f(X) = a + bX$, где a, b – параметры, подлежащие определению. Их можно определить различными способами. Наиболее употребительным является метод наименьших квадратов (МНК). Функцию $f(X)$ называют *наилучшим приближением Y* в смысле МНК, если математическое ожидание $M[Y - f(X)]^2$ принимает наименьшее возможное значение. Поэтому $f(X)$ называют *среднеквадратической регрессией Y* на X.

Теорема 1 [10]. *Линейная среднеквадратическая регрессия Y на X имеет вид*

$$f(x) = m_y + r\frac{\sigma_y}{\sigma_x}(x - m_x),$$

где $m_x = M(X);\ m_y = M(Y);\ \sigma_x = \sqrt{D(X)};\ \sigma_y = \sqrt{D(Y)};\ r = \dfrac{K(X,\ Y)}{\sigma_x \sigma_y}$ – *коэффициент корреляции СВ X и Y.*

Из теоремы 1 следует, что параметры a, b функции $f(X) = a + bX$ имеют вид: $a = m_y - r\dfrac{\sigma_y}{\sigma_x}m_x$, $b = r\dfrac{\sigma_y}{\sigma_x}$. Коэффициент $b = r\dfrac{\sigma_y}{\sigma_x}$ называется *коэффициентом регрессии Y на X*, а прямая

$$y - m_y = r\frac{\sigma_y}{\sigma_x}(x - m_x)$$

прямой среднеквадратической регрессии Y на X. Аналогично можно получить уравнение $x - m_x = r\frac{\sigma_x}{\sigma_y}(y - m_y)$ прямой среднеквадратической регрессии X на Y, где $r\frac{\sigma_x}{\sigma_y}$ – *коэффициент регрессии X на Y.* Из сопоставления уравнений линейной регрессии Y на X и X на Y видно, что при $r = \pm1$ они совпадают. Кроме того, очевидно, что прямые $y - m_y = r\frac{\sigma_y}{\sigma_x}(x - m_x)$ и $x - m_x = r\frac{\sigma_x}{\sigma_y}(y - m_y)$ проходят через точку (m_x; m_y), называемую *центром совместного распределения СВ X и Y.*

Выше были введены уравнения регрессии Y на X: $M(Y|X = x) = f(x)$ и X на Y: $M(X|Y = y) = \varphi(y)$. Поскольку условное математическое ожидание $M(Y|X = x)$ является функцией от x, то и его оценка – условное выборочное среднее $\overline{y}_x$ – также является функцией от x: $\overline{y}_x = f^*(x)$, где $f^*(x)$, в силу теоремы 1, имеет вид $f^*(x) = \overline{y}_в + r_в\frac{s_y}{s_x}(x - \overline{x}_в)$. Уравнение $\overline{y}_x = f^*(x)$ (иногда пишут $y_x = f^*(x)$) называют *выборочным уравнением регрессии Y на X*; функцию $f^*(x)$ – *выборочной регрессией Y на X*, а ее график – *выборочной линией регрессии Y на X.* Аналогично уравнение $\overline{x}_y = \varphi^*(y)$, где $\varphi^*(y) = \overline{x}_в + r_в\frac{s_x}{s_y}(y - \overline{y}_в)$, называют *выборочным уравнением регрессии X на Y*; функцию $\varphi^*(y)$ – *выборочной регрессией X на Y*, а ее график – *выборочной линией регрессии X на Y.*

Рассмотрим теперь, как по данным наблюдений (x_1, y_1), (x_2, y_2), …, (x_n, y_n), полученным в результате n независимых опытов, найти оценки неизвестных параметров функций $f^*(x)$, $\varphi^*(y)$, вид которых известен, и как оценить тесноту связи между СВ X и Y, а также установить, коррелированны ли эти величины.

Пусть принята гипотеза о линейной зависимости между СВ X и Y. По данным наблюдений найдем выборочное уравнение прямой линии среднеквадратической регрессии Y на X: $f^*(x) = \overline{y}_x = a + bx$. Будем полагать, что все результаты измерений $(x_i;\ y_i)$, $i = \overline{1, n}$ различны:

X	x_1	x_2	…	x_n
Y	y_1	y_2	…	y_n

Подберем параметры a, b так, чтобы точки $(x_1;\ y_1)$, $(x_2;\ y_2)$, …, $(x_n;\ y_n)$, построенные на плоскости xOy по данным наблюдений, лежали как можно ближе к прямой $\overline{y}_x = a + bx$ в смысле МНК. Сформулированное требование означает, что параметры a, b будем выбирать из условия минимума суммы квадратов отклонений $\overline{y}_{x_i} - y_i,\ i = \overline{1,\ n}$. Здесь $\overline{y}_{x_i}$ – ордината, вычисленная по эмпирическому (выборочному) уравнению $\overline{y}_{x_i} = a + bx_i$ и соответствующая наблюдаемому значению x_i, а y_i – наблюдаемая ордината, соответствующая x_i. Следовательно, рассмотрим функцию

$$F(a,\ b) = \sum_{i=1}^{n}\left(\overline{y}_{x_i} - y_i\right)^2 = \sum_{i=1}^{n}\left((a + bx_i) - y_i\right)^2 \to \min_{a,\ b}.$$

Минимум достигается в единственной точке $(a;\ b)$, которую находят из условия:

$$\begin{cases} \dfrac{\partial F(a;\ b)}{\partial a} = 0, \\ \dfrac{\partial F(a;\ b)}{\partial b} = 0. \end{cases}$$

Найдя соответствующие частные производные и приравняв их к нулю, получим

$$\begin{cases} na + b\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i, \\ a\sum\limits_{i=1}^{n} x_i + b\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i \end{cases}$$

или, после деления обеих частей уравнений полученной системы на n, имеем

$$\begin{cases} a + b\overline{x}_\text{в} = \overline{y}_\text{в}, \\ a\overline{x}_\text{в} + b\overline{x^2_\text{в}} = \overline{xy}_\text{в}, \end{cases}$$

где $\overline{x}_\text{в} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$, $\overline{y}_\text{в} = \dfrac{1}{n}\sum\limits_{i=1}^{n} y_i$, $\overline{x^2_\text{в}} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^2$, $\overline{xy}_\text{в} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i y_i$.

Вычисляя определитель Δ данной системы

$$\Delta = \begin{vmatrix} 1 & \overline{x}_\text{в} \\ \overline{x}_\text{в} & \overline{x^2_\text{в}} \end{vmatrix} = \overline{x^2_\text{в}} - \left(\overline{x}_\text{в}\right)^2 = D_\text{в},$$

находим неизвестные параметры a, b:

$$a=\frac{\Delta a}{\Delta}=\frac{\begin{vmatrix}\overline{y}_{\text{в}} & \overline{x}_{\text{в}}\\ \overline{xy}_{\text{в}} & \overline{x_{\text{в}}^2}\end{vmatrix}}{\overline{x_{\text{в}}^2}-\left(\overline{x}_{\text{в}}\right)^2}=\frac{\overline{y}_{\text{в}}\overline{x_{\text{в}}^2}-\overline{x}_{\text{в}}\overline{xy}_{\text{в}}}{\overline{x_{\text{в}}^2}-\left(\overline{x}_{\text{в}}\right)^2}=$$

$$=\frac{(\overline{y}_{\text{в}}\overline{x_{\text{в}}^2}-\overline{y}_{\text{в}}\left(\overline{x}_{\text{в}}\right)^2)+\overline{y}_{\text{в}}\left(\overline{x}_{\text{в}}\right)^2-\overline{x}_{\text{в}}\overline{xy}_{\text{в}}}{\overline{x_{\text{в}}^2}-\left(\overline{x}_{\text{в}}\right)^2}=\overline{y}_{\text{в}}-\frac{\overline{x}_{\text{в}}\overline{xy}_{\text{в}}-\overline{y}_{\text{в}}\left(\overline{x}_{\text{в}}\right)^2}{\overline{x_{\text{в}}^2}-\left(\overline{x}_{\text{в}}\right)^2}=$$

$$=\overline{y}_{\text{в}}-\frac{\overline{x}_{\text{в}}(\overline{xy}_{\text{в}}-\overline{y}_{\text{в}}\overline{x}_{\text{в}})}{\overline{x_{\text{в}}^2}-\left(\overline{x}_{\text{в}}\right)^2}\overset{(27)}{=}\overline{y}_{\text{в}}-\frac{\overline{x}_{\text{в}}K_{\text{в}}(X,\,Y)}{\sigma_{\text{в}}^2(X)}=$$

$$=\overline{y}_{\text{в}}-\frac{\overline{x}_{\text{в}}r_{\text{в}}(X,\,Y)\sigma_{\text{в}}(X)\sigma_{\text{в}}(Y)}{\sigma_{\text{в}}^2(X)}=\overline{y}_{\text{в}}-r_{\text{в}}(X,\,Y)\frac{\overline{x}_{\text{в}}\sigma_{\text{в}}(Y)}{\sigma_{\text{в}}(X)},$$

$$b=\frac{\Delta b}{\Delta}=\frac{\begin{vmatrix}1 & \overline{y}_{\text{в}}\\ \overline{x}_{\text{в}} & \overline{xy}_{\text{в}}\end{vmatrix}}{\overline{x_{\text{в}}^2}-\left(\overline{x}_{\text{в}}\right)^2}=\frac{\overline{xy}_{\text{в}}-\overline{y}_{\text{в}}\overline{x}_{\text{в}}}{\overline{x_{\text{в}}^2}-\left(\overline{x}_{\text{в}}\right)^2}=\frac{K_{\text{в}}(X,\,Y)}{\sigma_{\text{в}}^2(X)}\overset{(27)}{=}$$

$$=\frac{r_{\text{в}}(X,\,Y)\sigma_{\text{в}}(X)\sigma_{\text{в}}(Y)}{\sigma_{\text{в}}^2(X)}=r_{\text{в}}(X,\,Y)\frac{\sigma_{\text{в}}(Y)}{\sigma_{\text{в}}(X)}.$$

Таким образом, выборочное уравнение линейной регрессии Y на X имеет вид:

$$\overline{y}_x=\overline{y}_{\text{в}}-r_{\text{в}}(X,\,Y)\frac{\overline{x}_{\text{в}}\cdot\sigma_{\text{в}}(Y)}{\sigma_{\text{в}}(X)}+r_{\text{в}}(X,\,Y)\frac{\sigma_{\text{в}}(Y)}{\sigma_{\text{в}}(X)}x$$

или в иной записи: $\overline{y}_x-\overline{y}_{\text{в}}=r_{\text{в}}(X,\,Y)\dfrac{\sigma_{\text{в}}(Y)}{\sigma_{\text{в}}(X)}(x-\overline{x}_{\text{в}})$.

Аналогичным образом можно получить уравнение линейной регрессии X на Y:

$$\overline{x}_y-\overline{x}_{\text{в}}=r_{\text{в}}(X,\,Y)\frac{\sigma_{\text{в}}(X)}{\sigma_{\text{в}}(Y)}\left(y-\overline{y}_{\text{в}}\right).$$

На практике зачастую выборочные дисперсии $\sigma_{\text{в}}^2(X)$, $\sigma_{\text{в}}^2(Y)$ заменяются их несмещенными аналогами

$$s_x^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x}_{\text{в}})^2=\frac{1}{n-1}\sum_{i=1}^{n}x_i^2-\frac{n}{n-1}\overline{x}_{\text{в}}^2,$$

$$s_y^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y}_{\text{в}})^2=\frac{1}{n-1}\sum_{i=1}^{n}y_i^2-\frac{n}{n-1}\overline{y}_{\text{в}}^2.$$

Тогда *эмпирические (выборочные) уравнения линейной регрессии* приобретают вид:

$$\overline{y}_x - \overline{y}_{\text{в}} = r_{\text{в}} \frac{s_y}{s_x}(x - \overline{x}_{\text{в}}), \quad \overline{x}_y - \overline{x}_{\text{в}} = r_{\text{в}} \frac{s_x}{s_y}\left(y - \overline{y}_{\text{в}}\right). \qquad (29)$$

Заметим, что если нанести обе линии регрессии на корреляционное поле, то прямые пересекутся в точке $\left(\overline{x}_{\text{в}}; \overline{y}_{\text{в}}\right)$.

Уравнения линейной регрессии (29) получены в предположении, что все измерения встречаются по одному разу. При большом числе наблюдений одно и то же значение СВ X может повторяться n_x раз, а СВ Y – n_y раз. Одинаковая пара чисел (x; y) может наблюдаться n_{xy} раз. Поэтому результаты наблюдений группируют, подсчитывая *частоты* n_x, n_y, n_{xy}. Все сгруппированные данные записывают в *корреляционную таблицу 3.*

Таблица 3

x_i \ y_j	x_1	x_2	…	x_k	n_{y_j}
y_1	n_{11}	n_{12}	…	n_{1k}	$\sum\limits_{i=1}^{k} n_{1i}$
y_2	n_{21}	n_{22}	…	n_{2k}	$\sum\limits_{i=1}^{k} n_{2i}$
…	…	…	…	…	…
y_m	n_{m1}	n_{m2}	…	n_{mk}	$\sum\limits_{i=1}^{k} n_{mi}$
n_{x_i}	$\sum\limits_{j=1}^{m} n_{j1}$	$\sum\limits_{j=1}^{m} n_{j2}$	…	$\sum\limits_{j=1}^{m} n_{jk}$	$n = \sum\limits_{i=1}^{k} n_{x_i} = \sum\limits_{j=1}^{m} n_{y_j}$

В этом случае $r_{\text{в}} = \dfrac{\sum\limits_{i,j} n_{x_i y_j} x_i y_j - n\overline{x}_{\text{в}}\overline{y}_{\text{в}}}{n\sigma_{\text{в}x}\sigma_{\text{в}y}}$. При построении корреляционной таблицы в клетки верхней строки записывают наблюдаемые значения x_i, $i = \overline{1,\ k}$, в первый столбец – наблюдаемые значения y_j, $j = \overline{1,\ m}$; на пересечении строк и столбцов записывают частоты $n_{x_i y_j}$, $i = \overline{1,\ k}$, $j = \overline{1,\ m}$, наблюдаемых пар значений признаков; в правом нижнем углу помещают сумму всех частот n_{x_i}, $i = \overline{1,\ k}$ и n_{y_j}, $j = \overline{1,\ m}$, равную числу n всех наблюдений.

Для непрерывных СВ корреляционная таблица имеет иной вид (табл. 4), где n_{ij} – частоты (кратности) наблюдаемых пар значений признаков, попавших в соответствующие интервалы $[x_i; x_{i+1})$, $[y_j; y_{j+1})$, $i=\overline{1,\ k}$, $j=\overline{1,\ m}$, длиной h_x, h_y соответственно.

Таблица 4

x_i / y_j	$[x_1; x_2)$	$[x_2; x_3)$	…	$[x_k; x_{k+1}]$	n_{y_j}
$[y_1; y_2)$	n_{11}	n_{12}	…	n_{1k}	$\sum_{i=1}^{k} n_{1i}$
$[y_2; y_3)$	n_{21}	n_{22}	…	n_{2k}	$\sum_{i=1}^{k} n_{2i}$
…	…	…	…	…	…
$[y_m; y_{m+1}]$	$n_{m,1}$	$n_{n,2}$	…	$n_{m,k}$	$\sum_{i=1}^{k} n_{m,i}$
n_{x_i}	$\sum_{j=1}^{m} n_{j1}$	$\sum_{j=1}^{m} n_{j2}$	…	$\sum_{j=1}^{m} n_{jk}$	$n=\sum_{i=1}^{k} n_{x_i}=\sum_{j=1}^{m} n_{y_j}$

В описанном случае табл. 4 сводится к табл. 3 путем перехода к серединам x_i^*, $i=\overline{1,\ k}$, y_j^*, $j=\overline{1,\ m}$, интервалов группировки статистических данных и соответствующим им частотам.

Раздел II. МЕТОДЫ ОПТИМИЗАЦИИ

9. КЛАССИФИКАЦИЯ ОПТИМИЗАЦИОННЫХ ЗАДАЧ

Понятие «оптимизация» уже прочно вошло в жизнь: если у рассматриваемой задачи существует несколько решений, то естественно выбрать то из них, которое лучше. Основные методы оптимизации связаны с вопросами нахождения экстремумов функций и функционалов. Пусть, например, на некотором множестве U задана функция I со значениями в множестве $\mathbf{R}$, т. е. I: $U \to \mathbf{R}$. Следующая задача относится к числу оптимизационных: найти $u \in U$ такое, чтобы $I(u) \le I(\nu)$ для любых $\nu \in U$, что записывается в виде

$$\begin{cases} I(u) \to \min, \\ u \in U. \end{cases} \tag{30}$$

Здесь U представляет собой множество возможных способов достижения определенной цели или множество способов функционирования какой-нибудь системы. В прикладных задачах принадлежность к множеству U выражается с помощью некоторых ограничений. Поэтому U часто называют *множеством* (областью) *ограничений.* Функция I представляет собой *критерий качества*, который предопределяет наш выбор из множества возможных решений. В задаче (30) речь идет о том, чтобы выбрать решение u, которое *минимизирует* значение критерия I на множестве U. Если в (30) заменить I на $-I$, то *задача минимизации* преобразуется в *задачу максимизации.* Задачи оптимизации встречаются в различных сферах человеческой деятельности и подразумевают выбор *оптимального* (в некотором смысле наилучшего) решения по сравнению с другими возможными вариантами. На практике критерием I может быть стоимость, выработка, прибыль, время и т. д.

В данном разделе изучаются различные задачи нахождения *экстремума* (минимума, максимума) *функций* и *функционалов*. Приведем классификацию *оптимизационных задач*.

9.1. Задачи математического программирования

Математическое программирование (МП) представляет собой математическую дисциплину, содержание которой составляют теория и методы решения задач о нахождении экстремумов функций на множествах, определяемых линейными и нелинейными ограничениями (равенствами и неравенствами). МП является одним из разделов *науки об исследовании операций*. В МП можно выделить два направления:

– к первому относятся *детерминированные задачи*, когда вся исходная информация является полностью определенной;

– ко второму, так называемому *стохастическому программированию*, относятся задачи, в которых либо исходная информация содержит элементы неопределенности, либо некоторые параметры носят случайный характер с известными вероятностными характеристиками.

Наименование «математическое *программирование*» связано с тем, что целью решения задач является выбор *программы* действий. Например, ряд производственных и экономических задач связан с распределением ресурсов, которые, как правило, являются ограниченными. Кроме того, само распределение возможно не единственным образом. Необходимую для выпуска продукцию можно получить (за счет выбора вида сырья, технологии производства, применяемого оборудования, организации работы на рабочих местах и всего производственного процесса) различными способами. При этом каждый способ распределения ресурсов, оцениваемый с точки зрения некоторого *критерия* (например, прибыли, объема выпускаемой продукции, удовлетворения спроса потребителей и др.), характеризуется определенным его значением. При решении задачи необходимо найти максимальное (или минимальное) значение критерия. Таким образом, требуется выбрать такое распределение ресурсов (программу, план), которое доставляло бы *экстремальное* значение выбранному критерию. Такую программу (план) называют *оптимальной* [2, 13, 16–18].

Общая задача МП формулируется следующим образом: найти вектор $x = (x_1, x_2, \ldots, x_n)$, удовлетворяющий системе ограничений в виде равенств

$$g_i(x_1, x_2, \ldots, x_n) = b_i,\ i = \overline{1, k},$$

и неравенств

$$g_j\left(x_1, x_2, \ldots, x_n\right) \le b_j,\ \ j = \overline{k+1, m},$$

и доставляющий экстремум целевой функции $f(x_1, x_2, \ldots, x_n)$ (скалярной функции векторного аргумента):

$$f(x_1, x_2, \ldots, x_n) \to \text{extr}.$$

Предполагается, что функции $f(x_1, x_2, \ldots, x_n)$, $g_l(x_1, x_2, \ldots, x_n)$ и постоянные b_l, $l = \overline{1, m}$, известны. Зачастую на некоторые переменные $x_1, x_2, \ldots, x_n$ накладываются условия неотрицательности: $x_j \ge 0$. Если

$$g_i\left(x_1, x_2, \ldots, x_n\right) = \sum_{j=1}^{n} a_{ij} x_j,\ i = \overline{1, m}, \tag{31}$$

$$f\left(x_1, x_2, ..., x_n\right) = \sum_{j=1}^{n} c_j x_j \,, \tag{32}$$

где a_{ij}, c_j – известные постоянные, то при условии неотрицательности решения $x_1, x_2, \ldots, x_n$ задачи (31), (32) (т. е. при $x_j \geq 0,\ j = \overline{1, n}$) получаем *задачу линейного программирования* (ЗЛП). Если хотя бы одна из функций $f(x_1, x_2, \ldots, x_n)$, $g_i(x_1, x_2, \ldots, x_n)$ нелинейная, то соответствующая задача является *задачей нелинейного программирования.*

На рис. 9 приведена классификация задач математического программирования (ЗМП).

Рис. 9. Классификация задач математического программирования

Наиболее изученным разделом МП является ***линейное программирование*** (ЛП). Для решения задач ЛП разработан целый ряд эффективных методов, алгоритмов и программ. Среди задач ***нелинейного программирования*** наиболее глубоко изучены *задачи выпуклого программирования*. Это задачи минимизации выпуклой (или максимизации вогнутой) функции, заданной на выпуклом замкнутом множестве. В свою очередь, среди задач выпуклого программирования наиболее подробно исследованы *задачи квадратичного программирования*. В результате решения таких задач требуется в общем случае найти максимум (минимум) квадратичной функции при условии, что ее переменные удовлетворяют некоторой системе линейных неравенств или линейных уравнений либо некоторой системе, содержащей как линейные неравенства, так и линейные уравнения. Отдельными классами ЗМП являются ***задачи целочисленного*** (ЗЦП), ***параметрического*** (ЗПП) и ***дробно-линейного*** (ЗДЛП) ***программирования***. В ЗЦП неизвестные могут принимать только целочисленные значения. В ЗПП целевая функция и/или функции, определяющие область возможных изменений переменных, зависят от некоторых параметров. В ЗДЛП целевая функция представляет собой отношение двух линейных функций; линейными являются и функции, определяющие область возможных изменений переменных.

Выделяют также отдельные классы задач ***стохастического*** и ***динамического программирования***. Если в целевой функции или в функциях, определяющих область возможных изменений переменных, содержатся случайные величины, то такая задача относится к задаче *стохастического программирования*. Задачи, процесс поиска решения которых является многоэтапным, относятся к задачам *динамического программирования*.

Типичная ЗМП в векторной форме может быть записана в виде

$$\begin{cases} f(x) \to \text{extr (min, max)}, \\ g(x) = 0, \\ h(x) \le 0, \end{cases} \tag{33}$$

где $f(x)$, $g(x)$, $h(x)$ – функции векторного аргумента $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$: $f : \mathbf{R}^n \to \mathbf{R}$, $g : \mathbf{R}^n \to \mathbf{R}^k$, $h : \mathbf{R}^n \to \mathbf{R}^{m-k}$, или более общо (ср. с задачей (30)):

$$\begin{cases} f(x) \to \text{extr}, \\ x \in X \subset \mathbf{R}^n. \end{cases}$$

Если ищутся только целочисленные значения компонент вектора x, то ЗМП (33) – *задача целочисленного программирования*. Если рассматриваются задачи оптимизации, где требуется одновременно найти экстремум нескольких целевых функций $f_i(x), i = \overline{1, p}$, то исследуемая задача относится к *многокритериальным задачам оптимизации*.

9.2. Задачи вариационного исчисления

К задачам вариационного исчисления (ЗВИ) [4–8, 14, 21, 24] относят задачи нахождения ***экстремумов функционалов*** при различных ограничениях (связях).

Если каждой функции $y = y(x)$, $x \in [a; b]$, из некоторой совокупности $Y = \{y = y(x),\ x \in [a; b]\}$ функций, заданных на отрезке [a; b], ставится в соответствие некоторое число $I(y)$, то говорят, что задан *функционал.*

Например, $I(y) = y^2(a) + 2$, $I(y) = \int\limits_a^b F(x, y(x), y'(x))dx$ – функционалы.

Простейшая ЗВИ состоит в нахождении такой *функции* $y = y(x) \in Y$, $x \in [a; b]$, которая доставляет экстремум функционалу $I(y)$ и для заданных чисел y_0, y_1 удовлетворяет граничным условиям $y(a) = y_0$, $y(b) = y_1$:

$$\begin{cases} I(y) = \int\limits_a^b F(x, y, y')dx \to \mathrm{extr}, \\ y(a) = y_0, \quad y(b) = y_1. \end{cases}$$

9.3. Задачи оптимального управления

Чаще всего задачи оптимального управления (ЗОУ) [5–7, 22] связаны с оптимизацией *функционалов* на *траекториях динамических систем.* Рассмотрим *задачу оптимального быстродействия.* Пусть имеется динамическая система (т. е. некоторый объект, поведение которого описывается, например, системой обыкновенных дифференциальных уравнений)

$$\frac{dx}{dt} = f(x(t), u(t), t), \quad t > t_0, \tag{34}$$

где $x(t) \in \mathbf{R}^n$ – *вектор состояния* объекта, $u(t) \in \mathbf{R}^r$ – *вектор управляющих воздействий (управление)*, $u(\cdot) \in U$, т. е. значения управляющего воздействия принадлежат множеству U кусочно-непрерывных r-вектор-функций, $f : \mathbf{R}^n \times \mathbf{R}^r \times \tilde{\mathbf{R}} \to \mathbf{R}^n$, $\tilde{\mathbf{R}} = \{t \in \mathbf{R},\ t > t_0\}$, f – заданная n-вектор-функция, t – время, $t \in \mathbf{R}$, $t > t_0$. В пространстве состояний $\mathbf{R}^n$ системы (34) заданы две «точки»: *начальное* $x(t_0) = x_0 \in \mathbf{R}^n$ и *конечное* $x(t_1) = x_1 \in \mathbf{R}^n$ положения системы (заметим, что конечный момент времени t_1 не задан). Требуется найти управление $u(t)$ такое, что траектория (решение) $x(t)$, порожденная в силу (34) этим управлением и начальным условием $x(t_0) = x_0$, попадет из начального положения x_0 в конечное положение x_1 за минимально возможное время:

$$I(u) = t_1 - t_0 \to \min.$$

Минимизируемый функционал $I(u)$ называется *целевой функцией* или *критерием качества* переходного процесса.

Многие ЗОУ решаются с помощью *принципа максимума Понтрягина* либо *методом динамического программирования Беллмана.*

10. ПРИМЕРЫ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ОПТИМИЗАЦИОННЫХ ЗАДАЧ

Математической моделью называется приближенное описание какого-либо явления или процесса внешнего мира, выраженного с помощью математической символики. Под *математическим моделированием* понимают изучение свойств объекта на математической модели. Цель математического моделирования состоит в исследовании каких-либо явлений, процессов или систем объектов путем построения и изучения их моделей, в использовании моделей для определения или уточнения характеристик и рационализации способов построения вновь конструируемых объектов. В модели должны быть учтены все наиболее существенные факторы, влияющие на исследуемый процесс, и вместе с тем она не должна быть загромождена второстепенными факторами, учет которых только усложнит анализ.

Модели ЗМП включают:

- *план задачи* – совокупность искомых переменных $x = (x_1, x_2, \ldots, x_n)$, которые удовлетворяют имеющимся ограничениям;
- *целевую функцию* (прибыль, объем выпуска или реализации продукции, затраты на производство, отходы и т. д.);
- *условия* (системы ограничений), *налагаемые на неизвестные величины* (в виде равенств и неравенств), совокупность которых образует область допустимых значений.

Ниже будут рассмотрены экономические модели, приводящие к задачам линейного программирования.

10.1. Задача о наилучшем использовании ресурсов

Пусть имеется предприятие, специализирующееся на выпуске n видов продукции, для производства которых используется m видов ресурсов. Заданы нормы расходов ресурсов на производство всех видов продукции, требуемые объемы выпускаемой продукции, а также прибыль от реализации единицы продукции каждого вида.

Задача. Организовать производство (т. е. составить план выпуска продукции) таким образом, чтобы при выполнении всех заданных ограничений предприятие получило максимальную прибыль.

Составим математическую модель задачи. Пусть $x_1, x_2, \ldots, x_n$ – *план* производства, показывающий, какие виды продукции необходимо производить и в каких количествах, $b_1, b_2, \ldots, b_m$ – имеющиеся объемы ресурсов, c_1, c_2, …, c_n – прибыль от реализации единицы продукции вида x_1, x_2, …, x_n; a_{ij} – расход i-го вида ресурсов на производство j-го вида продукции ($i = \overline{1, m}$, $j = \overline{1, n}$). В силу введенных обозначений *целевая функция* (прибыль от реализации продукции) будет иметь вид

$$z = c_1x_1 + c_2x_2 + \ldots + c_nx_n = \sum_{j=1}^{n} c_jx_j \to \max,$$

а ограничения на ресурсы –

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

при этом по смыслу введенных переменных x_j, $j = \overline{1, n}$, следует

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

В векторно-матричной форме математическая модель задачи примет вид

$$z = c'x \to \max,$$

$$Ax \le b, x \ge 0,$$

где

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ c_2 \\ \ldots \\ c_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

знак «'» означает транспонирование, т. е. если, например, c – вектор-столбец, то $c' = (c_1, c_2, \ldots, c_n)$ – вектор-строка. Полученная математическая модель относится к моделям ЗЛП.

10.2. Задача о смесях (о рационе, о выборе диеты)

Пусть имеется n пищевых продуктов № 1, 2, 3, …, n, в которых содержится m различных питательных веществ № 1, 2, …, m. Единица j-го продукта содержит a_{ij} единиц i-го питательного вещества. Для нормальной жизнедеятельности в заданный промежуток времени t необходимо употреблять не менее b_i единиц питательного вещества i-го вида. Пусть c_j – стоимость единицы продукта j-го вида.

Задача. Составить рацион минимальной стоимости, содержащий необходимое количество питательных веществ.

Составим математическую модель задачи. *Планом* задачи является количество $x_j,\ j=\overline{1,n}$, продуктов № 1, 2, 3, …, n каждого вида, которое должно составить рацион в заданный промежуток времени t. Тогда математическая модель задачи имеет вид

$$z = c_1x_1 + c_2x_2 + \ldots + c_nx_n = \sum_{j=1}^{n} c_jx_j \to \min$$

– стоимость рациона,

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

– ограничения по содержанию питательных веществ в рационе,

$$x_j \geq 0, \quad j=\overline{1,n},$$

– использование (> 0) или отсутствие (= 0) продуктов x_j, $j=\overline{1,n}$, в рационе. Таким образом, математическая модель задачи о смесях (рационе, выборе диеты) также представляет собой ЗЛП. В векторно-матричной форме эта задача имеет вид $z = c'x \to \min$, $Ax \geq b$, $x \geq 0$.

10.3. Транспортная задача

Имеется m предприятий $A_1, A_2,\ldots, A_m$, на которых производится a_i, $i=\overline{1,m}$, единиц однородной продукции, и n пунктов $B_1, B_2,\ldots, B_n$ ее потребления с потребностями b_j, $j=\overline{1,n}$, соответственно. Известна стоимость c_{ij} (затраты) перевозки единицы продукции из i-го пункта производства в j-й пункт потребления.

Задача. Определить такой план перевозок продукции по пунктам ее потребления, при котором весь продукт из пунктов производства будет вывезен, спрос всех потребителей удовлетворен, а транспортные расходы минимальны.

Составим математическую модель задачи. Рассмотрим *план* задачи в виде *матрицы планирования* (матрицы перевозок) $X = (x_{ij})$, $i=\overline{1,m}$, $j=\overline{1,n}$, где x_{ij} количество продукции, перевозимой из i-го пункта ее производства в j-й пункт ее потребления. Тогда $c_{ij}x_{ij}$ – стоимость перевозки продукта от i-го производителя к j-му потребителю. Математическая модель задачи принимает вид:

$$z = c_{11}x_{11} + c_{12}x_{12} + \ldots + c_{m,n-1}x_{m,n-1} + c_{mn}x_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \to \min,$$

т. е. требуется минимизировать целевую функцию z при условии, что вся продукция из пунктов производства вывезена:

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

все запросы потребителей удовлетворены:

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

и при этом

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

Следовательно, математическая модель транспортной задачи также представляет собой ЗЛП.

10.4. Задачи целочисленного программирования

К данному типу задач [11, 14, 15] относится *задача о контейнерных перевозках.* Контейнер оборудован m отсеками вместимостью b_i, $i = \overline{1, m}$, для перевозки n видов продукции x_j. Все виды продукции характеризуются неделимостью (т. е. их можно брать в количестве 0, 1, 2, … единиц). Пусть x_j – количество единиц j-го вида продукции, погруженной в контейнер, a_{ij} – используемый объем i-го отсека для перевозки j-го вида продукции, c_j – прибыль от перевозки единицы j-го вида продукции.

Задача. Найти план $x_1, x_2, \ldots, x_n$ перевозки, при котором общая прибыль от рейса максимальна.

Составим математическую модель задачи. *Целевая функция* задачи имеет вид $\sum_{j=1}^{n} c_j x_j \to \max$, *основные ограничения*: $\sum_{j=1}^{n} a_{ij}x_j = b_i, i = \overline{1, m},$ *прямые ограничения:* $x_j \geq 0$. При этом должно выполняться условие целочисленности перевозимой продукции x_j, $j = \overline{1, n}$.

10.5. Типичная задача вариационного исчисления

Такой задачей является [4, 6, 8, 14, 21, 24] *задача о брахистохроне* (от греческого: *брахистос* – наискорейший, *хронос* – время). В вертикальной плоскости заданы две точки A, B, расположенные

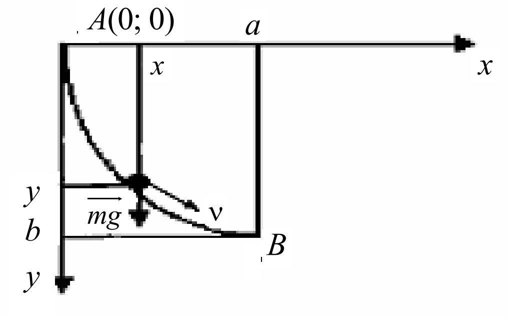


Рис. 10. Задача о брахистохроне

на разных уровнях (рис. 10). Требуется соединить их гладкой кривой так, чтобы, двигаясь по ней только под действием силы тяжести с нулевой начальной скоростью, материальная точка массой m прошла путь от A до B за минимальное время (трением и сопротивлением среды пренебрегаем).

Составим математическую модель задачи. Введем декартову систему координат, поместив начало координат в точку A. По закону сохранения энергии суммы потенциальной и кинетической энергий в точках A и B (и в любой точке $(x; y)$ кривой) равны, следовательно,

$$mgy = \frac{mv^2}{2} \Rightarrow v = \sqrt{2gy}. \qquad (35)$$

Пусть $y = y(x), x \in [0; a]$, – уравнение искомой гладкой линии, соединяющей точки A и B. Известно, что если s – путь, t – время, то скорость и длина пути определяются по формулам

$$v = \frac{ds}{dt}, \quad ds = \sqrt{dx^2 + dy^2}, \qquad (36)$$

где $dy = y'_x dx$. Подставив (35) в (36), получим

$$\sqrt{2gy} = \frac{\sqrt{dx^2 + dy^2}}{dt} \Rightarrow \sqrt{2gy} = \frac{\sqrt{1 + y_x'^2}}{dt} dx \Rightarrow dt = \frac{\sqrt{1 + y_x'^2}}{\sqrt{2gy}} dx.$$

Проинтегрировав последнее соотношение от 0 до a, найдем, что время движения от точки A до точки B по линии $y = y(x)$, равно

$$T\left(y(x)\right) = \frac{1}{\sqrt{2g}} \int_0^a \sqrt{\frac{1 + y'_x(x)^2}{y(x)}} dx.$$

Таким образом, задача о брахистохроне свелась к поиску такой *дифференцируемой функции* $y(x)$, $0 \le x \le a$, которая на концах интервала $0 \le x \le a$ принимает заданные значения $y(0) = 0$, $y(a) = b$ и доставляет минимум функционалу $T(y(x))$.

Решением задачи о брахистохроне, как будет показано в подразд. 17.3, является *дуга циклоиды*. Эта линия называется *брахистохроной*. На практике к такому решению пришли строители зданий в тропических странах (при строительстве буддийских пагод), где в условиях затяжных дождей быстрейший скат воды с крыши существенно влиял на ее долговечность.

В отличие от задач, рассмотренных ранее, в задаче о брахистохроне находится не конечномерный вектор, минимизирующий функцию, а *функция*, на которой достигается минимум *функционала*, т. е. задача

$$T\big(y(x)\big) = \frac{1}{\sqrt{2g}} \int_0^a \sqrt{\frac{1 + y_x'(x)^2}{y(x)}}\,dx \to \min,\; y(0) = 0,\; y(a) = b,$$

рассматривается ***на элементах*** $y(\cdot)$ ***функционального пространства***.

● **Задача 15.** Для производства чугунного литья используется n различных шихтовых материалов (чугун различных марок, стальной лом, феррофосфор и др.). Химический состав чугунного литья определяется содержанием в нем химических элементов (кремния, марганца, фосфора и др.). Готовый чугун должен иметь строго определенный химический состав, который задается величинами H_i, представляющими собой доли (в процентах) j-го химического элемента в готовом продукте. При этом известны величины: h_{ij} – содержание (в процентах) j-го химического элемента в i-м исходном шихтовом материале; c_i – цена единицы i-го шихтового материала.

Составьте математическую модель определения состава шихты, обеспечивающего получение литья заданного качества при минимальной общей стоимости используемых шихтовых материалов.

11. ПОСТАНОВКА ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Распространены два вида записи задач линейного программирования: *ЗЛП в нормальной форме* и *ЗЛП в канонической форме*.

ЗЛП в нормальной форме называется задача максимизации (минимизации) линейной функции (линейной формы)

$$z(x) = z(x_1, x_2, \ldots, x_n) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max\ (\min) \qquad (37)$$

по переменным $x_1, x_2, \ldots, x_n$, удовлетворяющим неравенствам:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m; \end{cases} \qquad (38)$$

$$x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0, \qquad (39)$$

где a_{ij}, b_i, c_j – заданные действительные числа, $i = \overline{1, m}$, $j = \overline{1, n}$. Ограничения (38) называются *основными*, а (39) – *прямыми ограничениями* ЗЛП (37)–(39). ЗЛП (37)–(39) может быть записана в более компактной форме:

$$z(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} c_j x_j \to \max\ (\min); \tag{37*}$$

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = \overline{1, m}; \tag{38*}$$

$$x_j \ge 0, \quad j = \overline{1, n}\ . \tag{39*}$$

ЗЛП в канонической форме имеет вид:

$$z(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} c_j x_j \to \max\ (\min); \tag{40}$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = \overline{1, m}; \tag{41}$$

$$x_j \ge 0, \quad j = \overline{1, n}\ . \tag{42}$$

Введем в рассмотрение векторы

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}, c = \begin{pmatrix} c_1 \\ c_2 \\ \ldots \\ c_n \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}, A_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \ldots \\ a_{mj} \end{pmatrix}, \ j = \overline{1, n}, \tag{43}$$

где c – *вектор стоимостей*, b – *вектор ограничений*, A_1, A_2, …, A_n – *векторы-столбцы условий*,

$$A = [A_1, A_2, \ldots, A_n] = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

– *матрица условий* (матрица затрат). Тогда ЗЛП (37)–(39) может быть переписана в *векторном виде*:

$$z(x) = c'x \to \max\ (\min); \tag{44}$$

$$A_1 x_1 + A_2 x_2 + \ldots + A_n x_n \le b, \tag{45}$$

$$x \ge 0, \tag{46}$$

где штрих «'» в (44) означает операцию транспонирования. Задача (44)–(46) может быть представлена в более компактной *векторно-матричной форме*:

$$z(x) = c'x \to \max\ (\min); \tag{44*}$$

$$Ax \le b; \tag{45*}$$

$$x \ge 0, \tag{46*}$$

где $x \in \mathbf{R}^n$, $c \in \mathbf{R}^n$, $A \in \mathbf{R}^{m \times n}$, $\mathbf{R}^{m \times n}$ – множество действительных ($m \times n$) – матриц, а неравенство $x \geq \underline{0}$ означает неотрицательность всех компонент n-вектора x: $x_j \geq 0,\ j = \overline{1, n}$.

Обсуждение. Две формы записи ЗЛП (*нормальная* и *каноническая*) отличаются лишь типом *основных* ограничений. В нормальной форме основные ограничения заданы неравенствами (38) (или, что то же самое, (38*), (45), (45*)), в канонической форме – равенствами (41). Нетрудно заметить, что ***любое ограничение в виде неравенства путем введения дополнительной неотрицательной переменной можно свести к ограничению типа равенства и наоборот.***

Действительно, пусть дано неравенство

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

Путем введения дополнительной неотрицательной переменной $x_{n+i} \geq 0$ из него можно получить равенство $a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n + x_{n+i} = b_i$. И наоборот, каждое ограничение типа равенства

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

можно записать в виде двух ограничений типа неравенства:

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

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

Поскольку последнее соотношение эквивалентно

$$-(a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n) \leq -b_i,$$

то исходное ограничение типа равенства свелось к двум ограничениям типа неравенства со знаком $\leq$:

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

$$-(a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n) \leq -b_i\,,$$

т. е. к ограничению вида (38). Поэтому ЗЛП в нормальной и канонической формах эквивалентны.

Нетрудно также видеть, что ***задачу минимизации целевой функции можно заменить задачей максимизации и наоборот***. Действительно, поскольку для любой функции $y = f(x)$ (рис. 11) справедливо равенство $\min f(x) = -\max(-f(x))$, то для функции $z(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} c_j x_j$ ее минимум равен максимуму функции $-z(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n}(-c_j)x_j$, взятому с противоположным знаком. Заметим, что оба экстремума достигаются при одних и тех же значениях переменных x, т. е. $\min z(x) = -\max(-z(x))$.

ЗЛП может содержать ограничения в виде линейных равенств и неравенств одновременно; возможно, что некоторые переменные не подчинены условию неотрицательности. В последнем случае каждую из таких переменных заменяют разностью двух новых переменных, на которые налагают условие неотрицательности. Если, например, переменная $x_k \le 0$, то всегда можно выбрать такие $x'_k \ge 0$, $x''_k \ge 0$, что $x_k \le 0$ представимо в виде $x_k = x'_k - x''_k$.

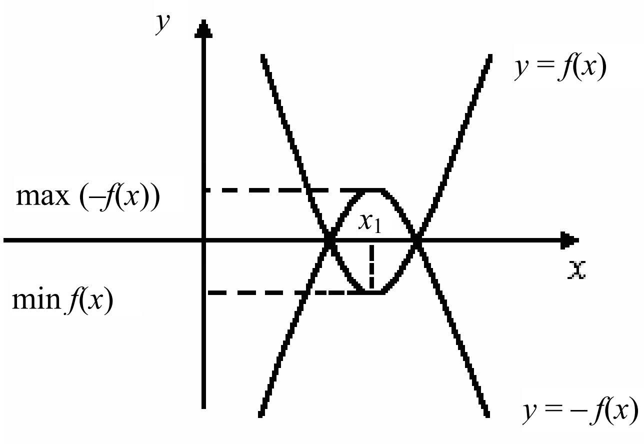


Рис. 11. Интерпретация задачи на экстремум:
$\min f(x) = -\max(-f(x))$

Пример 12. Имеется два вида химического оборудования: А и Б. Каждое оборудование выпускает в смену 10 единиц химической продукции. Для обслуживания оборудования А в смену требуется 3 человека и 100 кВт-ч электроэнергии, для обслуживания оборудования Б – 4 человека и 60 кВт-ч электроэнергии. Трудовые ресурсы составляют 90 человек, ресурсы электроэнергии – 2000 кВт-ч, затраты на эксплуатацию оборудования типа А – 12 денежных единиц в смену, для Б – 10 денежных единиц в смену. Запланирован выпуск 250 единиц продукции. Требуется выполнить план выпуска химической продукции с минимальными затратами. Составить математическую модель задачи и привести ее к канонической форме.

Решение. Пусть x_1 – количество используемого в смену оборудования типа А; x_2 – количество используемого в смену оборудования типа Б. Тогда целевая функция задачи будет иметь вид

$$z(x_1, x_2) = 12\,x_1 + 10\,x_2 \to \min.$$

Ограничения по трудовым ресурсам можно выразить неравенством $3x_1 + 4x_2 \le 90$, ограничения по использованию электроэнергии: $100x_1 + 60x_2 \le 2000$. Требование выполнения плана по выпуску продукции: $10x_1 + 10x_2 = 250$. Очевидно, что переменные x_1, x_2 должны удовлетворять условиям $x_1 \ge 0$, $x_2 \ge 0$. Таким образом, математическая модель *задачи* имеет вид

$$z(x_1, x_2) = 12x_1 + 10x_2 \to \min;$$

$$\begin{cases} 3x_1 + 4x_2 \le 90, \\ 100x_1 + 60x_2 \le 2000, \\ 10x_1 + 10x_2 = 250; \end{cases}$$

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

Сведем задачу к канонической форме. Для этого добавим к левым частям первого и второго неравенств основных ограничений неотрицательные переменные x_3 и x_4 соответственно. В результате получим *ЗЛП в канонической форме*:

$$z(x_1, x_2) = -12x_1 - 10x_2 + 0x_3 + 0x_4 \to \max;$$

$$\begin{cases} 3x_1 + 4x_2 + x_3 = 90, \\ 100x_1 + 60x_2 + x_4 = 2000, \\ 10x_1 + 10x_2 = 250; \end{cases}$$

$$x_1 \ge 0, x_2 \ge 0, x_3 \ge 0, x_4 \ge 0. \ \blacktriangle$$

12. ГЕОМЕТРИЧЕСКИЙ МЕТОД РЕШЕНИЯ ЗЛП

Рассмотрим ЗЛП (37)–(39) в нормальной форме. Каждый вектор $x \in \mathbf{R}^n$, удовлетворяющий ограничениям (45*), (46*), называется *планом* (допустимым решением) ЗЛП (44*)–(46*). Множество

$$X = \left\{ x \in \mathbf{R}^n : \ Ax \le b, \ x \ge 0 \right\} \tag{47}$$

называется *множеством планов* (допустимых решений) ЗЛП (44*)–(46*). План $x^0 \in X$, доставляющий максимум целевой функции ($c'x^0 = \max c'x$), называется *оптимальным планом* ЗЛП (44*)–(46*).

ЗЛП с двумя переменными ($n = 2$) всегда можно решить геометрически. При $n = 3$ решение ЗЛП задачи существенно усложняется, а при $n > 3$ геометрическое решение, вообще говоря, невозможно.

Рассмотрим *геометрический* (графический) *метод* решения ЗЛП при $n = 2$:

$$z(x_1, x_2) = c_1x_1 + c_2x_2 \to \max\ (\min); \tag{48}$$

$$a_{i1}x_1 + a_{i2}x_2 \le b_i, \quad i = \overline{1, m}; \tag{49}$$

$$x_1 \ge 0, x_2 \ge 0. \tag{50}$$

Дадим геометрическую интерпретацию элементов этой задачи. Область *D допустимых планов* задачи (48)–(50) образуется пересечением m множеств, каждое из которых определяется соответствующим неравенством вида $a_{i1}x_1 + a_{i2}x_2 \le b_i, \quad i = \overline{1, m},$ и представляет собой полуплоскость, лежащую по одну сторону от прямой

$$a_{i1}x_1 + a_{i2}x_2 = b_i, \quad i = \overline{1, m}. \tag{51}$$

Чтобы определить, какую полуплоскость задает неравенство (49) для каждого i, $i = \overline{1, m}$, достаточно взять произвольную точку, не лежащую на граничной прямой (51), и проверить, удовлетворяют ли координаты этой точки неравенству (49). Если удовлетворяют, то (49) определяет полуплоскость, содержащую данную точку. Если не удовлетворяют, то неравенство (49) определяет полуплоскость, не содержащую данную точку. Часто в качестве такой точки берут начало координат $O(0;\ 0)$. Нужная полуплоскость помечается с помощью нанесения *штриховки* у ее границы. Пересечение D указанных полуплоскостей образует многоугольную область, которая является *выпуклым множеством*. *Выпуклым множеством* называется множество, которое вместе с любыми своими точками x_1 и x_2 содержит и все точки отрезка $[x_1;\ x_2]$, т. е. точки вида $x = \lambda x_1 + (1 - \lambda)x_2$, $\lambda \in [0; 1]$.

Область D может оказаться пустым, ограниченным или неограниченным множеством (рис. 12).

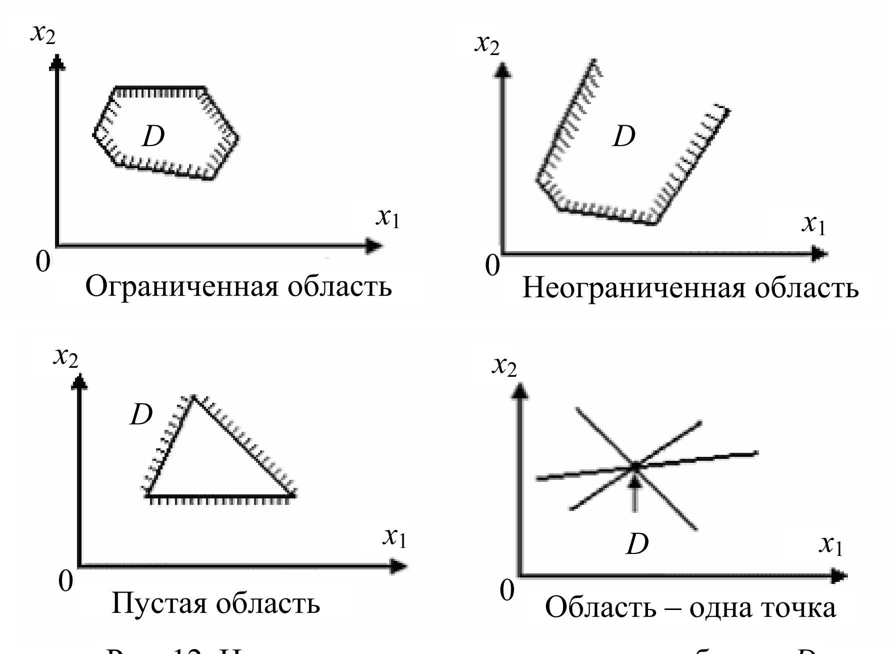


Рис. 12. Некоторые возможные варианты области D

Область D называют *многоугольником допустимых решений* ЗЛП (48)–(50), а вершины многоугольника – *крайними* (угловыми) точками.

Рассмотрим целевую функцию $z(x_1, x_2) = c_1x_1 + c_2x_2$ ЗЛП (48)–(50). *Линии уровня* (линии постоянного значения) целевой функции $z(x_1, x_2)$ – это линии, описываемые уравнением вида

$$c_1x_1 + c_2x_2 = C, \tag{52}$$

которые (C – постоянная) образуют семейство параллельных прямых. Чтобы установить направление возрастания (или убывания) целевой функции z, находим ее частные производные: $\frac{\partial z}{\partial x_1} = c_1$, $\frac{\partial z}{\partial x_2} = c_2$, указывающие скорость возрастания вдоль соответствующих осей. Тогда вектор с координатами c_1, c_2 является *градиентом* функции z. Он показывает направление *наискорейшего возрастания* целевой функции и служит вектором нормали к прямым $c_1x_1 + c_2x_2 = C$. Вектор $(–c)$ (*антиградиент*) указывает направление *наискорейшего убывания* целевой функции.

Геометрическая интерпретация ЗЛП. ЗЛП (48)–(50) представляет собой отыскание такой точки области D, через которую проходит прямая (52), соответствующая наибольшему (при решении задачи на максимум) значению функции z. Тогда выберем из семейства прямых (52) прямую, имеющую общую точку с областью D, и будем смещать ее в направлении вектора $c = \text{grad}z$. Найдется такое предельное положение прямой, когда D окажется по одну сторону от прямой и будет иметь с D хотя бы одну общую точку. Тогда все точки области D, лежащие на предельной прямой, будут решениями ЗЛП. Полученная предельная прямая называется *опорной прямой* к области D.

Алгоритм геометрического метода решения ЗЛП:

1. Строим многоугольник (многоугольную область) допустимых решений ЗЛП.

2. Строим вектор $c = \text{grad}z$ и одну из прямых семейства (52), в частности прямую $z = 0$: $c_1x_1 + c_2x_2 = 0$, проходящую через начало координат и перпендикулярную вектору $\text{grad}z$.

3. Прямую $c_1x_1 + c_2x_2 = 0$ перемещаем параллельно самой себе (т. е. рассматриваем семейство $c_1x_1 + c_2x_2 = C$) в направлении вектора $\text{grad}z$ при решении задачи на максимум (или $(–\text{grad}z)$ в задаче на минимум) до того момента, когда при некотором $C = C_0$ область D окажется полностью расположенной по одну сторону от прямой $c_1x_1 + c_2x_2 = C_0$, причем будет иметь с перемещаемой прямой по крайней мере одну общую точку. Возможны следующие случаи:

а) *прямая $c_1x_1 + c_2x_2 = C_0$ проходит через вершину области D* и имеет с ней единственную общую точку $A_{\max}$ ($B_{\min}$ в задаче на минимум). Тогда ЗЛП разрешима и имеет единственное решение;

б) *прямая* $c_1x_1 + c_2x_2 = C_0$ *совпадает с одной из сторон области D.* В этом случае ЗЛП разрешима и имеет бесконечное множество решений: целевая функция достигает одного и того же экстремального значения во всех точках области D, лежащих на прямой $c_1x_1 + c_2x_2 = C_0$. В этом случае говорят, что имеет место *альтернативный оптимум*;

в) при любом значении *C прямая* $c_1x_1 + c_2x_2 = C_0$ *имеет общие точки с областью D.* В этом случае ЗЛП не имеет решения, точнее, целевая функция не ограничена на множестве планов: $z_{\max} = +\infty (z_{\min} = -\infty$ в ЗЛП на минимум).

4. В случае «3а» находим крайнюю точку $A_{\max}$ ($B_{\min}$), в которой достигается максимум (минимум) целевой функции (48). Ее координаты x_1^0, x_2^0 определяют оптимальный план $x^0 = (x_1^0, x_2^0)'$ и экстремальное значение целевой функции $z_{\max}(A) = c'x^0$ (или $z_{\min}(B) = c'x^0$); в случае «3б» достаточно найти координаты какой-либо точки отрезка прямой $c_1x_1 + c_2x_2 = C_0$, совпадающего с одной из сторон области D, на котором (в каждой точке которого) достигается максимум (минимум) целевой функции (48). Координаты x_1^0, x_2^0 точки определят *альтернативный оптимум*, т. е. оптимальный план $x^0 = (x_1^0, x_2^0)'$ и экстремальное значение $c'x^0$ целевой функции.

Задача о рационе. При откорме каждое животное ежедневно должно получать не менее 9 единиц питательного вещества S_1, не менее 8 единиц питательного вещества S_2 и не менее 12 единиц питательного вещества S_3. Для составления рациона используют два вида кормов. Содержание количества единиц каждого питательного вещества в 1 кг каждого вида кормов и стоимость 1 кг корма приведены в табл. 5.

Таблица 5

Показатель на 1 кг корма	Корм		Ежедневная норма
	I	II	
Количество питательных единиц S_1	3	2	≥9
S_2	1	2	≥8
S_3	1	6	≥12
Стоимость, тыс. руб.	4	6	–

Требуется составить дневной рацион с минимальными денежными затратами.

Решение. Составим математическую модель задачи. Пусть x_1 и x_2 – количество килограммов корма I и II вида в дневном рационе. На основании условий задачи получим систему ограничений:

$$\begin{cases} 3x_1 + 2x_2 \ge 9, \\ x_1 + 2x_2 \ge 8, \\ x_1 + 6x_2 \ge 12. \end{cases}$$

Если корм I не используется в рационе, то $x_1 = 0$ в противном случае $x_1 > 0$. Аналогично имеем $x_2 \ge 0$. Следовательно, прямые ограничения задачи имеют вид $x_1 \ge 0$, $x_2 \ge 0$. Общая стоимость дневного рациона

$$z(x_1, x_2) = 4x_1 + 6\,x_2 \to \min.$$

Так как $n = 2$, то задачу можно решить геометрическим методом. Построим многоугольную область D допустимых решений. Для этого на плоскости x_1Ox_2 изобразим граничные прямые

$$l_1\!: 3x_1 + 2x_2 = 9,\;\; l_2\!: x_1 + 2x_2 = 8,\;\; l_3\!: x_1 + 6x_2 = 12.$$

Получим неограниченную многоугольную область с угловыми точками A, B, C, E (рис. 13). Построим вектор-градиент целевой функции $\mathrm{grad}z = \vec{c}$ с координатами $(4; 6)$.

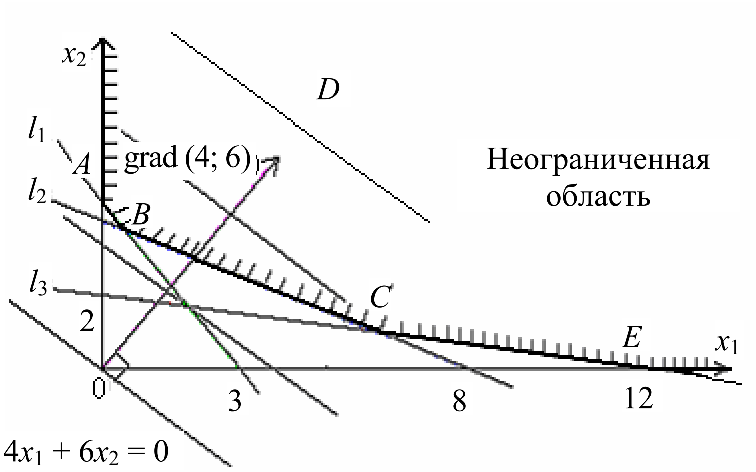


Рис. 13. Область допустимых значений в задаче о рационе

Прямую вида $4x_1 + 6x_2 = C$, параллельную прямой $4x_1 + 6x_2 = 0$ и имеющую общие точки с D перемещаем в направлении антиградиента целевой функции. Из рис. 13 видно, что минимальное значение целевой функции достигается в точке B – точке пересечения прямых l_1 и l_2:

$$B\!: \begin{cases} 3x_1 + 2x_2 = 9, \\ x_1 + 2x_2 = 8, \end{cases}$$

откуда находим $x_1 = 0{,}5$, $x_2 = 3{,}75$. Следовательно, точка B имеет координаты (0,5; 3,75). Таким образом, дневной рацион должен состоять из 0,5 кг корма I вида и 3,75 кг корма II вида, при этом минимальные денежные затраты составят $z_{\min}(B) = 4 \cdot 0{,}5 + 6 \cdot 3{,}75 = 24{,}5$ тыс. руб.

При возникновении сомнений, в какой именно точке достигается экстремум (на рис. 13 точки A, B), необходимо вычислить значение целевой функции во всех точках, относительно которых есть сомнения.

Найдем значение целевой функции в точке A пересечения прямой l_1 и оси Ox_2.

$$A: \begin{cases} 3x_1 + 2x_2 = 9, \\ x_1 = 0, \end{cases} \Rightarrow x_1 = 0,\ \ x_2 = 4{,}5, \quad z(A) = 4 \cdot 0 + 6 \cdot 4{,}5 = 27 > z(B).$$

Так как $z(A) > z(B)$, то точка $A(0;\ 4{,}5)$ не является решением задачи о рационе, поскольку не обеспечивает минимум денежных затрат.

Ответ: суточный рацион состоит из 0,5 кг корма первого типа и 3,75 кг второго типа. Минимальная стоимость рациона $z_{\min}(0{,}5;\ 3{,}75) = 24{,}5$ тыс. руб. ▲

● **Задача 16.** Сформулировать задачу о рационе для случая m видов питательных веществ и n видов кормов.

Замечание 9. С помощью геометрического метода может быть решена ЗЛП, система ограничений которой содержит n неизвестных и m линейно независимых уравнений, причем $n - m = 2$.

Теорема 2 (*основная теорема линейного программирования*). *Если ЗЛП имеет решение, то целевая функция достигает экстремального значения хотя бы в одной из угловых (крайних) точек многогранника допустимых решений (планов). Если же целевая функция достигает экстремального значения в различных угловых точках, то она достигает того же значения в любой точке, являющейся их выпуклой линейной комбинацией.*

В ЗЛП, система ограничений которой представлена на рис. 14, $z_{\min}$ достигается в единственной точке C, а $z_{\max}$ – на отрезке AE. В последнем случае говорят, что имеет место *альтернативный оптимум*. Значение $z_{\max}$ целевой функции в любой точке X отрезка AE одно и то же: $z_{\max}(A) = z_{\max}(E) = z_{\max}(X)$.

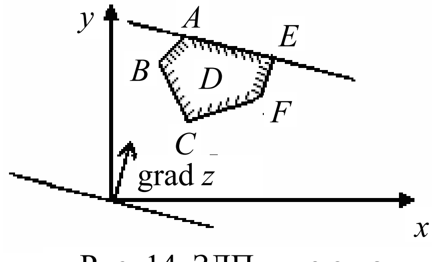


Рис. 14. ЗЛП, имеющая «альтернативный оптимум»

• **Задача 17.** Решить ЗЛП геометрическим методом:

$$z(x_1, x_2) = -x_1 + 2x_2 \to \min (\max);$$

$$\begin{cases} 3x_1 - x_2 \ge 8, \\ -x_1 + 2x_2 \le 8, \\ 0 \le x_1 \le 10; \end{cases}$$

$$x_2 \ge 0.$$

13. СИМПЛЕКС-МЕТОД РЕШЕНИЯ ЗЛП

В связи с основной теоремой ЛП возникает идея решения ЗЛП путем упорядоченного перебора угловых точек многоугольника планов (в общем случае – многогранной области) и сравнения значений целевой функции в этих точках. Простой перебор всех угловых точек даже с относительно небольшим числом переменных и ограничений практически неосуществим, так как требует большого числа вычислений. Одним из универсальных методов решения ЗЛП является ***симплекс-метод*** (*метод последовательного улучшения плана*). Симплекс-метод позволяет вести расчеты как вручную, так и на ЭВМ. В нем реализуется целенаправленный переход по ребрам многогранника из вершины в соседнюю вершину в направлении наискорейшего возрастания целевой функции в задаче на максимум. Этот метод применяется к ЗЛП в *канонической форме* при условии, что известен *базисный* план.

Планом или *допустимым решением* ЗЛП в канонической форме

$$z(x) = c'x \to \max (\min); \tag{53}$$

$$Ax = b; \tag{54}$$

$$x \ge 0, \tag{55}$$

где $x \in \mathbf{R}^n$, $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, называют вектор $x \in \mathbf{R}^n$, удовлетворяющий ограничениям (54), (55). При этом, не ограничивая общности, полагают $b \ge 0$, $\mathrm{rank}A = m$. Действительно, предполагая противное, т. е. $\mathrm{rank}A < m$, можно заключить, что либо в ограничениях-равенствах (54) ЗЛП (53)–(55) есть уравнения-следствия и их можно исключить, либо система ограничений (54) несовместна (не имеет решения). В последнем случае ЗЛП (53)–(55) не имеет допустимых решений.

План $x \in \mathbf{R}^n$ ЗЛП (53)–(55) называют *базисным*, если у него $(n - m)$ компонент (координат) равны нулю, а остальные m компонент соответствуют линейно-независимым векторам условий.

Базисный план называют *невырожденным*, если он содержит ровно m положительных компонент (например, $x = (x_1, x_2, ..., x_m, 0, ..., 0)'$, $x_i > 0,\ i = \overline{1, m}$), и *вырожденным* – в противном случае. Первые m компонент называются *базисными*, а остальные ($n - m$) – *небазисными* (или свободными) переменными. Базисный план часто называют *опорным планом.*

Оптимальным планом ЗЛП называется план, доставляющий экстремум (максимальное или минимальное значение) целевой функции.

Решение ЗЛП симплекс-методом включает в себя три этапа:

1) построение первоначального базисного плана;

2) проверка плана на оптимальность;

3) в случае неоптимальности плана – переход к новому базисному плану.

Последовательно рассмотрим все этапы.

13.1. Построение первоначального базисного плана

Рассмотрим процедуру построения первоначального базисного плана в зависимости от вида основных ограничений ЗЛП. Исследуем три случая.

1. Пусть среди векторов ограничений ЗЛП имеется единичный базис:

$$\begin{aligned} x_1 \qquad\qquad &+ a_{1,m+1}x_{m+1} + \ldots + a_{1,n}x_n = b_1, \\ x_2 \qquad &+ a_{2,m+1}x_{m+1} + \ldots + a_{2,n}x_n = b_2, \\ &\ldots, \\ x_m &+ a_{m,m+1}x_{m+1} + \ldots + a_{m,n}x_n = b_m, \end{aligned} \tag{56}$$

где $x_j \geq 0, j = \overline{1, n}, b_i \geq 0, i = \overline{1, m}$. Система ограничений вида (56) называется *системой в предпочтительном виде*. Здесь базис образуют векторы

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix},\ A_2 = \begin{pmatrix} 0 \\ 1 \\ \ldots \\ 0 \end{pmatrix},\ \ldots,\ A_m = \begin{pmatrix} 0 \\ \ldots \\ 0 \\ 1 \end{pmatrix}, \text{ т. е. } A_i = \begin{pmatrix} 0 \\ \ldots \\ 1 \\ \ldots \\ 0 \end{pmatrix} \leftarrow i,\ i = \overline{1, m}. \tag{57}$$

Переменные $x_1, x_2, \ldots, x_m$, соответствующие базисным векторам $A_1, A_2, \ldots, A_m$, являются *базисными переменными* (БП), остальные компоненты $x_{m+1}, \ldots, x_n$ – *свободными переменными* (СП).

Так как $b \geq 0$, то, полагая в (56) СП равными нулю (т. е. $x_{m+1} = x_{m+2} = \ldots = x_n = 0$), получаем значение для БП: $x_1 = b_1, x_2 = b_2, \ldots, x_m = b_m$, т. е. $x_i = b_i$, $i = \overline{1, m}$. Тогда $x^1 = (b_1, b_2, \ldots, b_m, 0, \ldots, 0)'$ – *первоначальный базисный план* для задачи (53)–(56).

2. Пусть ЗЛП задана в нормальной форме, т. е. система ограничений (54) имеет вид $Ax \leq b$, где $b \geq 0$. Представим эти ограничения в координатной форме

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i, \; i = \overline{1, m}. \tag{58}$$

С помощью введения *дополнительных* неотрицательных *переменных* $x_{n+1} \geq 0$, $x_{n+2} \geq 0$, …, $x_{n+m} \geq 0$ приведем (58) к канонической форме:

$$\begin{cases} \sum\limits_{j=1}^{n} a_{1j} x_j + x_{n+1} & = b_1, \\ \sum\limits_{j=1}^{n} a_{2j} x_j \qquad + x_{n+2} & = b_2, \\ \ldots & \\ \sum\limits_{j=1}^{n} a_{mj} x_j \qquad\qquad + x_{n+m} & = b_m. \end{cases} \tag{59}$$

Очевидно, здесь переменные $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$ являются БП, а все остальные – СП. Полагая в (59) СП $x_1 = x_2 = \ldots = x_n = 0$, получаем значения БП: $x_{n+1} = b_1$, $x_{n+2} = b_2$, …, $x_{n+m} = b_m$. Тогда из системы ограничений (59) следует, что *первоначальный базисный план* ЗЛП (53), (55), (58) имеет вид

$$x^1 = \left(\underbrace{0,\ 0,\ \ldots,\ 0}_{n},\ b_1,\ b_2,\ \ldots,\ b_m \right)'.$$

3. Пусть ЗЛП задана в нормальной форме, а система основных ограничений имеет вид

$$Ax \geq b, \; b \geq 0. \tag{60}$$

Ограничения (60), записанные в координатной форме, $\sum\limits_{j=1}^{n} a_{ij} x_j \geq b_i$, $i = \overline{1, m}$, приведем к канонической форме с помощью введения дополнительных неотрицательных переменных $x_{n+i} \geq 0$, $i = \overline{1, m}$:

$$\begin{cases} \sum_{j=1}^{n} a_{1j}x_j - x_{n+1} = b_1, \\ \sum_{j=1}^{n} a_{2j}x_j \quad - x_{n+2} = b_2, \\ \dots \\ \sum_{j=1}^{n} a_{mj}x_j \quad - x_{n+m} = b_m. \end{cases} \tag{61}$$

Тогда целевая функция $z(x)$ в (53) для ЗЛП (53), (61), (55) примет вид

$$z(x, x_{n+1}, x_{n+2}, \dots, x_{n+m}) = c'x + 0x_{n+1} + 0x_{n+2} + \dots + 0x_{n+m} \to \max (\min),$$

а прямые ограничения – $x_j \geq 0$, $j = \overline{1, n}$, $x_{n+1} \geq 0$, $x_{n+2} \geq 0$, …, $x_{n+m} \geq 0$. В общем случае система ограничений (61) не содержит единичного базиса и не является системой в предпочтительном виде. В этом случае для решения ЗЛП вводят *искусственный базис* и переходят к *М-задаче*. Вместо целевой функции вида (53) рассматривают функцию $\tilde{z}$ вида:

$$\tilde{z}(x, x_{n+1}, \dots, x_{n+m}, x_{n+m+1}, \dots, x_{n+2m}) = c'x - (+)M\sum_{l=1}^{m} x_{l+n+m} \to \max (\min), \tag{62}$$

а для образования базиса в систему ограничений (61) вводят неотрицательные *искусственные переменные* x_{n+m+i}, $i = \overline{1, m}$. В результате вместо ЗЛП (50), (57), (52) получаем ЗЛП (62) с основными ограничениями

$$\begin{cases} \sum_{j=1}^{n} a_{1j}x_j - x_{n+1} \quad + x_{n+m+1} \quad = b_1, \\ \sum_{j=1}^{n} a_{2j}x_j \quad - x_{n+2} \quad + x_{n+m+2} \quad = b_2, \\ \dots \\ \sum_{j=1}^{n} a_{mj}x_j \quad - x_{n+m} \quad + x_{n+2m} = b_m \end{cases} \tag{63}$$

и прямыми ограничениями

$$x_j \geq 0, \ j = \overline{1, n+2m}. \tag{64}$$

В новой целевой функции $\tilde{z}$ в (62) знак минус относится к задаче (53), (60), (55) при исследовании ЗЛП на максимум; плюс – к этой же задаче на минимум. Здесь M – большое положительное число (штраф за нарушение ограничений), $x_{n+m+i}, i = \overline{1, m}$, – *искусственные* переменные (за них полагается штраф). Система ограничений (63), в отличие от (61), имеет предпочтительный вид (так как векторы A_{n+m+1}, A_{n+m+2}, …, A_{n+2m}, соответствующие *искусственным* переменным x_{n+m+i}, $i=\overline{1,\ m}$, образуют единичный базис). Положив СП $x_1 = x_2 = \ldots = x_{n+m} = 0$, из системы (63) основных ограничений получим первоначальный базисный план x^1 M-задачи (61)–(63): $x^1 = \left(\underbrace{0,\ 0,\ ...,\ 0}_{n+m},\ b_1,\ b_2,\ ...,\ b_m\right)'$. Если некоторые из уравнений системы ограничений (60) имеют предпочтительный вид, т. е. содержат базисные векторы, то в эти уравнения искусственные переменные можно не вводить.

Пример 13. Составить первоначальный базисный план для ЗЛП

$$z\ (x_1, x_2, x_3, x_4) = -\ 5\ x_1 + x_2 + 3\ x_3 - 2x_4 \to \max;$$

$$\begin{cases} 3x_1 - x_2 + 2x_3 + x_4 \le 1, \\ 2x_1 + 3x_2 - x_3 + 4x_4 \ge 3, \\ x_1 - 2x_2 + 3x_3 - 2x_4 \ge 6; \end{cases} \tag{65}$$

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

Решение. Путем введения неотрицательных дополнительных переменных $x_5 \ge 0$, $x_6 \ge 0$, $x_7 \ge 0$ приведем ЗЛП (65) к каноническому виду:

$$z\ (x_1, x_2, x_3, x_4, x_5, x_6, x_7) = -\ 5\ x_1 + x_2 + 3x_3 - 2x_4 + 0x_5 + 0x_6 + 0\ x_7 \to \max;$$

$$\begin{cases} 3x_1 - x_2 + 2x_3 + x_4 + x_5 = 1, \\ 2x_1 + 3x_2 - x_3 + 4x_4 - x_6 = 3, \\ x_1 - 2x_2 + 3x_3 - 2x_4 - x_7 = 6; \end{cases} \tag{66}$$

$$x_j \ge 0,\ j = \overline{1,\ 7}.$$

Система основных ограничений ЗЛП (66) содержит лишь один базисный вектор $A_5 = (1, 0, 0)'$ (векторы $A_6 = (0, -1, 0)'$, $A_7 = (0, 0, -1)'$ базисными не являются!). До базиса в системе ограничений (66) не хватает двух единичных векторов. Поэтому во второе и третье уравнения системы основных ограничений (66) вводим две *искусственные* переменные $x_8 \ge 0$, $x_9 \ge 0$. В результате возникает M-задача с целевой функцией

$$\tilde{z}(x_1, \ldots, x_9) = -5x_1 + x_2 + 3x_3 - 2x_4 + 0 \cdot x_5 + 0 \cdot x_6 + 0 \cdot x_7 - M\left(x_8 + x_9\right) \to \max$$

и ограничениями

$$\begin{cases} 3x_1 - x_2 + 2x_3 + x_4 + x_5 = 1, \\ 2x_1 + 3x_2 - x_3 + 4x_4 - x_6 + x_8 = 3, \\ x_1 - 2x_2 + 3x_3 - 2x_4 - x_7 + x_9 = 6; \end{cases} \tag{67}$$

$$x_j \geq 0, \;\; j = \overline{1,\ 9}.$$

В (67) векторы A_5, A_8, A_9 являются базисными. Тогда x_5, x_8, x_9 – БП, x_1, x_2, x_3, x_4, x_6, x_7 – СП. Полагая для СП $x_1 = x_2 = x_3 = x_4 = x_6 = x_7 = 0$, из (67) получаем первоначальный базисный план: $x^1 = (0, 0, 0, 0, 1, 0, 0, 3, 6)'$. ▲

13.2. Симплекс-таблица

После приведения системы ограничений ЗЛП к предпочтительному виду ее данные заносят в *симплексную таблицу*. Для ЗЛП с основными ограничениями (56) она имеет вид, представленный в табл. 6.

Таблица 6

i	Базис	$c_б$	b	c_1	c_2	…	c_k	…	c_m	c_{m+1}	…	c_n	θ	
				A_1	A_2	…	A_k	…	A_m	A_{m+1}	…	A_n		
1	A_1	c_1	b_1	1	0	…	0	…	0	$a_{1,m+1}$	…	$a_{1,n}$		
2	A_2	c_2	b_2	0	1	…	0	…	0	$a_{2,m+1}$	…	$a_{2,n}$		
…	…	…	…	…	…	…	…	…	…	…	…	…		№ит
k	A_k	c_k	b_k	0	0	…	1	…	0	$a_{k,m+1}$	…	$a_{k,n}$		
…	…	…	…	…	…	…	…	…	…	…	…	…		
m	A_m	c_m	b_m	0	0	…	0	…	1	$a_{m,m+1}$	…	$a_{m,n}$		
$m+1$	$\Delta_j = z_j - c_j$		Δ_0	Δ_1	Δ_2	…	Δ_k	…	Δ_m	Δ_{m+1}	…	Δ_n		

В столбце «Базис» записывают базисные векторы A_1, A_2, …, A_m, соответствующие первоначальному базисному плану $x^1 = \left(b_1, b_2, ..., b_m, \underbrace{0, ..., 0}_{n-m}\right)'$. В столбце «$c_б$» – коэффициенты c_1, c_2, …, c_m целевой функции $z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$, соответствующие векторам базиса A_1, A_2, …, A_m. В столбце «b» записывают координаты b_1, b_2, …, b_m вектора b в базисе A_1, A_2, …, A_m. В этом столбце на каждой итерации получаем *координаты текущего базисного плана*. В столбцах A_j, $j = \overline{1,\ n}$, записывают координаты вектора A_j в выбранном на данной итерации базисе. В частности, на первой итерации имеем:

$$A_i = \begin{pmatrix} 0 \\ \dots \\ 1 \\ \dots \\ 0 \end{pmatrix} \leftarrow i; \; A_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{mj} \end{pmatrix}, \; i = \overline{1,\, m}, \; j = \overline{1,\, n},$$

так как на первом шаге векторы-столбцы $A_1, A_2, \dots, A_m$ образуют единичный базис.

В строке (*индексной*) с номером $m + 1$ записывают: в столбце «b» – *значение целевой функции* на базисном плане: $z(x^1) = \Delta_0 = c'_б b$, а в столбцах «$A_j$» – *оценки индексной строки:* $\Delta_j = z_j - c_j = c'_б A_j - c_j$, где $z_j = c'_б A_j$. На первом шаге $z_j = \sum_{i=1}^{m} c_i a_{ij}$, $j = \overline{1,\, n}$. Построенный первоначальный базисный план x_0 проверяется на оптимальность.

Теорема 3. Критерий оптимальности (*при решении ЗЛП на максимум*):

1. Если $\Delta_j \geq 0$ *для всех* j, $j = \overline{1,\, n}$, *то базисный план* x_0 *оптимален.*

2. Если найдется хотя бы одна оценка $\Delta_j < 0$, *такая, что в столбце* A_j *все числа* $x_{ij} \leq 0, i = \overline{1,\, m}$, *то целевая функция не ограничена сверху на множестве допустимых планов задачи и ЗЛП не имеет решения.*

3. Если среди оценок Δ_j *есть отрицательные, причем для каждого номера* j *с* $\Delta_j < 0$ *существует хотя бы один номер* i, *при котором* $x_{ij} > 0$, *то план* x_0 *не является оптимальным.*

В этом случае осуществляется следующая итерация симплекс-метода.

13.3. Переход к новому базисному плану

Если первоначальный базисный план $x^1 = \left(x_1^1,\, x_2^1,\, \dots,\, x_n^1\right)'$ не является оптимальным, то при исследовании *ЗЛП* (*на максимум*) переходят к нехудшему базисному плану $x^2 = \left(x_1^2,\, x_2^2,\, \dots,\, x_n^2\right)$ в направлении наискорейшего возрастания целевой функции. Для этого среди отрицательных оценок $\Delta_j < 0$ находят оценку Δ_{j_0}, максимальную по абсолютной величине:

$$\max_{\Delta j < 0} \left|\Delta_j\right| = \left|\Delta_{j_0}\right|.$$

Столбец A_{j_0}, соответствующий номеру j_0, называют *разрешающим* столбцом. При решении задачи на *минимум* из всех положительных оценок $\Delta_j > 0$ выбирают $\max\Delta_j > 0$, и в качестве разрешающего столбца A_{j_0} берут столбец с номером j_0 из условия $\max\limits_{\Delta j>0} \Delta_j = \Delta_{j_0}$. Вектор-столбец A_{j_0} вводят в базис. Для всех положительных элементов $x_{ij_0} > 0$, $i = \overline{1,\ m}$, вычисляют величину θ_i: $\theta_i = \dfrac{b_i}{x_{ij_0}}$, $i = \overline{1,\ m}$, которая называется *симплексным отношением*. Симплексные отношения θ_i, $i = \overline{1,\ m}$, заносят в столбец θ симплекс-таблицы 6. Среди всех симплексных отношений θ_i находим наименьшее θ_{i_0}:

$$\min_{i,\, x_{ij_0}>0} \frac{b_i}{x_{ij_0}} = \frac{b_{i_0}}{x_{i_0 j_0}} = \theta_{i_0}.$$

Строка i_0, в которой находится наименьшее симплексное отношение θ_{i_0}, называется *разрешающей строкой*. Базисный вектор, стоящий в этой строке, исключается из базиса. Элемент $x_{i_0 j_0}$, стоящий на пересечении *разрешающего столбца* и *разрешающей строки*, называется *разрешающим элементом*.

Разрешающая строка и разрешающий столбец отмечаются в симплекс-таблице стрелками. Для завершения шага преобразований, ведущих к новому базисному плану x^2, составляют ***новую симплекс-таблицу по следующим правилам:***

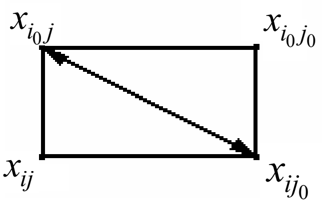
1. Элементы строки i_0 новой таблицы равны соответствующим элементам разрешающей строки старой таблицы, деленным на разрешающий элемент $x_{i_0 j_0}$;

2. Все элементы столбца j_0 в новой таблице равны 0, за исключением элемента с номером (i_0, j_0): $a_{i_0 j_0} = 1$;

3. Чтобы получить все остальные элементы (включая элементы индексной строки в новой таблице), нужно из соответствующего элемента прежней таблицы вычесть произведение элемента разрешающей строки на элемент разрешающего столбца, разделенного на разрешающий элемент. Для контроля вычислений элементов индексной строки применяют формулы

$$\Delta_0 = c'_б b;\quad \Delta_j = c'_б A_j - c_j,\quad j = \overline{1,\ n}.$$

Заметим, что любой из оставшихся элементов $(x_{ij})_{нов}$ новой симплекс-таблицы можно получить через элементы $x_{ij}, x_{i_0 j}, x_{ij_0}, x_{i_0 j_0}$ старой симплекс-таблицы по *правилу прямоугольника*:



$$\left(x_{ij}\right)_{\text{нов}} = x_{ij} - \frac{x_{i_0 j} x_{ij_0}}{x_{i_0 j_0}}.$$

Новый базисный план x^2 опять проверяется на оптимальность по описанной выше схеме с использованием критерия оптимальности. Если план x^2 оптимален, то задача решена, если нет – строим новый базисный план. Этот процесс продолжаем до тех пор, пока не будет получено решение (п. 1 *критерия оптимальности*) или установлена неограниченность целевой функции (п. 2 *критерия оптимальности*). Переход от старого базисного плана к новому называется *симплекс-итерацией.*

С помощью симплекс-метода может быть решена ЗЛП с любым числом переменных, в частности и с $n = 2$. Возникает вопрос: как в симплекс-таблице отражается ситуация, изображенная на рис. 14, когда $z_{\max}$ достигается не в одной точке, а на отрезке AE, т. е. когда имеет место *альтернативный оптимум*?

Теорема 4 (*альтернативный оптимум* – *признак бесконечности множества оптимальных планов*). *Если в индексной строке последней симплекс-таблицы, содержащей оптимальный план, имеется хотя бы одна нулевая оценка, соответствующая свободной (не базисной) переменной, то ЗЛП имеет бесконечное множество оптимальных планов.*

Ввиду особого значения п. 2 критерия оптимальности (см. теорему 3) сформулируем его в виде отдельного утверждения.

Теорема 5 (*признак неограниченности целевой функции*). *Если в разрешающем столбце симплекс-таблицы нет ни одного положительного элемента, то ЗЛП неразрешима в силу неограниченности целевой функции на множестве допустимых планов.*

Обсуждение. Описанный алгоритм симплекс-метода для ЗЛП с невырожденными базисными планами позволяет за конечное число итераций получить оптимальный план либо установить неограниченность значений целевой функции на множестве планов. Этот алгоритм можно применять и для вырожденных ЗЛП, т. е. задач с вырожденными базисными планами. Теоретически здесь возможно «зацикливание», т. е. возвращение к уже использованным базисным векторам, однако, как правило, на практике это явление не наблюдается.

13.4. Метод искусственного базиса (M-задача)

Рассмотрим ЗЛП в канонической форме:

$$z(x_1, x_2, \ldots, x_n) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max; \qquad (68)$$

$$\begin{cases} a_{11}x_1 + ... + a_{12}x_2 + ... + a_{1n}x_n = b_1, \\ ... \\ a_{m1}x_1 + ... + a_{m2}x_2 + ... + a_{mn}x_n = b_m; \end{cases} \tag{69}$$

$$x_j \geq 0, \quad j = \overline{1,\ n}, \tag{70}$$

где, не ограничивая общности, полагаем $b_i \geq 0,\ i = \overline{1,\ m}$. Если среди векторов условий системы основных ограничений (69) единичный базис отсутствует, то в ЗЛП (68)–(70) вводим *искусственные* переменные $x_{n+i} \geq 0,\ i = \overline{1,\ m}$, и переходим к расширенной задаче (*М-задаче*):

$$\begin{gathered} \tilde{z}(x_1,\ \ldots,\ x_n,\ x_{n+1},\ \ldots,\ x_{n+m}) = \\ = c_1x_1 + ... + c_nx_n - M\left(x_{n+1} + ... + x_{n+m}\right) \to \max; \end{gathered} \tag{71}$$

$$\begin{cases} a_{11}x_1 + ... + a_{1n}x_n + x_{n+1} & = b_1, \\ a_{21}x_1 + ... + a_{2n}x_n \quad + x_{n+2} & = b_2, \\ \ldots \\ a_{m1}x_1 + ... + a_{mn}x_n \qquad + x_{n+m} & = b_m; \end{cases} \tag{72}$$

$$x_j \geq 0, \quad j = \overline{1,\ n+m}. \tag{73}$$

В (71) M – произвольное достаточно большое положительное число, а $x_{n+i},\ i = \overline{1,\ m}$, – *искусственные* переменные. Единичные векторы $A_{n+1},\ \ldots,\ A_{n+m}$, соответствующие искусственным переменным, $x_{n+i},\ i = \overline{1,\ m}$, образуют базис. Тогда $x_1,\ x_2,\ \ldots,\ x_n$ – СП. Положив в системе (72) основных ограничений СП $x_1 = x_2 = \ldots = x_n = 0$, получим первоначальный базисный план $\tilde{x}^1 = (\underbrace{0,\ \ldots,\ 0}_{n},\ b_1,\ b_2,\ \ldots,\ b_m)'$. Решая M-задачу симплекс-методом, через конечное число итераций придем к оптимальному плану либо установим неразрешимость M-задачи.

Рассмотрим связь между решениями ЗЛП (68)–(71), содержащей n переменных $x_i,\ i = \overline{1,\ n}$, и M-задачей (71)–(73) с переменными x_i, $i = \overline{1,\ n+m}$.

Теорема 6. *Если в оптимальном плане* $\tilde{x}^0 = (x_1,\ x_2,\ \ldots,\ x_n,\ \underbrace{0,\ \ldots,\ 0}_{m})'$ *М-задачи (71)–(73) все искусственные переменные равны нулю, то план* $x^0 = (x_1, x_2, \ldots, x_n)'$ *является оптимальным планом ЗЛП (68)–(70).*

Теорема 7. *Если в оптимальном плане М-задачи (71)–(73) не все искусственные переменные равны нулю, то исходная ЗЛП (68)–(70) не имеет планов (допустимых решений). Множество планов пусто.*

Теорема 8. *Если в М-задаче (71)–(73)целевая функция* $\tilde{z}$ *не ограничена, то и в исходной ЗЛП (68)–(70) целевая функция z не ограничена (на множестве планов).*

13.5. Симплекс-таблица для *М*-задачи

Все данные *М*-задачи (71)–(73) заносим в симплекс-таблицу 6 обычным образом за исключением последней (m + 1)-й строки. Поскольку в *М*-задаче оценки $\Delta_j = z_j - c_j$ являются линейными функциями от *М*, то их удобно записывать в двух строках. В строке (m + 1) записывают свободные члены, а в строке (m + 2) – коэффициенты при *М*. В столбце «*b*» записывают значение целевой функции $\tilde{z}$ на базисном плане также в двух строках. При решении задачи на максимум по наименьшему отрицательному элементу строки (m + 2) определяем вектор, подлежащий включению в базис. В случае равенства нулю элементов строки (m + 2) правило выбора разрешающего столбца применяем к строке (m + 1).

Замечание 10. При наличии в матрице условий единичных векторов число искусственных переменных может быть уменьшено. Искусственные переменные достаточно ввести так, чтобы соответствующие им базисные векторы составили вместе с выбранными единичными векторами условий единичную матрицу. Симплекс-метод начинается, вообще говоря, с искусственного базиса. Оказывается, что *если искусственный вектор на некоторой итерации выйдет из базиса, то обратно в базис он уже никогда не войдет.* В этом случае из симплекс-таблицы может быть удален соответствующий столбец.

● **Задача 18.** Решите симплекс-методом задачу о рационе (см. с. 80).

● **Задача 19.** Решите симплекс-методом задачу 16 (см. с. 82), задав при этом необходимые параметры.

14. ТРАНСПОРТНАЯ ЗАДАЧА

Транспортная задача относится к специальным задачам ЛП. Для некоторых видов ЗЛП созданы специальные методы решения, значительно сокращающие объем вычислений по сравнению с симплекс-методом. Одну из таких задач (транспортную задачу для двух поставщиков) в 30-е годы исследовал математик А. Н. Толстой. В общем случае транспортная задача была решена академиком Л. В. Канторовичем.

14.1. Постановка транспортной задачи. Теорема существования

Общая постановка транспортной задачи состоит в определении оптимального плана перевозок некоторого однородного товара (груза) из m пунктов отправления в n пунктов назначения. 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.

Имеется m пунктов производства $A_1, A_2, \ldots, A_m$ однородного груза и n пунктов $B_1, B_2, \ldots, B_n$ его потребления. В i-м пункте производства A_i произведено $a_i, i=\overline{1, m}$, единиц груза, а в j-м пункте потребления $B_j, j=\overline{1, n}$, реализуется $b_j, j=\overline{1, n}$, единиц груза. Известна стоимость c_{ij} (*тариф*) перевозки единицы груза из пункта A_i в пункт B_j. Матрица $C = (c_{ij})_{m\times n}$ называется *матрицей тарифов*. Полагаем, что выполняется *условие баланса*

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j, \tag{74}$$

т. е. запасы всего произведенного груза полностью востребованы потребителями.

Задача. Составить план перевозок, при котором весь груз из пунктов производства A_i, $i=\overline{1, m}$, будет вывезен, все запросы потребителей B_j, $j=\overline{1, n}$, удовлетворены, а транспортные расходы по перевозке груза окажутся минимальными.

Составим математическую модель задачи. Пусть x_{ij} ($x_{ij} \geq 0$) – количество груза, запланированного к перевозке из пункта A_i в пункт B_j, $i=\overline{1, m}$, $j=\overline{1, n}$.

Планом транспортной задачи называется матрица перевозок $X = (x_{ij})_{m\times n}$, где x_{ij} – количество единиц груза, которое необходимо доставить из i-го пункта производства в j-й пункт потребления при условии, что все запасы вывезены и все потребности удовлетворены.

Тогда общая стоимость z всех перевозок выразится функцией

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

Определение 9. План $X^0 = (x_{ij}^0)_{m\times n}$, при котором функция z принимает свое минимальное значение, называется *оптимальным планом* транспортной задачи.

Таким образом, приходим к следующей ЗЛП: указать количество $x_{ij} \geq 0$, $i=\overline{1, m}$, $j=\overline{1, n}$, груза, перевозимого из пункта A_i в пункт B_j, при условиях

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

$$\sum_{j=1}^{n} x_{ij} = a_i,\ i = \overline{1,\ m} \quad \text{(все запасы вывезены);} \qquad (75)$$

$$\sum_{i=1}^{m} x_{ij} = b_j,\ j = \overline{1,\ n} \quad \text{(все потребности удовлетворены).}$$

Все данные транспортной задачи удобно записывать в виде табл. 7, объединяя матрицы тарифов и перевозок в одной *матрице планирования*.

Таблица 7

Поставщики	**Потребители**				**Запасы**
	B_1	B_2	…	B_n	
A_1	c_{11} x_{11}	c_{12} x_{12}	…	c_{1n} x_{1n}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}	…	c_{2n} x_{2n}	a_2
…	…	…	…	…	…
A_m	c_{m1} x_{m1}	c_{m2} x_{m2}	…	c_{mn} x_{mn}	a_m
Потребности	b_1	b_1	…	b_n	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

В правом верхнем углу каждой клетки матрицы планирования записывают стоимость c_{ij} перевозки единицы груза от i-го поставщика к j-му потребителю, которую называют *тарифом* клетки $(i,\ j)$.

Теорема существования решения. *Для того чтобы транспортная задача имела решение, необходимо и достаточно, чтобы выполнялось условие баланса (74).*

Решение транспортной задачи, как и любой ЗЛП, состоит из трех этапов:

1) построение первоначального базисного плана перевозок;

2) проверка построенного плана на оптимальность;

3) в случае неоптимальности плана – переход к нехудшему базисному плану.

14.2. Построение первоначального базисного плана

Решение транспортной задачи начинаем с построения *первоначального базисного плана* $X_1 = (x_{ij}),\ i = \overline{1,\ m},\ j = \overline{1,\ n}$. Система ограничений задачи (75) содержит $m \times n$ неизвестных x_{ij}, а число уравнений в системах основных ограничений (75) – $(m + n)$. Так как по предположению выполняется условие баланса (74), то число линейно независимых уравнений составляет $m + n - 1$. Следовательно, план транспортной задачи может иметь не более $m + n - 1$ отличных от нуля перевозок.

Рассмотрим методы построения *первоначального базисного плана.*

Метод северо-западного угла. Сначала заполняют клетки первой строки матрицы планирования слева направо по порядку, пока не будут исчерпаны запасы a_1, при этом по возможности полностью удовлетворяют потребности B_1, B_2, …, B_n. Затем последовательно заполняют клетки второй строки, начиная с того потребителя, которому не хватило запасов a_1, до полного исчерпания запасов a_2 и т. д.

Метод минимальной стоимости. Заполнение таблицы начинают с клетки, тариф которой минимален, и затем продолжают в порядке неубывания стоимостей. Клетки, в которых стоят отличные от нуля перевозки x_{ij}, называются *загруженными* (занятыми), а остальные – *свободными*. План, содержащий $m + n - 1$ загруженную клетку, называется *невырожденным*. *Базисность* плана состоит в его *ацикличности*, т. е. в невозможности построения в матрице планирования замкнутого *цикла*, все вершины которого расположены в загруженных клетках.

Определение 10. ***Циклом*** называется ломаная линия, вершины которой расположены в загруженных клетках матрицы планирования, 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.

Загруженные клетки соответствуют базисным переменным и для невырожденного плана их число равно $m + n - 1$. Если число загруженных клеток построенного первоначального плана меньше, чем $m + n - 1$, то он называется *вырожденным*. Тогда для получения невырожденного плана в свободную клетку (обычно в ту, которой соответствует наименьший тариф) заносится *базисный нуль*, и эта клетка считается загруженной. Построенный план будет базисным, если он *ацикличен.*

Отметим важные ***свойства планов транспортной задачи***:

1) план транспортной задачи является базисным тогда и только тогда, когда из занятых этим планом клеток нельзя образовать ни одного цикла;

2) для любого невырожденного базисного плана и для каждой свободной клетки существует (и при том только один) цикл, содержащий данную клетку и некоторые загруженные клетки.

После нахождения первоначального невырожденного ациклического плана необходимо исследовать его на оптимальность и, если он не оптимален, перейти к улучшенному плану. Для определения оптимальности плана транспортной задачи разработано несколько методов. Однако наиболее часто используется *метод потенциалов*.

14.3. Метод потенциалов. Проверка на оптимальность

После того как найден первоначальный невырожденный базисный план перевозок, каждому поставщику A_i ставится в соответствие некоторое число u_i, $i = \overline{1, m}$, а каждому потребителю B_j – некоторое число v_j, $j = \overline{1, n}$. Числа ui и vj называются *потенциалами* соответственно A_i (строк) и B_j (столбцов). Числа u_i и v_j выбирают так, чтобы для *любой загруженной клетки* их сумма была равна тарифу этой клетки, т. е.

$$u_i + v_j = c_{ij}, \ i = \overline{1, m}, \ j = \overline{1, n}. \quad (76)$$

Так как чисел u_i и v_j $m + n$, а число загруженных клеток $m + n - 1$, то для нахождения потенциалов u_i и v_j получаем систему (76) $m + n - 1$ уравнений с $m + n$ неизвестными. Поэтому один потенциал можно задать произвольно, например положив его равным нулю (обычно для строки или столбца, где больше занятых клеток). Тогда остальные потенциалы определяются однозначно. При проверке оптимальности плана для *каждой свободной клетки* вычисляют разность Δ_{ij} между тарифом этой клетки и суммой потенциалов u_i и v_j, соответствующих данной клетке:

$$\Delta_{ij} = c_{ij} - (u_i + v_j). \quad (77)$$

Теорема 9 (критерий оптимальности). *План является оптимальным, если для всех свободных клеток* $\Delta_{ij} \geq 0$. *Если хотя бы для одной клетки* $\Delta_{ij} < 0$, *то план не оптимален, и необходимо переходить к новому базисному плану.*

14.4. Переход к новому плану

Переход к новому базисному плану осуществляется по правилам:

1. Свободная клетка, для которой значение Δ_{ij} минимально, называется *перспективной* и подлежит загрузке. Если таких клеток несколько, то выбираем клетку с наименьшим тарифом и ее загружаем.

2. Для выбранной перспективной клетки строим *цикл*, т. е. составляем *замкнутый контур*, по которому следует перераспределить груз. *Контур* представляет собой замкнутую ломаную линию, составленную из направленных горизонтальных и вертикальных отрезков, соединяющих клетки, из которых одна (перспективная) свобод-

на, а все остальные загружены. Для каждой свободной клетки такой ***контур существует и является единственным***. Точка, в которой меняется направление контура (горизонтальное на вертикальное и наоборот), называется *вершиной цикла*. В одной строке или столбце могут находиться две и только две вершины цикла. Точки самопересечения контура вершинами цикла не являются.

Примеры циклов приведены на рис. 15.

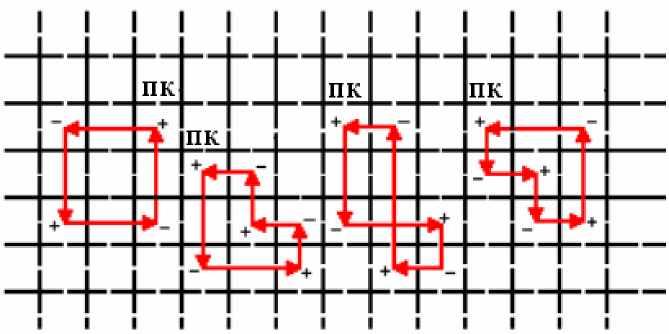


Рис. 15. Примеры циклов для транспортной задачи

3. Расставляем поочередно знаки «+» и «–» в вершинах цикла, начиная с перспективной клетки. Перспективной клетке приписы-ваем знак «+», а остальным вершинам цикла поочередно знаки «–» и «+», пока не придем в исходную вершину.

4. В клетках, соответствующих отрицательным вершинам, отыскиваем наименьший груз θ, который перемещаем по клеткам цикла, т. е. прибавляем к поставкам x_{ij} в клетках со знаком «+» (включая перспективную клетку) и вычитаем из поставок в клетках со знаком «–». В итоге получаем новый план, для которого составляем новую систему потенциалов и проверяем план на оптимальность. Процесс продолжаем до тех пор, пока не получим оптимальный план.

Замечание 11. Наличие в оптимальном плане свободных клеток, в которых $\Delta_{ij} = 0$, говорит о неединственности оптимального плана, что полезно учитывать при составлении конкретных плановых заданий.

Замечание 12. При выполнении *условия баланса* $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$ имеем ***закрытую модель*** транспортной задачи. Если условие баланса не выполнено, получаем ***открытую модель*** транспортной задачи. При этом возможно выполнение двух условий: а) запасы превосходят потребности: $\sum_{i=1}^{m} a_i > \sum_{i=1}^{n} b_j$ (*перепроизводство товаров*); б) запасы не обеспечивают потребности: $\sum_{i=1}^{m} a_i < \sum_{i=1}^{n} b_j$ (*дефицит товаров*). Для сведения открытой задачи к закрытой вводят фиктивного потребителя в случае «а» или фиктивного поставщика в случае «б». Все дополнительные тарифы при этом полагают равными нулю.

Пример 14. Имеется четыре поставщика некоторых товаров с запасами $a_1 = 400$, $a_2 = 200$, $a_3 = 300$, $a_4 = 200$ и четыре потребителя, потребности которых в данном товаре составляют $b_1 = 240$, $b_2 = 360$, $b_3 = 180$, $b_4 = 400$. Известна матрица тарифов $C = \begin{pmatrix} 4 & 7 & 1 & 3 \\ 2 & 8 & 4 & 4 \\ 6 & 5 & 1 & 3 \\ 4 & 3 & 7 & 6 \end{pmatrix}$ – матрица стоимости перевозок единицы товара от каждого поставщика к каждому потребителю. Составить оптимальный план перевозок.

Решение. Прежде чем составлять матрицу планирования, проверим выполнение условия баланса. Так как $\sum_{i=1}^{4} a_i = 1100 < \sum_{j=1}^{4} b_j = 1180$, то условие баланса не выполняется, следовательно, имеем открытую модель транспортной задачи типа «б». Поэтому для решения задачи необходимо ввести фиктивного поставщика A_5 с запасом $a_5 = \sum_{j=1}^{4} b_j - \sum_{i=`1}^{4} a_i = 80$ и нулевыми тарифами. Первоначальный базисный план, построенный по методу минимальной стоимости, представлен в табл. 8.

Таблица 8

Постав-щики	**Потребители**				**Запасы**
	B_1	B_2	B_3	B_4	
A_1	4 –	7 –	1 180	3 220	400
A_2	2 200	8 –	4 –	4 –	200
A_3	6 –	5 120	1 –	3 180	300
A_4	4 –	3 200	7 –	6 –	200
A_5	0 40	0 40	0 –	0 –	80
Потребности	240	360	180	400	1180 1180

Полученный первоначальный базисный план

$$X_1 = \begin{pmatrix} 0 & 0 & 180 & 220 \\ 200 & 0 & 0 & 0 \\ 0 & 120 & 0 & 180 \\ 0 & 200 & 0 & 0 \\ 40 & 40 & 0 & 0 \end{pmatrix}$$

является невырожденным. Действительно, число загруженных клеток в матрице планирования X_1 совпадает с числом $m + n - 1 = 8$. Стоимость перевозки товаров по этому плану составляет $z(X_1) = 1 \cdot 180 + 3 \cdot 220 + 2 \cdot 200 + 5 \cdot 120 + 3 \cdot 180 + 3 \cdot 200 + 0 \cdot 40 + 0 \cdot 40 =$ $= 2980$ ден. ед. Выясним, является ли этот план оптимальным. Согласно формуле (76), составим систему потенциалов, положив один из них, например u_1, равным нулю. Тогда остальные потенциалы определяются однозначно (табл. 9).

Таблица 9

Поставщики	**Потребители**				**Запасы**	**Потенциалы** u_i
	B_1	B_2	B_3	B_4		
A_1	+ 4 –	7 –	1 180	- 3 220	400	$u_1 = 0$
A_2	2 200	8 –	4 –	4 –	200	$u_2 = -3$
A_3	6 –	5 - 120	1 –	3 180 +	300	$u_3 = 0$
A_4	4 –	3 200	7 –	6 –	200	$u_4 = -2$
A_5	- 0 40	+ 0 40	0 –	0 –	80	$u_5 = -5$
Потребности	240	360	180	400	1180 1180	
Потенциалы v_j	$v_1 = 5$	$v_2 = 5$	$v_3 = 1$	$v_4 = 3$		

По всем свободным клеткам, согласно формуле (77), вычислим Δ_{ij}:

$$\Delta_{11} = 4 - (5 + 0) = -1 < 0;\ \Delta_{12} = 7 - (5 + 0) = 2 > 0;$$

$$\Delta_{22} = 8 - (5 + (-3)) = 6 > 0;\ \Delta_{23} = 4 - (1 + (-3)) = 6 > 0;$$

$$\Delta_{24} = 4 - (3 + (-3)) = 4 > 0;\ \Delta_{31} = 6 - (5 + 0) = 1 > 0;$$

$$\Delta_{33} = 1 - (1 + 0) = 0;\ \Delta_{34} = 3 - (3 + 0) = 0;$$

$$\Delta_{41} = 4 - (5 + (-2)) = 1 > 0;\ \Delta_{43} = 7 - (1 + (-2)) = 8 > 0;$$

$$\Delta_{44} = 6 - (3 + (-2)) = 5 > 0;\ \Delta_{53} = 0 - (1 + (-5)) = 4 > 0;$$

$$\Delta_{54} = 0 - (3 + (-5)) = 2 > 0.$$

Среди оценок Δ_{ij} значение Δ_{11} оказалось отрицательным. Это означает, что план X_1 не является оптимальным, стоимость перевозки товаров можно уменьшить. Свободная клетка (1; 1) является перспективной. Ее следует загрузить. Согласно второму пункту правил перехода к новому плану, необходимо построить цикл (см. табл. 9) с клеткой (1; 1) в качестве одной из вершин замкнутого контура, в котором только одна клетка (перспективная) является незагруженной, а остальные – загруженные. Таким замкнутым контуром является контур, проходящий через клетки (1; 1), (1; 4), (3; 1), (3; 2), (5; 2), (5; 1), (1; 1). Расставим в вершинах циклапоочередно знаки «+» и «–», начиная с перспективной. Из клеток со знаком «–» выберем наименьший груз: $\theta = \min(220; 120; 40) = 40$, который прибавим ко всем перевозкам в вершинах цикла со знаком «+» и вычтем из всех перевозок в вершинах цикла со знаком «–». В результате получим новую таблицу 10 и новый план

$$X_2 = \begin{pmatrix} 40 & 0 & 180 & 180 \\ 200 & 0 & 0 & 0 \\ 0 & 80 & 0 & 220 \\ 0 & 200 & 0 & 0 \\ 0 & 80 & 0 & 0 \end{pmatrix},$$ стоимость перевозок по которому равна $z(X_2) = 4 \cdot 40 + 1 \cdot 180 + 3 \cdot 180 + 2 \cdot 200 + 5 \cdot 80 + 3 \cdot 220 + 3 \cdot 200 + 0 \cdot 80 = 2940$ ден. ед.

Для того чтобы узнать, является ли данный план оптимальным, по загруженным клеткам составим новую систему потенциалов (см. табл. 10). Затем по свободным клеткам вычислим оценки Δ_{ij}:

$$\Delta_{12} = 7 - (5 + 0) = 2 > 0;\ \Delta_{22} = 8 - (5 + (-2)) = 5 > 0;$$

$$\Delta_{23} = 4 - (1 + (-2)) = 5 > 0;\ \Delta_{24} = 4 - (3 + (-2)) = 3 > 0;$$

$$\Delta_{31} = 6 - (4 + 0) = 2 > 0;\ \Delta_{33} = 1 - (1 + 0) = 0;$$

$$\Delta_{41} = 4 - (4 + (-2)) = 2 > 0;\ \Delta_{43} = 7 - (1 + (-2)) = 8 > 0;$$
$$\Delta_{44} = 6 - (3 + (-2)) = 5 > 0;\ \Delta_{51} = 0 - (4 + (-5)) = 1 > 0;$$
$$\Delta_{53} = 0 - (1 + (-5)) = 4 > 0;\ \Delta_{54} = 0 - (3 + (-5)) = 2 > 0.$$

Таблица 10

Постав-щики	**Потребители**				**Запа-сы**	**Потен-циалы** u_i
	B_1	B_2	B_3	B_4		
A_1	4 40	7 –	1 180	3 180	400	$u_1 = 0$
A_2	2 200	8 –	4 –	4 –	200	$u_2 = -2$
A_3	6 –	5 80	1 –	3 220	300	$u_3 = 0$
A_4	4 –	3 200	7 –	6 –	200	$u_4 = -2$
A_5	0 –	0 80	0 –	0 –	80	$u_5 = -5$
Потребно-сти	240	360	180	400	1180 1180	
Потенциа-лы v_j	$v_1 = 4$	$v_2 = 5$	$v_3 = 1$	$v_4 = 3$		

Так как все $\Delta_{ij} \geq 0$, то, согласно теореме 9, план

$$X^0 = \begin{pmatrix} 40 & 0 & 180 & 180 \\ 200 & 0 & 0 & 0 \\ 0 & 80 & 0 & 220 \\ 0 & 200 & 0 & 0 \end{pmatrix}$$

является оптимальным. Стоимость перевозок по нему составляет $z(X^0) = 4 \cdot 40 + 1 \cdot 180 + 3 \cdot 180 + 2 \cdot 200 + 5 \cdot 80 + 3 \cdot 220 + 3 \cdot 200 = 2940$ ден. ед. В силу замечания 11 клетка (3; 3) с $\Delta_{33} = 0$ говорит о том, что полученный план не является единственным: альтернативный план той же стоимости можно получить, если построить цикл с перспективной клеткой (3; 3). ▲

● **Задача 20.** Постройте альтернативный план для примера 14 и убедитесь, что стоимость перевозки продуктов по нему равна 2940 ден. ед.

● **Задача 21.** В трех хранилищах горючего ежедневно хранится 175, 125 и 140 т бензина. Этот бензин ежедневно получают четыре заправочные станции в количествах, равных соответственно 180, 110, 60 и 40 т. Стоимости перевозок 1 т бензина с хранилищ к заправочным станциям задаются матрицей $C=\begin{pmatrix} 9 & 7 & 5 & 3 \\ 1 & 2 & 4 & 6 \\ 8 & 10 & 12 & 1 \end{pmatrix}$. Составьте план перевозок бензина, при котором общая стоимость перевозок является минимальной.

15. ДВОЙСТВЕННЫЕ ЗАДАЧИ. ТЕОРИЯ ДВОЙСТВЕННОСТИ

С каждой ЗЛП, которая называется *прямой* или исходной, тесно связана другая ЗЛП, называемая *двойственной*. Многие ЗЛП первоначально ставятся в виде исходных или двойственных задач, поэтому говорят о *паре взаимно двойственных задач*. Рассмотрим экономическую задачу, приводящую к двойственной ЗЛП.

15.1. Задача о распределении ресурсов

Имеются два предприятия A и B. Предприятие A располагает m видами ресурсов в объеме $b_1, b_2, \ldots, b_m$ и производит n видов продукции в количестве $x_1, x_2, \ldots, x_n$ соответственно. Известна норма a_{ij} затрат i-го, $i=\overline{1,\ m}$, вида ресурсов на производство единицы j-го, $j=\overline{1,\ n}$, вида продукции и выручка c_j от реализации единицы продукции j-го вида, $j=\overline{1,\ n}$. Требуется спланировать производство продукции на предприятии A таким образом, чтобы суммарная прибыль от ее реализации была максимальной.

Составим математическую модель задачи:

$$z(x) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max; \tag{78}$$

$$a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \ldots + a_{in}x_n \le b_i,\ \ i=\overline{1,\ m}; \tag{79}$$

$$x_j \ge 0, \quad j=\overline{1,\ n}. \tag{80}$$

Предположим, что предприятие A решило реализовать имеющиеся ресурсы, не выпуская готовой продукции. В его интересах получить прибыль, не меньшую, чем при реализации готовой продукции. Предприятие B, желающее закупить ресурсы, заинтересовано в наименьшей их стоимости. Каковы должны быть установлены оптимальные цены $y_1, y_2, \ldots, y_m$ на эти ресурсы?

Очевидно, предприятию A выгодно продать ресурсы, идущие на изготовление единицы продукции j-го, $j=\overline{1,\ n}$, вида, если стоимость $a_{1j}y_1 + a_{2j}y_1 + \ldots + a_{mj}y_m$, $j=\overline{1,\ n}$, ресурсов не меньше прибыли c_j от реализации единицы этого вида продукции. В интересах предприятия B, чтобы стоимость $b_1y_1 + b_2y_2 + \ldots + b_my_m$ ресурсов была минимальна. Таким образом, приходим к ЗЛП следующего вида:

$$f(y) = b_1y_1 + b_2y_2 + \ldots + b_my_m \to \min; \tag{81}$$

$$a_{1j}y_1 + a_{2j}y_1 + \ldots + a_{mj}y_m \geq c_j, \quad j = \overline{1,\ n}; \tag{82}$$

$$y_i \geq 0,\ i = \overline{1,\ m}. \tag{83}$$

Задача (81)–(83) называется *двойственной* к ЗЛП (78)–(80). Переменные y_1, y_2, …, y_m считаются *двойственными* (объективно обусловленными) *оценками.*

Таким образом, *прямая ЗЛП* (78)–(80) в векторно-матричной форме имеет вид

$$z(x) = c'x \to \max; \tag{78*}$$

$$Ax \leq b; \tag{79*}$$

$$x \geq 0, \tag{80*}$$

а *двойственная задача* (81)–(83) может быть записана как

$$f(y) = b'y \to \min; \tag{81*}$$

$$A'y \geq c; \tag{82*}$$

$$y \geq 0. \tag{83*}$$

Сопоставляя задачи (78*)–(80*) и (81*)–(83*), можно сформулировать следующее ***правило перехода от прямой ЗЛП в нормальной форме*** (78*)–(80*) ***к двойственной ЗЛП*** (81*)–(83*):

$$x \to y,\ c \to b,\ \max \to \min,\ A \to A',\ \leq \to \geq,\ x \geq 0 \to y \geq 0.$$

Для прямой *ЗЛП в канонической форме*

$$z(x) = c'x \to \max,\ Ax = b,\ x \geq 0,$$

двойственной является ЗЛП следующего вида:

$$f(y) = b'y \to \min,\ A'y \geq c,\ y - \text{любого знака.}$$

15.2. Правила построения двойственной задачи

В общем случае ЗЛП можно рекомендовать следующие правила построения двойственной задачи.

1. Если прямая ЗЛП исследуется на максимум, то двойственная – на минимум и наоборот.

2. В задаче на максимум ограничения-неравенства имеют один смысл, а в задаче на минимум – противоположный.

3. Вектору ограничений прямой задачи соответствует вектор стоимостей двойственной задачи и наоборот.

4. Матрица системы ограничений двойственной задачи получается из матрицы системы исходной задачи транспонированием.

5. Если какое-либо ограничение прямой задачи записано как равенство, то на соответствующую переменную двойственной задачи условие неотрицательности не налагается.

Пример 15. Рассмотрим ЗЛП:

$$z(x_1, x_2, x_3, x_4) = -12x_1 + 10x_2 - 15x_3 - 11x_4 \to \max;$$

$$\begin{cases} 2x_1 - 4x_2 \qquad\quad + x_4 \geq 3, \\ 3x_1 + x_2 + 2x_3 + 2x_4 = 4, \\ 4x_1 - 5x_2 \qquad\quad + 2x_4 \geq 1, \\ x_1 \qquad + 3x_3 + 2x_4 \geq 2, \\ -2x_1 + x_2 \qquad\quad - 3x_4 = -5; \end{cases}$$

$$x_1 \geq 0,\ x_4 \geq 0,\ x_2, x_3 \text{ – произвольного знака.}$$

Построить двойственную ей задачу.

Решение. Приведем основные ограничения исходной ЗЛП к виду ≤ 0, вектору ограничений прямой ЗЛП, в силу правила 3, поставим в соответствие вектор стоимости двойственной ЗЛП ($y_i\,(i = \overline{1,\ 5})$ – переменные двойственной задачи):

$$z(x_1, x_2, x_3, x_4) = -12x_1 + 10x_2 - 15x_3 - 11x_4 \to \max;$$

$$\begin{cases} -2x_1 + 4x_2 \qquad\quad - x_4 \leq -3, & \Big| y_1 \\ 3x_1 + x_2 + 2x_3 + 2x_4 = 4, & \Big| y_2 \\ -4x_1 + 5x_2 \qquad\quad - 2x_4 \leq -1, & \Big| y_3 \\ -x_1 \qquad - 3x_3 - 2x_4 \leq -2, & \Big| y_4 \\ -2x_1 + x_2 \qquad\quad - 3x_4 = -5, & \Big| y_5 \end{cases}$$

$$x_1 \geq 0,\ x_4 \geq 0,\ x_2, x_3 \text{ – произвольного знака.}$$

Тогда целевая функция двойственной задачи, согласно правилам 1, 3, и основные ограничения, в силу правил 2, 4, представляются как

$$f(y_1, y_2, y_3, y_4, y_5) = -3y_1 + 4y_2 - y_3 - 2y_4 - 5y_5 \to \min;$$

$$\begin{cases} -2y_1 + 3y_2 - 4y_3 - y_4 - 2y_5 \geq -12, \\ 4y_1 + y_2 + 5y_3 \qquad + y_5 = 10, \\ 2y_2 \qquad - 3y_4 \qquad = -15, \\ -y_1 + 2y_2 - 2y_3 - 2y_4 - 3y_5 \geq -11, \end{cases}$$

а прямые ограничения, согласно правилу 5, записываются в виде

$$y_1 \ge 0,\ y_3 \ge 0,\ y_4 \ge 0,\ y_2,\ y_5 \text{ – произвольного знака.}$$

Следующие утверждения указывают на связь решений прямой и двойственной задач линейного программирования.

Теорема 10 (*достаточное условие оптимальности*). *Если для допустимых планов* x^*, y^* *прямой и двойственной задач справедливо равенство* $c'x^* = b'y^*$, *то* x^* *– решение прямой задачи,* y^* *– решение двойственной задачи.*

Теорема 11 (*теорема существования*). *Для того чтобы прямая и двойственная ЗЛП имели решение, необходимо и достаточно, чтобы каждая из них имела хотя бы один план* (*допустимое решение*).

Теорема 12 (*первая теорема двойственности*). *Если одна из двойственных задач имеет оптимальный план* x^*, *то и другая имеет оптимальный план* y^*, *причем значения целевых функций на этих планах* x^*, y^* *совпадают*: $c'x^* = b'\,y^*$.

Замечание 13. Для задачи о распределении ресурсов теорема 12 означает, что максимально возможный доход от реализации продукции совпадает с минимально возможной стоимостью сырья. Двойственные оценки $y_1, y_2, \ldots, y_m$ указывают величины назначаемых (экономически обусловленных) цен.

Двойственную задачу, как любую ЗЛП, можно решить симплекс-методом. Следует иметь в виду, что решение двойственной задачи можно получить по конечной симплекс-таблице решения прямой задачи. При этом решение y^* двойственной задачи может быть вычислено по формуле $y^{*\prime} = c_{\text{б}}^{*\prime} A_{\text{ст.б}}^{\ *}$. Здесь $c_{\text{б}}^{\ *}$ – вектор-столбец $c_{\text{б}}$ последней симплекс-таблицы прямой задачи, $A_{\text{ст.б}}^{\ *}$ – матрица, составленная из столбцов этой таблицы, соответствующих первоначальным базисным векторам, взятым в том же порядке.

16. УСЛОВНЫЙ ЭКСТРЕМУМ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ. МЕТОД МНОЖИТЕЛЕЙ ЛАГРАНЖА РЕШЕНИЯ ЗАДАЧ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Для иллюстрации понятия «условный экстремум» рассмотрим две задачи.

Пример 16. В пространстве $\mathbf{R}^3$ найти экстремум функции

$$z = ax + by + c \tag{84}$$

при условии, что переменные x и y связаны соотношением

$$x^2 + y^2 = R^2, \quad x \in \mathbf{R}, \;\; y \in \mathbf{R}. \tag{85}$$

Решение. Геометрически задача означает следующее (рис. 16): на эллипсе L_1 пересечения плоскости (84) с цилиндром (85) требуется найти максимальное (или минимальное) значение аппликаты z ($z_{\max}$ или $z_{\min}$).

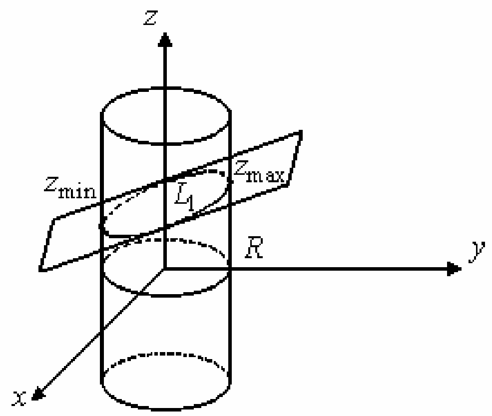


Рис. 16. Демонстрация понятия «условный экстремум»

Эту задачу можно решить следующим образом: из уравнения (85) найдем y как неявно заданную функцию переменной x и, подставив найденное значение $y = y(x)$ в уравнение (84), получим функцию одной переменной x:

$$z = ax \pm b\sqrt{R^2 - x^2} + c, \quad |x| \le R, \tag{86}$$

где знак выбирается в зависимости от параметров a и b.

Таким образом, задачу (84), (85) на условный экстремум сведем к задаче нахождения экстремума функции (86) одной переменной x на отрезке $[-R;\, R]$, которую можно решить стандартными средствами математического анализа. Лучше, однако, задачу решать с помощью замены $x = r\cos\varphi$, $y = r\sin\varphi$, $\varphi \in [0;\, 2\pi]$. ▲

Пример 17. Каковы должны быть размеры консервной банки максимального объема при заданной площади полной ее поверхности S.

Решение. Если предположить, что банка имеет форму цилиндра с радиусом основания r и высотой h, то объем банки будет равен

$$V = \pi r^2 h, \tag{87}$$

причем переменные r и h связаны условием $2\pi r^2 + 2\pi rh = S$, откуда следует

$$h = \frac{S}{2\pi r} - r. \tag{88}$$

Подставив (88) в (87), получим $V = \frac{S}{2}r - \pi r^3$. Таким образом, искомый объем является функцией одной переменной r: $V = V(r)$. Исследуем эту функцию на экстремум: $V'(r) = \frac{S}{2} - 3\pi r^2 = 0$. Следовательно, $r_0 = \sqrt{\frac{S}{6\pi}}$ – стационарная точка функции $V(r)$. Так как $V''(r_0) = -6\pi r_0 < 0$, то в стационарной точке r_0 функция $V(r)$ достигает максимума. Найденному радиусу основания банки r_0, согласно (88), отвечает высота

$$h = \frac{S - 2\pi r_0^{\ 2}}{2\pi r_0} = 2\sqrt{\frac{S}{6\pi}} = 2r_0.$$

Итак, при заданной площади S полной поверхности объем V консервной банки будет максимальным, если ее высота h равна диаметру круга основания, т. е. $h = 2r_0$. ▲

Рассмотренные примеры 16, 17 являются частными случаями более общей задачи математического программирования об отыскании экстремума функции n переменных $f(x_1, x_2, \ldots, x_n)$ при условии, что ее аргументы $x_1, x_2, \ldots, x_n$ удовлетворяют некоторым дополнительным ограничениям. Задача заключается в нахождении такой точки $M_0(x_1^0, x_2^0, \ldots, x_n^0)$, в которой функция $u = f(x_1, x_2, \ldots, x_n)$, называемая *целевой функцией*, достигает *экстремума* (максимума или минимума).

Математическая постановка задачи: найти экстремум функции

$$u = f(x_1, x_2, \ldots, x_n) \to \text{extr} \qquad (89)$$

при условии, что переменные $x_1, x_2, \ldots, x_n$ подчиняются ограничениям

$$F_i(x_1, x_2, \ldots, x_n) = b_i, \quad i = \overline{1, m}, \qquad (90)$$

называемым *уравнениями связи* или просто *связями*.

Задача (89), (90) относится к задачам нелинейного программирования. В курсе математического анализа [11, 20, 23] ее называют задачей на *условный экстремум.* При некоторых предположениях на функции (90) для нахождения решения задачи (89), (90) вводят набор переменных $\lambda_1, \lambda_2, \ldots, \lambda_m$, называемых *множителями Лагранжа*, и составляют *функцию Лагранжа*

$$L(x_1, x_2, \ldots, x_n; \lambda_1, \lambda_2, \ldots, \lambda_m) = f(x_1, x_2, \ldots, x_n) + \\ + \sum_{i=1}^{m} \lambda_i [b_i - F_i(x_1, x_2, \ldots, x_n)], \qquad (91)$$

находят частные производные $\frac{\partial L}{\partial x_j},\ j=\overline{1,\ n},\ \frac{\partial L}{\partial \text{л}_i},\ i=\overline{1,\ m}$, и решают систему $n+m$ уравнений

$$\begin{cases} \frac{\partial L}{\partial x_j}=\frac{\partial f}{\partial x_j}-\sum_{i=1}^{m}\lambda_i\frac{\partial F_i}{\partial x_j}=0, \quad j=\overline{1,\ n}, \\ \frac{\partial L}{\partial \lambda_i}=b_i-F_i(x_1,\ x_2,\ \ldots,\ x_n)=0, \quad i=\overline{1,\ m}, \end{cases} \tag{92}$$

с $n+m$ неизвестными $x_1, x_2, \ldots, x_n; \lambda_1, \lambda_2, \ldots, \lambda_m$. Всякое решение системы уравнений (92) определяет точку $x_0=(x_1^0,\ x_2^0,\ \ldots,\ x_n^0)$, «подозри-тельную» на экстремум функции u из (89). Следовательно, решив систему уравнений (92), получим все точки, в которых функция u может иметь экстремальные значения.

Суть *метода множителей Лагранжа*, справедливого для функций (89), (90) многих переменных, изложим для случая $n = 2$:

$$u = f(x, y) \to \text{extr}, \tag{93}$$

$$F(x, y) = 0, \tag{94}$$

и $n = 3$:

$$u = f(x, y, z) \to \text{extr}, \tag{95}$$

при условии, что переменные x, y, z подчиняются ограничению

$$F(x, y, z) = 0. \tag{96}$$

С геометрической точки зрения задача отыскания экстремума функции $u = f(x, y, z)$ при связи (96) является задачей об экстремуме функции $f(x, y, z)$ в точках поверхности $F(x, y, z) = 0$.

Рассмотрим задачу (93), (94).

Определение 11. Будем говорить, что точка $M_0(x_0; y_0)$, удовлетворяющая условиям связи (94), является точкой *локального условного максимума* (*минимума*) функции $u = f(x, y)$, если существует окрестность $U(M_0)$ точки M_0, такая, что для всех точек $M \in U(M_0)$, удовлетворяющих связи, выполнено неравенство $f(M) \le f(M_0)$ ($f(M) \ge f(M_0)$).

Для задачи (93), (94) составим функцию Лагранжа

$$L(x, y; \lambda) = f(x, y) + \lambda F(x, y).$$

Теорема 13 (*необходимое условие экстремума функции при наличии связи*). *Пусть функции $f(M) = f(x, y)$, $F(M) = F(x, y)$ непрерывно дифференцируемы в некоторой окрестности точки $M_0(x_0; y_0)$, для которой $F(x_0, y_0) = 0$. Если функция $f(x, y)$ в точке $M_0(x_0; y_0)$ при-*

нимает условный экстремум и $\mathrm{grad}F(M_0) \neq 0$, *то существует число* λ, *такое, что в точке* $M_0(x_0; y_0)$ *выполнены равенства*

$$\begin{cases} L'_x(x,y;\lambda) = f'_x(x,y) + \lambda F'_x(x,y) = 0, \\ L'_y(x,y;\lambda) = f'_y(x,y) + \lambda F'_y(x,y) = 0, \\ L'_\lambda(x,y;\lambda) = F(x,y) = 0. \end{cases} \tag{97}$$

Точка $(x_0,\ y_0;\ \lambda_0)$, удовлетворяющая равенствам (97), является *стационарной точкой функции Лагранжа* $L(x, y; \lambda)$.

Выясним достаточные условия существования экстремума функции $u = f(x, y)$ в точке $M_0(x_0; y_0)$ при связи $F(x, y) = 0$.

Введем символ $d^2L = d^2L(x, y; \lambda)$, обозначающий дифференциал второго порядка функции Лагранжа, т. е.

$$d^2L = (dx, dy)\begin{bmatrix} L''_{xx} & L''_{xy} \\ L''_{yx} & L''_{yy} \end{bmatrix}\begin{pmatrix} dx \\ dy \end{pmatrix}. \tag{98}$$

Матрица $\begin{bmatrix} L''_{xx} & L''_{xy} \\ L''_{yx} & L''_{yy} \end{bmatrix}$ вторых частных производных функции Лагранжа в (98) называется *матрицей Гессе* (Л. О. Гессе (1811–1874) – немецкий математик). Если $d^2L > 0$ ($d^2L < 0$) при $(dx)^2 + (dy)^2 \neq 0$, то квадратичная форма (98) считается *положительно* (*отрицательно*) *определенной.*

Теорема 14 (*достаточное условие экстремума функции при наличии связи*). *Пусть функции* $f(x, y)$ *и* $F(x, y)$ *дважды непрерывно дифференцируемы в окрестности стационарной точки* $(x_0,\ y_0;\ \lambda_0)$ *функции Лагранжа* $L(x, y; \lambda) = f(x, y) + \lambda F(x, y)$. *Тогда если*

$$\begin{cases} d^2L(x_0,y_0;\lambda_0) < 0 \quad \text{при } (dx)^2 + (dy)^2 \neq 0, \\ F(x_0,y_0) = 0, \end{cases} \tag{99}$$

то в точке $M_0(x_0; y_0)$ *функция* $u = f(x, y)$ *имеет* ***локальный условный максимум***. *Если же в стационарной точке* $(x_0, y_0; \lambda_0)$

$$\begin{cases} d^2L(x_0,y_0;\lambda_0) > 0 \quad \text{при } (dx)^2 + (dy)^2 \neq 0, \\ F(x_0,y_0) = 0, \end{cases} \tag{100}$$

то в точке $M_0(x_0; y_0)$ *функция* $u = f(x, y)$ *имеет* ***локальный условный минимум.***

Покажем, что достаточные условия экстремума (99), (100) равносильны следующим более простым условиям.

Теорема 15. *Если в стационарной точке* (x_0, y_0; λ_0) *функции Лагранжа* $L(x, y; \lambda)$ *число*

$$Q(x_0, y_0; \lambda_0) = \left(F'_y, \ -F'_x\right)\begin{bmatrix} L''_{xx} & L''_{xy} \\ L''_{yx} & L''_{yy} \end{bmatrix}\begin{pmatrix} F'_y \\ -F'_x \end{pmatrix} \tag{101}$$

меньше нуля, то в точке $M_0(x_0;\ y_0)$ *функция* $u = f(x,\ y)$ *имеет* ***локальный условный максимум.*** *Если* $Q(x_0, y_0; \lambda_0) > 0$, *то в точке* $M_0(x_0;\ y_0)$ *функция* $u = f(x,\ y)$ *имеет* ***локальный условный минимум.***

Доказательство. Первый дифференциал dL функции $L(x, y; \lambda)$, равный $dL(x, y; л) = L'_x dx + L'_y dy$, может быть представлен в виде линейной формы

$$dL(x, y; \lambda) = L'_x dx + L'_y dy = (dx, dy)\begin{bmatrix} L'_x \\ L'_y \end{bmatrix} = (L'_x, L'_y)\begin{bmatrix} dx \\ dy \end{bmatrix},$$

а второй дифференциал d^2L представляет собой квадратичную форму (98), в которой дифференциалы dx и dy связаны соотношением

$$F'_x dx + F'_y dy = 0, \tag{102}$$

вытекающим из уравнения связи $F(x, y) = 0$. Найдя из уравнения (102) дифференциал dy и подставив его в (98), получим

$$d^2L = \left(dx, -\frac{F'_x}{F'_y}dx\right)\begin{bmatrix} L''_{xx} & L''_{xy} \\ L''_{yx} & L''_{yy} \end{bmatrix}\begin{pmatrix} dx \\ -\dfrac{F'_x}{F'_y}dx \end{pmatrix} = (dx)^2\left(1, -\frac{F'_x}{F'_y}\right)\begin{bmatrix} L''_{xx} & L''_{xy} \\ L''_{yx} & L''_{yy} \end{bmatrix}\begin{pmatrix} 1 \\ \dfrac{F'_x}{-F'_y} \end{pmatrix}.$$

Знак d^2L не изменится от умножения обеих частей полученного соотношения на $\left(F'_y\right)^2 > 0$:

$$d^2L \cdot \left(F'_y\right)^2 = (dx)^2\left(F'_y, -F'_x\right)\begin{bmatrix} L''_{xx} & L''_{xy} \\ L''_{yx} & L''_{yy} \end{bmatrix}\begin{pmatrix} F'_y \\ -F'_x \end{pmatrix}.$$

Так как $(dx)^2 > 0$, то знак d^2L совпадает со знаком выражения Q из (101), откуда и следует справедливость теоремы. ▲

Пример 18. Исследовать на экстремум функцию $u = xy$ при условии, что $x^2 + y^2 = 2a^2$.

Решение. Составим функцию Лагранжа:

$$L(x, y; \lambda) = xy + \lambda(x^2 + y^2 - 2a^2).$$

Система уравнений (97) для этой функции имеет вид

$$\begin{cases} L'_x(x, y; \lambda) = y + 2\lambda x = 0, \\ L'_y(x, y; \lambda) = x + 2\lambda y = 0, \\ L'_\lambda(x, y; \lambda) = x^2 + y^2 - 2a^2 = 0, \end{cases}$$

из первого уравнения которой находим $y = -2\lambda x$. После подстановки y во второе уравнение получаем $x(1 - 4\lambda^2) = 0$, откуда следует $л_1 = \frac{1}{2}$, $л_2 = -\frac{1}{2}$. Значению $л_1 = \frac{1}{2}$ соответствуют стационарные точки $\left(a,\ -a;\ \frac{1}{2}\right), \left(-a,\ a;\ \frac{1}{2}\right)$ функции Лагранжа, а значению $л_2 = -\frac{1}{2}$ – стационарные точки $\left(a,\ a;\ -\frac{1}{2}\right)$, $\left(-a,\ -a;\ -\frac{1}{2}\right)$. Это значит, что точками возможного локального условного экстремума функции $u = xy$ являются точки $M_1(a; -a)$, $M_2(a; a)$, $M_3(a; a)$, $M_4(-a; -a)$. Вычислим число Q, определяемое формулой (101), в этих точках. В точке $M_3(a; a)$ имеем: $\left(F'_y, -F'_x\right)\Big|_{M_3} = \left(2y, -2x\right)\Big|_{M_3} = (2a, -2a)$;

$$\begin{bmatrix} L''_{xx} & L''_{xy} \\ L''_{yx} & L''_{yy} \end{bmatrix}\Bigg|_{M_3} = \begin{bmatrix} 2\lambda & 1 \\ 1 & 2\lambda \end{bmatrix}\Bigg|_{\lambda = -\frac{1}{2}} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Следовательно,

$$Q(M_3) = \left(2a, -2a\right)\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{pmatrix} 2a \\ -2a \end{pmatrix} = \left(-4a, 4a\right)\begin{pmatrix} 2a \\ -2a \end{pmatrix} = -16a^2 < 0,$$

т. е. в точке $M_3(a; a)$ функция $u = xy$ имеет *локальный условный максимум*, равный $u_{\max}(M_3) = a^2$. Аналогично найдем: $Q(M_4) = -16a^2 < 0$, и $u_{\max}(M_4) = u_{\max}(M_3) = a^2$. В точках $M_1(a; -a)$, $M_2(-a; a)$ получаем $Q(M_1) = Q(M_2) = 16a^2 > 0$, значит в точках M_1, M_2 функция $u = xy$ имеет *локальный условный минимум:* $u_{\min}(M_1) = u_{\min}(M_2) = -a^2$.

Рассмотрим задачу нахождения экстремума функции (95) при связи (96).

Определение точки *локального условного максимума* (*минимума*) функции $u = f(x, y, z)$ трех переменных для задачи (95), (96) вводится аналогично определению 11.

Для задачи (95), (96) составим функцию Лагранжа

$$L(x, y, z; \lambda) = f(x, y, z) + \lambda F(x, y, z).$$

Стационарная точка $(x_0,\ y_0,\ z_0;\ \lambda_0)$ этой функции определяется системой уравнений

$$\begin{cases} L'_x(x,y,z;\lambda) = f'_x(x,y,z) + \lambda F'_x(x,y,z) = 0, \\ L'_y(x,y,z;\lambda) = f'_y(x,y,z) + \lambda F'_y(x,y,z) = 0, \\ L'_z(x,y,z;\lambda) = f'_z(x,y,z) + \lambda F'_z(x,y,z) = 0, \\ L'_\lambda(x,y,z;\lambda) = F(x,y,z) = 0. \end{cases}$$

Теорема 16. *Если в стационарной точке* $(x_0, y_0, z_0; \lambda_0)$ *функции Лагранжа* $L(x, y, z; \lambda)$ *для задачи (95), (96) матрица* Q *второго порядка*

$$Q(x_0, y_0, z_0; \lambda_0) = \begin{bmatrix} -F'_z & 0 & F'_x \\ 0 & -F'_z & F'_y \end{bmatrix} \begin{bmatrix} L''_{xx} & L''_{xy} & L''_{xz} \\ L''_{yx} & L''_{yy} & L''_{yz} \\ L''_{zx} & L''_{zy} & L''_{zz} \end{bmatrix} \begin{bmatrix} -F'_z & 0 \\ 0 & -F'_z \\ F'_x & F'_y \end{bmatrix}$$

(*соответствующая квадратичная форма*) *положительно (отрицательно) определена, то в этой точке функция* $u = f(x, y, z)$ *имеет локальный условный минимум (максимум).*

• **Задача 22.** Найдите экстремум функции $u = f(x, y, z) = x^2 + y^2 + z^2$ при связи $F(x, y, z) = x + y + z - 4 = 0$.

• **Задача 23.** Сформулируйте задачу математического программирования на условный экстремум при ограничениях типа неравенства.

17. ВАРИАЦИОННОЕ ИСЧИСЛЕНИЕ

Вариационное исчисление – раздел математики, посвященный исследованию методов отыскания экстремумов функционалов, зависящих от одной или нескольких функций, при различного рода ограничениях (фазовых, дифференциальных, интегральных и т. п.), накладываемых на эти функции [4, 8, 14, 21, 24].

Определение 13. Если каждой функции $x = x(\cdot)$ из некоторого множества X функций поставлено в соответствие единственное действительное число $I(x) = I(x(\cdot))$, то говорят, что на множестве X задан *функционал* I, т. е. $I\colon X \to \mathbf{R}$.

Например, наибольшее значение $I(x(\cdot))$ функции $x(\cdot)$ на отрезке $[a, b]$, $I\left(x(\cdot)\right) = \int\limits_0^1 x(t)dt$ и т.д. – функционалы.

Замечание 14. Чтобы различать *элемент* $x(\cdot)$ соответствующего функционального пространства и *значение функции* $x(t)$ при фиксированном t, в литературе вместо $I(x(t))$ часто используются обозначения $I(x(\cdot))$ или $I(x)$.

Первой в историческом плане вариационной задачей была з*адача о брахистохроне*, сформулированная Я. Бернулли в 1696 г. (постановку задачи см. в подразд. 10). В качестве простейшего примера вариационной задачи ниже рассматривается з***адача о кратчайшем расстоянии***.

Задача. На плоскости (t; x) заданы две точки (t_0; x_0) и (T; x_T). Требуется соединить их гладкой кривой $x = x(t)$, $t \in [t_0, T]$, имеющей наименьшую длину.

Составим математическую модель задачи. Поскольку [20, 23] длина кривой $x(t)$, соединяющей две заданные точки (t_0; x_0), (T; x_T), находится по формуле $I\left(x(\cdot)\right) = \int\limits_{t_0}^{T} \sqrt{1 + \left(x'(t)\right)^2}\, dt$, то решение поставленной задачи сводится к определению такой непрерывной функции $x^*(\cdot)$, которая на отрезке [t_0; T] имеет непрерывную производную, удовлетворяет заданным граничным условиям $x^*(t_0) = x_0$, $x^*(T) = x_T$ и на которой функционал $I(x^*(\cdot))$ принимает минимальное значение. Очевидно, решением задачи является прямая (рис. 17), соединяющая две заданные точки (t_0; x_0), (T; x_T).

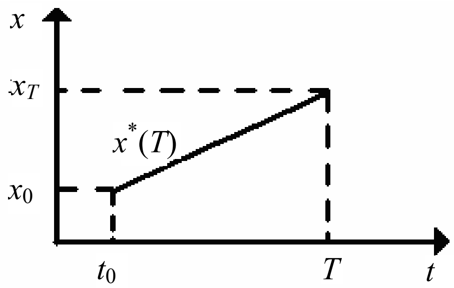


Рис. 17. Иллюстрация задачи о кратчайшем расстоянии

Определение 14. Функционал $I(x(\cdot))$ называется *непрерывным*, если малому приращению функции $x(\cdot)$ соответствует малое изменение функционала.

Понятие малости приращения функции требует, однако, уточнения. С этой целью введем понятие расстояния между функциями.

Пусть функционал $I(x(\cdot))$ определен на элементах $x(\cdot)$ *линейного нормированного пространства* функций, в котором каждому элементу $x(\cdot)$ поставлено в соответствие действительное число $\|x\|$, называемое *нормой* элемента. При этом для любых элементов x и y, принадлежащих пространству, и любого действительного числа λ выполнены следующие условия:

1) $\|x\| \geq 0$, причем $\|x\| = 0$ в том и только в том случае, когда $x = 0$ (0 – нулевой элемент);

2) $\|\lambda x\| = |\lambda| \cdot \|x\|$;

3) $\|x + y\| \leq \|x\| + \|y\|$ (неравенство треугольника).

Величину $\|x-y\|$ будем называть *расстоянием* между элементами x и y [7, 22].

Рассмотрим два пространства: $C^0([t_0;\ T])$ и $C^1([t_0;\ T])$. *Пространство* $C^0([t_0;\ T])$ состоит из *непрерывных функций* (кривых) $x(\cdot)$, определенных на отрезке $[t_0;\ T]$. *Норма* в пространстве $C^0([t_0;\ T])$ определяется соотношением $\|x\|_0 = \max\limits_{t\in[t_0;\,T]}|x(t)|$. *Пространство* $C^1([t_0;\ T])$ состоит из *непрерывных функций* (кривых) $x(\cdot)$, определенных на отрезке $[t_0;T]$ и *имеющих на этом отрезке непрерывную производную*. *Норма* в пространстве $C^1([t_0;\ T])$ вводится согласно формуле

$$\|x\|_1 = \max_{t\in[t_0;\,T]}|x(t)| + \max_{t\in[t_0;\,T]}|x'(t)|.$$

Аналогично вводится норма в пространстве $C^m([t_0;\ T])$ функций, имеющих *непрерывную производную до порядка m* включительно:

$$\|x\|_m = \sum_{p=0}^{m}\max_{t\in[t_0;\,T]}\left|x^p(t)\right|.$$

Пример 19. Найти расстояние $\|x-x^*\|_0$, $\|x-x^*\|_1$ между кривыми $x(t)=t^2$, $x^*(t)=t^3$, $t\in[0;\ 1]$, в пространствах $C^0([0;\ 1])$, $C^1([0;\ 1])$.

Решение. Расстояние в пространстве $C^0([t_0;\ T])$ от кривой x^* до кривой x определим формулой

$$\|x-x^*\|_0 = \max_{t\in[t_0;\,T]}|x(t)-x^*(t)|, \tag{103}$$

а в пространстве $C^1([t_0;\ T])$ –

$$\|x-x^*\|_1 = \max_{t\in[t_0;\,T]}\left|x(t)-x^*(t)\right| + \max_{t\in[t_0;\,T]}\left|x'(t)-x^{*\prime}(t)\right|. \tag{104}$$

Найдем расстояние между кривыми $x(t)=t^2$, $x^*(t)=t^3$, $t\in[0;\ 1]$, в пространстве $C^0([0;\ 1])$. Очевидно, в силу (103) $\|x-x^*\|_0 =$ $= \max\limits_{t\in[0;\,1]}|t^2-t^3| = \max\limits_{t\in[0;1]}(t^2-t^3)$. Поскольку для $u(t)=t^2-t^3$ имеем $u'(t) =$ $= (t^2-t^3)' = t(2-3t)$, то стационарными точками функции $u(\cdot)$ являются точки $t=0$, $t=\dfrac{2}{3}$. Вычисляя значения функции $u(\cdot)$ в этих точках и сравнивая их со значениями на концах промежутка [0; 1], убеждаемся в том, что в точке $t=\dfrac{2}{3}$ достигается *глобальный максимум* функции. Таким образом, расстояние между кривыми x и x^* в пространстве $C^0([0; 1])$ равно

$$\left\|x - x^*\right\|_0 = \left|\frac{4}{9} - \frac{8}{27}\right| = \frac{4}{27}.$$

Найдем расстояние между кривыми $x(t) = t^2$, $x^*(t) = t^3$, $t \in [0; 1]$, в пространстве $C^1([0; 1])$. Согласно (104),

$$\left\|x - x^*\right\|_1 = \max_{t\in[0;\,1]}\left|t^2 - t^3\right| + \max_{t\in[0;\,1]}\left|2t - 3t^2\right|.$$

Для первого слагаемого в силу предыдущих вычислений имеем $\max\limits_{t\in[0;\,1]}\left|t^2 - t^3\right| = \dfrac{4}{27}$. Нетрудно видеть, что глобальный максимум второго слагаемого достигается в точке $t = 1$. Тогда расстояние между кривыми x и x^* в пространстве $C^1([0; 1])$ равно $\left\|x - x^*\right\|_1 = \dfrac{4}{27} + 1 = \dfrac{31}{27}$. ▲

Вариационное исчисление имеет дело с функционалами. Кривые (графики функций $x(\cdot)$), на которых сравниваются значения функционала, называются *допустимыми кривыми* или *кривыми сравнения*. Обозначим через $x^0 = x^0(\cdot)$ допустимую кривую, на которой функционал достигает экстремума, а через $x = x(\cdot)$ – произвольную допустимую кривую. Разность $\delta x = x - x^0$ называется *вариацией кривой* (функции) x^0. Вариация $\delta x(t) = x(t) - x^0(t)$ является функцией аргумента t, $t \in [a; b]$, и принадлежит тому же функциональному пространству, что и функции x, x^0. Используя вариацию δx, допустимую кривую x можно задать уравнением

$$x = x^0 + \delta x \tag{105}$$

либо представить в виде

$$x(t) = x^0(t) + \alpha\delta x(t), \quad t \in [a; b], \tag{106}$$

где δx – фиксированная кривая (точнее функция), α – числовой параметр. Очевидно, при $\alpha = 0$ получаем $x = x^0$.

Приращением $\Delta I(x)$ функционала $I(x)$ называется разность

$$\Delta I(x) = I(x) - I(x^0) = I(x^0 + \alpha\delta x) - I(x^0). \tag{107}$$

Далее будем рассматривать *линейные функционалы на множестве допустимых кривых*, т. е. функционалы, удовлетворяющие следующим условиям:

1) $I(cx) = cI(x)$;

2) $I(x_1 + x_2) = I(x_1) + I(x_2)$,

где c – произвольная постоянная, x, x_1, x_2 – любые допустимые кривые.

Определим *первую вариацию* функционала. Так как функционал $I(x^0 + \alpha\delta x)$ зависит от параметра α, то можно записать $I(x^0 + \alpha\delta x) = \varphi(\alpha)$. Предположим, что функцию $\varphi(\alpha)$ можно представить в окрестности точки $\alpha = 0$ по формуле Тейлора:

$$\Delta I(x) = I(x^0 + \alpha\delta x) - I(x^0) = \alpha\delta I + \frac{1}{2}\alpha^2\delta^2 I + o(\alpha^2). \qquad (108)$$

Коэффициент

$$\delta I = \frac{d\varphi}{d\alpha}\Big|_{\alpha=0} = \frac{d(I(x^0 + \alpha\delta x)}{d\alpha}\Big|_{\alpha=0} \qquad (109)$$

при первой степени параметра α разложения (108) называется ***первой вариацией функционала*** на кривой x^0. Коэффициент

$$\delta^2 I = \frac{d^2\varphi}{d\alpha^2} = \frac{d^2 I(x^0(t) + \alpha\delta x(t))}{d\alpha^2}\Big|_{\alpha=0} \qquad (110)$$

при $\frac{1}{2}\alpha^2$ разложения (108) называется ***второй вариацией функционала*** на кривой x^0.

Аналогично определяются последующие вариации функционала.

Пример 20. Найти первую вариацию функционала $I(x) = \int\limits_a^b x^2(t)dt$ на кривой $x^0 \in C^0([a;\ b])$.

Решение. Согласно формуле (109), сначала найдем для функционала $I(x) = \int\limits_a^b x^2(t)dt$ величину $I(x^0 + \alpha\delta x)$:

$$I(x^0 + \alpha\delta x) = \int\limits_a^b \left(x^0(t) + \alpha\delta x(t)\right)^2 dt.$$

Тогда из (109) следует

$$\delta I = \frac{dI(x^0 + \alpha\delta x)}{d\alpha}\Big|_{\alpha=0} = \frac{d}{d\alpha}\int\limits_a^b \left(x^0(t) + \alpha\delta x(t)\right)^2 dt\Big|_{\alpha=0} =$$

$$= \int\limits_a^b 2\left(x^0(t) + \delta\alpha x(t)\right)\delta x(t)dt\Big|_{\alpha=0} = 2\int\limits_a^b x^0(t)\delta x(t)dt.$$

Таким образом, $\delta I = 2\int\limits_a^b x^0(t)\delta x(t)dt$ – первая вариация исследуемого функционала. ▲

• **Задача 24.** Найдите вторую вариацию функционала $I(x)=\int\limits_a^b x^2(t)dt$ на кривой $x^0 \in C^0([a;b])$.

Определение 15. Говорят, что функционал $I(x(\cdot))$, определенный на классе $\Re$ кривых $x(\cdot)$, достигает на кривой $x^0(\cdot)$ *глобального максимума* (*минимума*), если $I^0(x(\cdot)) \geq I(x(\cdot))$ ($I^0(x(\cdot)) \leq I(x(\cdot))$) для любых кривых $x(\cdot)\in\Re$.

Пример 21. Найти глобальные максимум и минимум функционала $I(x)=\int\limits_0^1 x(t)dt$, где $x\in\Re=\{x_1(t)=t,\ x_2(t)=t^2,\ x_3(t)=-(t-1)^2+1,\ t\in[0;1]\}$.

Решение. Нетрудно видеть, что все кривые $x_i(\cdot)$, $i=\overline{1,\ 3}$, принадлежащие классу функций $\Re$, обладают следующими свойствами: $x_i(0)=0$, $x_i(1)=1$, $i=\overline{1,\ 3}$, т. е. все кривые $x_i(t)$, $i=\overline{1,\ 3}$, при $t=0$ проходят через начало координат, а при $t=1$ – через точку (1; 1) (рис. 18).

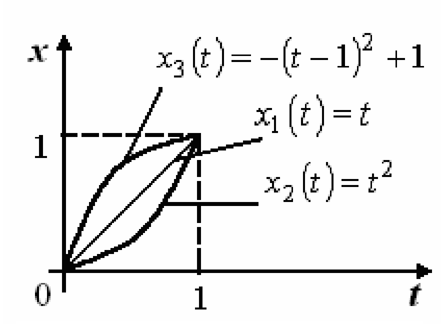


Рис. 18. Кривые, принадлежащие классу $\Re$

Очевидно, что функции $x_2(t)=t^2$, $t\in[0;\ 1]$, на заданном классе $\Re$ соответствует наименьшее значение функционала (на рис. 18 ей соответствует наименьшая площадь под кривой), а функции $x_3(t)$, $t\in[0;\ 1]$, – наибольшее значение (на рис. 18 ей соответствует наибольшая площадь). Следовательно, глобальный минимум, равный $\frac{1}{3}$, достигается на кривой $x_2(\cdot)$, а глобальный максимум, равный $\frac{2}{3}$, – на кривой $x_3(\cdot)$. ▲

Пример 22. Доказать, что на кривой $x^0(t)=t$, $t\in[0;\ 1]$, функционал

$$I(x)=\int\limits_0^1 (x'(t))^2\,dt \tag{111}$$

с граничными условиями

$$x(0)=0,\ x(1)=1, \tag{112}$$

достигает глобального минимума.

Решение. Очевидно, функция $x^0(\cdot)\in C^1([0;1])$. Рассмотрим вариации $\delta x(\cdot)\in C^1([0;1])$ кривой $x^0(\cdot)$, которые в силу (106) могут быть представлены в виде $\delta x(t)=\frac{1}{\alpha}(x(t)-x^0(t))$, $t\in[0;1]$. Из граничных условий (112) следует, что $\delta x(\cdot)$ удовлетворяют условиям $\delta x(0) = \delta x(1) = 0$. Исследуем приращение $I(x^0(\cdot) + \alpha\delta x(\cdot)) - I(x^0(\cdot))$ функционала (111). Поскольку $(\delta x(t))' = \delta x'(t)$, то

$$I(x^0(\cdot)+\alpha\delta x(\cdot))-I(x^0(\cdot))=\int_0^1((x^0(t)+\alpha\delta x(t))')^2dt-\int_0^1((x^0(t)')^2dt=$$

$$=\underbrace{2\alpha\int_0^1(x^0(t))'\delta x'(t)dt}_{=0}+\int_0^1(\alpha\delta x'(t))^2dt\ge 0,$$

так как $(x^0(t))' = 1$ и, следовательно, $\int_0^1(x^0(t))'\delta x'(t)dt=\int_0^1\delta x'(t)dt=$ $=\delta x(t)\Big|_0^1=0$. Поскольку кривая $x(\cdot) = x^0(\cdot) + \alpha\delta x(\cdot)\in C^1([0;1])$ произвольна и $I(x(\cdot)) = I(x^0(\cdot) + \alpha\delta x(\cdot)) \ge I(x^0(\cdot))$, то на функции $x^0(t) = t$, $t\in[0;1]$, достигается глобальный минимум. ▲

Понятие *локального минимума* (*максимума*) связано с исследованием поведения функционала на близких кривых. Различают *сильный* и *слабый локальный максимум* (*минимум*).

Пусть $x^0(\cdot)\in C^0([t_0;T])$ и $\varepsilon > 0$ – произвольное достаточно малое число.

Определение 16. ***ε-окрестностью нулевого порядка*** кривой $x^0(\cdot)$ называется совокупность $O_\varepsilon(x^0)$ кривых $x(\cdot)\in C^0([t_0;T])$, таких, что

$$\left\|x-x^0\right\|_0=\max_{t\in[t_0;\,T]}\left|x(t)-x^0(t)\right|<\varepsilon.$$

Это неравенство означает, что расстояние от кривой $x^0(\cdot)$, $x^0(t_0) = x_0$, $x^0(T) = x_T$, до кривых $x(\cdot)\in O_e(x^0)$ мало (меньше ε в смысле близости нулевого порядка) (рис. 19).

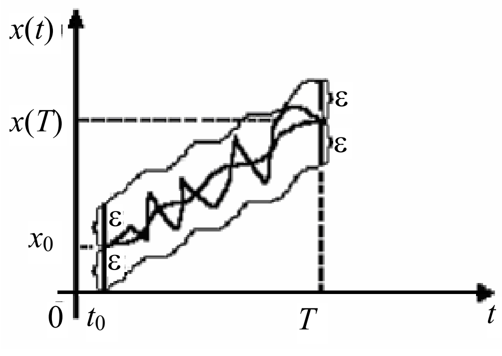


Рис. 19. ε-окрестность нулевого порядка кривой $x^0(\cdot)$

Определение 17. Говорят, что функционал $I(x(\cdot))$ достигает на кривой $x^0(\cdot)$ ***сильного локального минимума*** (***максимума***), если

$$I^0(x(\cdot)) \le I(x(\cdot))\ (I^0(x(\cdot)) \ge I(x(\cdot)))$$

для всех кривых $x(\cdot)$из некоторой ε-окрестности нулевого порядка кривой $x^0(\cdot)$.

Пусть $x^0(\cdot) \in C^1([t_0; T])$ и ε > 0 – произвольное достаточно малое число.

Определение 18. ***ε-окрестностью первого порядка*** кривой $x^0(\cdot)$ называется совокупность кривых $x(\cdot) \in C^1([t_0; T])$, таких, что

$$\left\| x - x^0 \right\|_1 = \max_{t \in [t_0;\, T]} \left| x(t) - x^0(t) \right| + \max_{t \in [t_0;\, T]} \left| x'(t) - x^{0\prime}(t) \right| < \mathrm{e}\,.$$

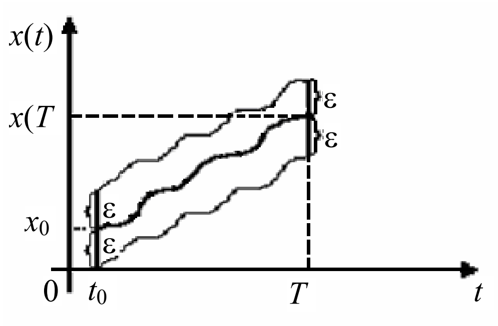


Рис. 20. ε-окрестность первого порядка кривой $x^0(\cdot)$

Согласно этому определению, у кривых $x^0(\cdot)$ и $x(\cdot)$ близки не только ординаты, но и значения производных в одинаковых точках отрезка $[t_0;\ T]$ (рис. 20). Отсюда следует, что ***кривая, принадлежащая ε-окрестности первого порядка, принадлежит и ε-окрестности нулевого порядка, но не наоборот.***

Определение 19. Говорят, что функционал $I(x(\cdot))$ достигает на кривой $x^0(\cdot)$ *слабого локального минимума* (*максимума*), если

$$I^0(x(\cdot)) \le I(x(\cdot))\ (I^0(x(\cdot)) \ge I(x(\cdot)))$$

для всех кривых $x(\cdot)$ из некоторой ε-окрестности первого порядка кривой $x^0(\cdot)$.

Локальные минимумы и максимумы функционала называются его *локальными экстремумами*.

Замечание 15. Всякий сильный экстремум функционала является и слабым экстремумом, а обратное, вообще говоря, не верно, так как сильный экстремум – это экстремум по отношению к *более широкому* классу допустимых кривых.

Необходимые условия локального экстремума на языке вариаций одинаковы для сильного и слабого минимума (максимума) и определяются следующей теоремой.

Теорема 17 (*необходимые условия локального экстремума*). *Если функционал $I(x(\cdot))$, имеющий первую вариацию, достигает минимума (максимума) на кривой $x^0(\cdot)$, где $x^0(\cdot)$ – внутренняя кривая области определения функционала, то на этой кривой первая вариация функционала равна нулю: $\delta I(x^0(\cdot)) = 0$.*

Доказательство теоремы, которое здесь не приводится, основано на следующей важной лемме.

Основная лемма вариационного исчисления [4, 21, 24]**.** *Если для каждой непрерывной функции* $\eta(\cdot)\in C^1([t_0;T])$, *удовлетворяющей условию* $\eta(t_0)=\eta(T)=0$, *выполняется равенство*

$$\int_{t_0}^{T} a(t)\eta(t)dt=0, \tag{113}$$

где $a(\cdot)$ *– непрерывная на отрезке* $[t_0;\,T]$ *функция, то* $a(t)=0,\ t\in[t_0;\ T]$.

17.1. Постановка простейшей задачи вариационного исчисления

Рассмотрим множество $\mathfrak{R}$ *допустимых функций (кривых)* $x(\cdot)$, $x(\cdot)\in\mathfrak{R}$, удовлетворяющих условиям:

1) $x(\cdot)\in C^1([t_0;T])$;

2) функции $x(\cdot)$ удовлетворяют *граничным условиям*

$$x(t_0)=x_0,\ x(T)=x_T, \tag{114}$$

где значения x_0 и x_T заданы (т. е. все кривые проходят через две закрепленные граничные точки x_0, x_T). На множестве кривых $x(\cdot)\in\mathfrak{R}$ задан функционал

$$I(x(\cdot))=\int_{t_0}^{T} F(t,\ x(t),\ x'(t))dt, \tag{115}$$

где подынтегральная функция $F(t,\ x(t),\ x'(t))$ трижды дифференцируема по своим переменным (достаточно существования непрерывных частных производных до второго порядка включительно).

Задача. Среди допустимых кривых $x(\cdot)\in\mathfrak{R}$ найти такую кривую $x^0(\cdot)$, на которой функционал (115) достигает экстремума.

Таким образом, рассматривается следующая экстремальная задача:

$$I(x(\cdot))\to \text{extr},$$
$$x(\cdot)\in\mathfrak{R}$$

или, что то же самое,

$$I(x^0(\cdot))=\underset{x(\cdot)\in\mathfrak{R}}{\text{extr}}\ I(x(\cdot)). \tag{116}$$

Так как на кривые $x(\cdot)$, образующие множество $\mathfrak{R}$, не наложено никаких дополнительных условий, кроме граничных условий (114), то задача (116) – *простейшая задача вариационного исчисления (ПЗВИ)* – называется ***задачей поиска безусловного экстремума.***

Стратегия поиска решения ПЗВИ (116) состоит в определении и приравнивании к нулю *первой вариации* δI функционала (115) (в соответствии с теоремой 17 о необходимых условиях экстремума функционала). В результате получаем соотношения, позволяющие найти кривые, «подозрительные» на экстремум функционала.

С помощью анализа *второй вариации* функционала выводятся достаточные условия экстремума – слабого минимума или максимума.

Теорема 18 (*необходимые условия экстремума в ПЗВИ*). *Если на кривой* $x^0(\cdot) \in C^1\left(\left[t_0; T\right]\right)$, *удовлетворяющей граничным условиям (114), достигается слабый локальный экстремум функционала в ПЗВИ (116), то* $x^0(\cdot)$ *удовлетворяет уравнению*

$$F'_x - \frac{d}{dt} F'_{x'} = 0, \tag{117}$$

где $F'_x = \dfrac{\partial F(t, x(t), x'(t))}{\partial x}$, $F'_{x'} = \dfrac{\partial F(t, x(t), x'(t))}{\partial x'}$.

Доказательство. Пусть $x^0(\cdot)$ – кривая, на которой достигается экстремум функционала (116). Тогда допустимая кривая $x(\cdot)$ в силу (106) может быть представлена в виде $x(\cdot) = x^0(\cdot) + \alpha\delta x(\cdot)$, а ее производная $x'(\cdot)$ – как $x'(\cdot) = x^{0'}(\cdot) + \delta x'(\cdot))$, где $\delta x(\cdot)$ – фиксированная вариация кривой, $\delta x'(\cdot) = (\delta x(\cdot))'$– производная вариации, α – числовой параметр. На основании предположения 2 для допустимых кривых (см. формулу (114)) и предположений теоремы 18 следует: $\delta x(\cdot) \in C^1\left(\left[t_0; T\right]\right)$,

$$\delta x(t_0) = \frac{1}{\alpha}(x(t_0) - x^0(t_0)) = \frac{1}{\alpha}(x_0 - x_0) = 0,$$

$$\delta x(T) = \frac{1}{\alpha}(x(T) - x^0(T)) = \frac{1}{\alpha}(x_T - x_T) = 0. \tag{118}$$

На допустимой кривой $x(\cdot)$ рассмотрим функционал:

$$I(x^0(\cdot)) + \alpha\delta x(\cdot)) = \int_{t_0}^{T} F(t, x^0(t) + \alpha\delta x(t),\ x^{0'}(t) + \alpha\delta x'(t))dt = \varphi(\alpha), \tag{119}$$

где $\varphi(\alpha)$ – функция числового параметра α. Используя формулу (110)

$$\delta I = \frac{d\varphi(\alpha)}{d\alpha}\Big|_{\alpha=0} = \frac{dI\left(x^0(t)+\alpha\delta x(t)\right)}{d\alpha}\Big|_{\alpha=0}$$

для вычисления первой вариации функционала, получим

$$\delta I = \frac{d}{d\alpha}\int_{t_0}^{T} F(t,\ \underbrace{x^0(t)+\alpha\delta x(t)}_{x(t)},\ \underbrace{x^{0\prime}(t)+\alpha\delta x'(t)}_{x'(t)})\Big|_{\alpha=0} dt =$$

$$= \int_{t_0}^{T}[F_x'(t,\ x^0(t)+\alpha\delta x(t),\ x^{0\prime}(t)+\alpha\delta x'(t))\Big|_{\alpha=0}\delta x(t) + \qquad (120)$$

$$+ F_{x'}'(t,\ x^0(t)+\alpha\delta x(t),\ x^{0\prime}(t)+\alpha\delta x'(t))\Big|_{\alpha=0}\delta x'(t))]dt =$$

$$= \int_{t_0}^{T}[F_x(t,\ x^0(t),\ x^{0\prime}(t))\delta x(t) + F_{x'}'(t, x^0(t),\ x^{0\prime}(t))\delta x'(t)]dt = \otimes.$$

В (120) проинтегрируем второе слагаемое по частям:

$$\int_{t_0}^{T}\underbrace{F_{x'}'(t,\ x^0(t),\ x^{0\prime}(t))}_{u}\delta x'(t)dt = \left|\begin{aligned} & u = F_{x'}'(t,\ x^0(t),\ x^{0\prime}(t)) \\ & dv = \delta x'(t)dt \\ & du = \frac{d}{dt}\ F_{x'}'dt \\ & v = \delta x(t) \end{aligned}\right| \overset{(117)}{=}$$

$$= F_{x'}'(t,\ x^0(t),\ x^{0\prime}(t))\underbrace{\delta x(t)\Big|_{t_0}^{T}}_{=0} - \int_{t_0}^{T}\frac{d}{dt}F_{x'}'(t,\ x^0(t),\ x^{0\prime}(t))\delta x(t)dt =$$

$$= -\int_{t_0}^{T}\frac{d}{dt}F_{x'}'(t,\ x^0(t),\ x^{0\prime}(t))\delta x(t)dt,$$

где учтено, что $\delta x'(\cdot) = (\delta x(\cdot))'$. Следовательно, выражение $\otimes$ в (120) принимает вид

$$\delta I = \otimes = \int_{t_0}^{T}\left[F_x'(t,\ x^0(t),\ x^{0\prime}(t)) - \frac{d}{dt}F_{x'}'(t,\ x^0(t),\ x^{0\prime}(t))\right]\delta x(t)dt.$$

В силу теоремы 17 должно выполняться равенство $\delta I = 0$, т. е.

$$\delta I = \int_{t_0}^{T}\left[F_x(t,\ x^0(t),\ x^{0\prime}(t)) - \frac{d}{dt}F_{x'}(t,\ x^0(t),\ x^{0\prime}(t))\right]\delta x(t)dt = 0. \qquad (121)$$

К (121) применима основная лемма вариационного исчисления, так как в силу наложенных в теореме 18 ограничений на кривой $x^0(\cdot)$

функция $F'_x - \dfrac{d}{dt}F'_x$ является непрерывной, а вариация $\delta x(\cdot)$ – *произвольной* непрерывно дифференцируемой функцией, удовлетворяющей условиям $\delta x(t_0) = 0$, $\delta x(T) = 0$. Следовательно, кривая $x^0(\cdot)$, на которой достигается экстремум функционала, в силу (113) удовлетворяет уравнению

$$F'_x(t,\, x^0(t),\, x^{0\prime}(t)) - \frac{d}{dt}F'_{x'}(t,\, x^0(t),\, x^{0\prime}(t)) = 0,$$

или, в более краткой записи,

$$F'_x - \frac{d}{dt}F'_{x'} = 0\,. \tag{122}$$

Теорема 18 полностью доказана. ▲

Уравнение (122) называется ***уравнением Эйлера***. *Интегральные кривые* уравнения Эйлера (т. е. графики решений уравнения (122)) называются ***экстремалями Эйлера.*** В развернутой форме уравнение Эйлера (122) принимает вид

$$F'_x - F''_{x't} - F''_{x'x}x' - F''_{x'x'}x'' = 0, \tag{123}$$

где $F'_x = \dfrac{\partial F}{\partial x}$, $F''_{x't} = \dfrac{\partial^2 F}{\partial t \partial x}$, $F''_{x'x} = \dfrac{\partial^2 F}{\partial x \partial x'}$, $F''_{x'x'} = \dfrac{\partial^2 F}{\partial x' \partial x'}$, и при $F''_{x'x'} \neq 0$ представляет собой *обыкновенное дифференциальное уравнение второго порядка* относительно неизвестной функции $x(\cdot)$. Его общее решение $x(t) = x(t, C_1, C_2)$, $t \in [t_0; T]$, зависит от двух произвольных постоянных C_1, C_2 и определяет двухпараметрическое *семейство экстремалей.* Граничные условия $x(t_0) = x_0$, $x(T) = x_T$ позволяют определить C_1, C_2, а следовательно, и кривую $x^0(\cdot)$, на которой может достигаться экстремум функционала. ***Только на удовлетворяющих граничным условиям экстремалях может реализоваться экстремум***. Чтобы выяснить, достигается ли на найденной экстремали экстремум функционала, и если достигается, то какой (максимум или минимум), следует использовать достаточные условия, связанные со второй вариацией функционала.

Замечание 16. Краевая задача (122), (114) не всегда имеет решение, а если решение существует, то оно может быть не единственным.

Замечание 17. На практике случается, что существование решения ПЗВИ очевидно. Тогда если решение краевой задачи (122), (114) единственно, то экстремаль $x^0(\cdot)$ будет решением ПЗВИ.

Замечание 18. Уравнение Эйлера (122) интегрируется в квадратурах лишь в исключительных случаях.

17.2. Простейшие случаи интегрируемости уравнения Эйлера

1. ***Функция*** $F(t, x, x')$ ***не зависит явно от*** x**,** т. е. $F(t, x, x') = F(t, x')$. Уравнение Эйлера (123) принимает вид $\frac{d}{dt}F'_{x'} = 0$, откуда следует

$$F'_{x'} = C. \tag{124}$$

Выражение (124) называется *первым интегралом уравнения Эйлера*.

2. ***Функция*** $F(t, x, x')$ ***не зависит явно от*** t ***и*** x, т. е. $F(t, x, x') = F(x')$. Тогда уравнение Эйлера (123) принимает вид $F''_{x'x'}x'' = 0$. Если $F''_{x'x'} \neq 0$, то $x'' = 0$, и общее решение уравнения (123) имеет вид $x(t) = C_1 t + C_2$, $t \in [t_0; T]$. В этом случае *экстремалями* являются прямые. В случае $F''_{x'x'} = 0$ требуется дополнительное исследование.

3. ***Функция*** $F(t, x, x')$ ***не зависит явно от*** t ***и*** x', т. е. $F(t, x, x') = F(x)$. Уравнение Эйлера (123) превращается в уравнение $F'_x = 0$ и не является дифференциальным. Его решение не содержит произвольных постоянных и поэтому может не удовлетворять граничным условиям. Однако, если соответствующая кривая проходит через граничные точки $(t_0; x_0)$ и $(T; x_T)$, то «подозрительная на экстремум» экстремаль существует.

4. ***Функция*** $F(t, x, x')$ ***линейно зависит от*** x', т. е. $F(t, x, x') = P(t, x) + x'Q(t, x)$. Уравнение Эйлера (123) принимает вид

$$\frac{\partial P}{\partial x} - \frac{\partial Q}{\partial t} = 0. \tag{125}$$

Уравнение (125) также не является дифференциальным, и его решение не всегда удовлетворяет граничным условиям. Если $\frac{\partial P}{\partial x} = \frac{\partial Q}{\partial t}$, то под знаком интеграла в (115) стоит *полный дифференциал* и, следовательно, величина интеграла не зависит от пути интегрирования. В этом случае вариационная задача теряет смысл.

5. ***Функция*** $F(t, x, x')$ ***не зависит явно от*** t, т. е. $F(t, x, x') = F(x, x')$. В этом случае уравнение Эйлера принимает вид

$$F'_x - F''_{x'x}x' - F''_{x'x'}x'' = 0. \tag{126}$$

Умножив обе части равенства (126) на $x' \neq 0$, получим

$$\frac{d}{dt}\left[F - x'F'_{x'}\right] = 0.$$

Действительно, $\frac{d}{dt}\left[F(x,\, x') - x'F'_{x'}(x,\, x')\right] \;=\; F_x x' + F'_{x'}x'' - x''F'_{x'} \;-$
$- x'F''_{x'x}x' - x'F''_{x'x'}x'' = x'\left(F'_x - F''_{x'x}x' - F''_{x'x'}x''\right)$. Тогда уравнение Эйлера (126) может быть переписано в виде

$$\frac{d}{dt}\left[F - x'F'_{x'}\right] = 0.$$

Проинтегрировав его, получим ***первый интеграл*** [23]

$$F - x'F'_{x'} = C\,. \qquad (127)$$

Замечание 19. Зачастую непосредственное применение уравнения Эйлера (117) оказывается проще использования первых интегралов.

6. ***Функция*** $F(t, x, x')$ ***не зависит явно от*** x', т. е. $F(t, x, x') = F(t, x)$. В этом случае уравнение Эйлера (123) сводится к уравнению $\frac{\partial F}{\partial x} = 0$, т. е. уравнение Эйлера не является дифференциальным и для него справедливы все рассуждения п. 3 настоящего подраздела.

17.3. Алгоритм применения необходимых условий экстремума

Теорема 18 дает необходимые условия экстремума функционала (115) на множестве допустимых функций, удовлетворяющих граничным условиям (114). Приведем алгоритм применения этих условий.

1. Найти F'_x, $F'_{x'}$, $\frac{d}{dt}F'_{x'}$ и записать уравнение Эйлера (122). Если функция $F(t, x, x')$ соответствует какому-либо случаю интегрируемости 1–6 подразд. 17.2, то используем соотношения (124)–(127).

2. Найти общее решение Эйлера $x = x(t, C_1, C_2)$, где C_1, C_2 – произвольные постоянные.

3. Определить C_1, C_2 из граничных условий $x(t_0, C_1, C_2) = x_0$, $x(T, C_1, C_2) = x_T$.

В результате получим экстремаль $x^0(\cdot)$, на которой может достигаться экстремум функционала.

Пример 23. Найти экстремаль функционала

$$I(x(\cdot)) = \int_0^1 \left[x^2(t) + x'^2(t)\right]dt,$$

удовлетворяющую граничным условиям $x(0) = 0$, $x(1) = 1$.

Решение. Действуем согласно описанному выше алгоритму применения необходимых условий экстремума функционалов.

1. Из вида подынтегральной функции $F(x, x') = x^2 + x'^2$ находим $F'_x = 2x, \quad F'_{x'} = 2x', \quad \frac{d}{dt}F'_{x'} = 2x''$. Следовательно, уравнение Эйлера (122) имеет вид $2x - 2x'' = 0$, или $x - x'' = 0$.

2. Находим общее решение уравнения Эйлера $x - x'' = 0$ – однородного дифференциального уравнения второго порядка с постоянными коэффициентами. Составляем для него характеристическое уравнение $\lambda^2 - 1 = 0$. Корнями этого уравнения являются числа $\lambda_1 = 1$, $\lambda_2 = -1$. Тогда общее решение имеет вид $x(t) = C_1e^{\lambda_1 t} + C_2e^{\lambda_2 t} = C_1e^t + C_2e^{-t}$, $t \in [0; 1]$.

3. Из граничных условий $x(0) = 0$, $x(1) = 1$ находим C_1, C_2:

$$\begin{cases} x(0) = C_1 + C_2 = 0, \\ x(1) = C_1e + C_2e^{-1} = 1, \end{cases} \Rightarrow C_1 = \frac{e}{e^2 - 1}, \quad C_2 = \frac{e}{1 - e^2}.$$

Следовательно, искомой экстремалью $x^0(\cdot)$ является кривая

$$x^0(t) = \frac{e}{e^2 - 1}e^t + \frac{e}{1 - e^2}e^{-t}, \quad t \in [0; 1]. \ \blacktriangle$$

Пример 24. **Задача о брахистохроне (*о линии быстрейшего ската*).** Среди всех гладких кривых, соединяющих точки $A(0; 0)$ и $B(t_1; x_1)$, найти такую, по которой материальная точка, двигаясь под действием силы тяжести из точки A с нулевой начальной скоростью, достигает точки B за кратчайшее время (см. рис. 10).

Решение. Математическая модель задачи была приведена ранее (см. подразд. 10) и в обозначениях подразд. 17.1 имеет вид

$$I(x(\cdot)) = \frac{1}{\sqrt{2g}}\int_0^{t_1}\frac{\sqrt{1 + x'^2(t)}}{\sqrt{x(t)}}dt \to \min, \ x(0) = 0, \ x(t_1) = x_1.$$

Поскольку постоянный коэффициент $\frac{1}{\sqrt{2g}}$ влияет только на величину функционала и не влияет на процесс нахождения экстремалей, в дальнейшем его учитывать не будем. Для решения задачи применим алгоритм подразд. 17.3.

1. Подынтегральная функция $F(x(t), x'(t)) = \frac{\sqrt{1 + x'^2(t)}}{\sqrt{x(t)}}$ не зависит явно от t. Следовательно, имеем случай 5 интегрируемости урав-

нения Эйлера и, стало быть, его первый интеграл вида $F - x'F'_{x'} = C_1$. Элементарные вычисления дают $\frac{1}{\sqrt{x(1+x'^2)}} = C_1$, откуда

$$x(1+x'^2) = \frac{1}{C_1^2} = C. \qquad (128)$$

2. Для нахождения общего решения полученного уравнения Эйлера введем параметр τ, полагая $\frac{dx}{dt} \equiv x'(t) = \operatorname{ctg}\tau$. Тогда из (128) имеем $x(1 + \operatorname{ctg}^2\tau) = C$ или $x = \frac{C}{1+\operatorname{ctg}^2\tau} = C\sin^2\tau$, откуда $dx = 2C\sin\tau \cdot \cos\tau \cdot d\tau$ и

$$dt = \frac{dx}{x'} = \frac{2C\sin\tau \cdot \cos\tau \cdot d\tau}{\operatorname{ctg}\tau} = 2C\sin^2\tau \cdot d\tau = C\left(1-\cos 2\tau\right)d\tau,$$

т. е. $dt = C(1 - \cos 2\tau)\, d\tau$. Интегрируя, получаем $t(\tau) = C\left(\tau - \frac{\sin 2\tau}{2}\right) + C_2 = \frac{C}{2}\left(2\tau - \sin 2\tau\right) + C_2$. Следовательно, параметрическое уравнение искомой линии имеет вид

$$x(\tau) = C\sin^2\tau = \frac{C}{2}\left(1-\cos 2\tau\right),\ \ t(\tau) = \frac{C}{2}\left(2\tau - \sin 2\tau\right) + C_2,$$

т. е.

$$x(\tau) = \frac{C}{2}\left(1-\cos 2\tau\right); \quad t(\tau) = \frac{C}{2}\left(2\tau - \sin 2\tau\right) + C_2. \qquad (129)$$

3. Так как по условию $x(0) = 0$, то в силу $x(\tau) = C\sin^2\tau$ следует $\tau = 0$, и из второго уравнения (129) получаем $C_2 = 0$. Вводя новую переменную $p = 2\tau$, приходим к выводу, что решением задачи о брахистохроне является ***семейство обыкновенных циклоид***:

$$\begin{cases} x = \dfrac{C}{2}\left(1-\cos p\right), \\ t = \dfrac{C}{2}\left(p - \sin p\right), \end{cases} \quad p \in [0;\, p_1],$$

где параметры C и p_1 определяются из условия прохождения циклоиды через заданные точки $A(0;\, 0)$ и $B(t_1;\, x_1)$ (рис. 21).

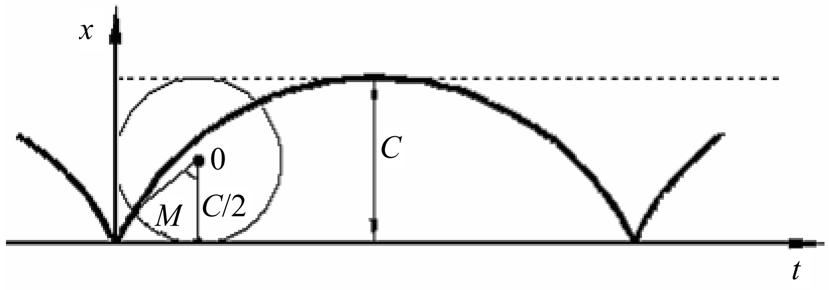


Рис. 21. Решение задачи о брахистохроне – обыкновенная циклоида

Циклоида – линия, которую описывает точка M, расположенная на расстоянии d от центра круга радиусом $r = \dfrac{C}{2}$, катящегося без скольжения по прямой – оси Ot, если в начальный момент точка M совпала с точкой отсчета – началом координат. Если $d = r$, циклоида называется *обыкновенной*, при $d > r$ – *удлиненной*, $d < r$ – *укороченной*. ▲

- **Задача 25.** Найдите экстремаль функционала

$$I(x(\cdot)) = \int_0^1 \left[x^2(t) + x'^2(t) + 2x(t)e^t \right] dt, \quad x(0) = 0, \ x(1) = 0.$$

- **Задача 26*.** Найдите экстремаль функционала

$$I(x(\cdot)) = \int_1^2 \left[3t(x'(t))^5 - 5x(t)(x'(t))^4 \right] dt, \quad x(1) = 1, \ x(2) = 4.$$

17.4. Функционалы, зависящие от производных высших порядков. Уравнение Эйлера – Пуассона

В подразд. 17.1 рассмотрен функционал (115), зависящий от функции $x(\cdot)$ и ее производной $x'(\cdot)$. На практике встречаются случаи зависимости функционала от производных функции $x(\cdot)$ более высоких порядков.

Задача. Среди кривых $x = x(\cdot)$, $x(\cdot) \in C^{(n)}([t_0;\ t_1])$, соединяющих две данные точки $A(t_0;\ x_0)$, $B(t_1;\ x_1)$, определить ту, вдоль которой функционал

$$I(x(\cdot)) = \int_{t_0}^{t_1} F(t,\ x(t),\ x'(t),\ \ldots,\ x^{(n)}(t))dt \qquad (130)$$

принимает экстремальное значение и удовлетворяет $2n$ граничным условиям

$$x(t_0) = x_0, x'(t_0) = x_{1,\,0},\ \ldots, x^{(n-1)}(t_0) = x_{n-1,\,0},$$
$$x(t_1) = x_1, x'(t_1) = x_{1,\,1},\ \ldots, x^{(n-1)}(t_1) = x_{n-1,\,1}, \qquad (131)$$

где искомая кривая $x(\cdot)\in C^{(n)}([t_0;t_1])$, а подынтегральная функция $F(t, x(t), x'(t), \ldots, x^{(n)}(t))$ дифференцируема $n+2$ раза по всем своим аргументам.

Теорема 19. *Если кривая $x = x(\cdot)$ принадлежит классу допустимых кривых (т. е. $x(\cdot)\in C^{(n)}([t_0;t_1])$) и доставляет экстремум функционалу (130), то она удовлетворяет* ***уравнению Эйлера – Пуассона***

$$F'_x-\frac{d}{dt}F'_{x'}+\frac{d^2}{dt^2}F'_{x''}-\ldots+(-1)^n\frac{d^n}{dt^n}F'_{x^{(n)}}=0. \qquad (132)$$

Доказательство этой теоремы аналогично доказательству теоремы 18 о необходимых условиях экстремума функционала $\int\limits_{t_0}^{t_1}F(t,x(t),x'(t))dt$ с применением многократного интегрирования по частям и основной леммы вариационного исчисления.

Уравнение (132) – дифференциальное уравнение порядка $2n$. Его решение дает *экстремали* рассматриваемой вариационной задачи. Общее решение содержит $2n$ произвольных постоянных, которые могут быть определены из $2n$ граничных условий (131).

Пример 25. Найти экстремаль функционала $I(x(\cdot))=\int\limits_0^1\left(1+(x''(t))^2\right)dt$, удовлетворяющую граничным условиям $x(0) = 0$, $x'(0) = 1$, $x(1) = 1$, $x'(1) = 1$.

Решение. Поскольку подынтегральная функция $F(t, x, x', x'') = 1 + (x''(t))^2$ не зависит явно от t, x, x' и $F'_{x''}=2x''$, то уравнение Эйлера – Пуассона (132) (для примера 25 $n=2$) сводится к виду $\frac{d^2}{dt^2}(2x'')=0$. Это уравнение приводит к дифференциальному уравнению $x^{(\mathrm{IV})}(t) = 0$ четвертого порядка, общее решение которого имеет вид $x(t) = C_1t^3 + C_2t^2 + C_3t + C_4$. Тогда

$$\begin{cases} x(t)=C_1t^3+C_2t^2+C_3t+C_4, \\ x'(t)=3C_1t^2+2C_2t+C_3, \end{cases} \qquad (133)$$

и, используя заданные граничные условия задачи, получаем

$$x(0) = C_4 = 0,\ x'(0) = C_3 = 1,$$

$$x(1) = C_1 + C_2 + C_3 + C_4 = 1,\ x'(1) = 3C_1 + 2C_2 + C_3 = 1,$$

откуда $C_4 = 0$, $C_3 = 1$, $C_1 = C_2 = 0$. Таким образом, искомой экстремалью является кривая $x(\cdot) = t$. ▲

18. ВВЕДЕНИЕ В ТЕОРИЮ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ

Теория оптимального управления (ТОУ) отражает современный этап развития вариационного исчисления. Она возникла в 50-е годы XX столетия в связи с необходимостью решения многочисленных прикладных задач. Эти задачи потребовали разработки новых методов решения.

К основным методам ТОУ относится *принцип максимума Понтрягина*, а также метод динамического программирования Р. Беллмана, который часто используется для получения *достаточных условий оптимальности*.

18.1. Простейшая задача оптимального управления

Пусть поведение движущегося объекта управления описывается уравнением (системой n дифференциальных уравнений)

$$\dot{x}(t) = f(x(t), u(t), t),\ \ t \in T = [t_0; t_1] \tag{134}$$

с начальным условием

$$x(t_0) = x_0, \tag{135}$$

где $x(t)$ – вектор *состояния* объекта, $x(t) \in \mathbf{R}^n$; $u(t)$ – вектор управляющих воздействий (*управление*), $u(t) \in \mathbf{R}^r$; t_0 и t_1 – начальный и конечный моменты переходного процесса (управления поведением объекта); x_0 – начальное состояние, $x_0 \in \mathbf{R}^n$, в котором находился объект в момент времени t_0. Управляющие воздействия $u = u(t)$ выбираются из класса *допустимых управлений*, состоящего из кусочно-непрерывных r-вектор-функций, удовлетворяющих условию

$$u(t) \in U \subset \mathbf{R}^r,\ \ t \in T. \tag{136}$$

Качество переходного процесса (допустимого управления) оценивается функционалом

$$I(u) = \varphi(x(t_1)), \tag{137}$$

который определен на конечных (*терминальных*) состояниях системы (134) и называется *терминальным критерием качества* [5, 6].

Задача. Среди множества допустимых управлений выбрать такое $u^0(t)$, $t \in T$, что порожденная им и начальным условием (135) *траектория* (решение $x(\cdot)$ системы (134)) доставляет минимум функционалу (137), т. е.

$$I(u) = \varphi\ (x(t_1)) \to \min. \tag{138}$$

Сформулированная задача является примером простейшей задачи оптимального управления и называется *задачей терминального управления* (ЗТУ) *со свободным правым концом*. Допустимое управление $u^0(t) \in U,\ t \in T$, и соответствующая ему в силу уравнения (134) траектория $x^0(t),\ t \in T$, доставляющие решение ЗТУ, называются *оптимальным управлением* и *оптимальной траекторией* соответственно.

Наряду с терминальным критерием (137) в ТОУ рассматриваются и другие критерии [5, 6]:

– *интегральный критерий*

$$I(u)=\int_{t_0}^{t_1} f_0(x,\ u,\ t)dt\,, \tag{139}$$

– *общий критерий*

$$I(u) = \varphi\big(x(t_1)\big)+\int_{t_0}^{t_1} f_0(x,\ u,\ t)dt\,, \tag{140}$$

– *критерий быстродействия*

$$I(u) = t_1 - t_0. \tag{141}$$

Можно показать, что критерии (137), (139), (140) эквивалентны в том смысле, что ЗТУ может быть сведена к задаче (134)–(136), (139), а также к задаче (134)–(136), (140) и наоборот. Поэтому достаточно исследовать одну из них, например, ЗТУ. Задача оптимального быстродействия, т. е. задача (134)–(136), (141), требует специального рассмотрения.

- **Задача 27.** Докажите эквивалентность критериев (137), (139), (140).

Исследуем ЗТУ в предположении, что функции $f(x, u, t)$, $\dfrac{\partial f(x,u,t)}{\partial x}$, $\dfrac{\partial \varphi(x(t_1))}{\partial x}$ непрерывны. Введем в рассмотрение векторную функцию ψ, $\psi(t) \in \mathbf{R}^n$, и построенный по правой части системы управления (134) *гамильтониан системы*

$$H(x(t), \psi\,(t), u(t), t) = \psi'(t)f(x, u, t), \tag{142}$$

где вектор-функция $\psi(\cdot)$ удовлетворяет дифференциальному уравнению

$$\dot{\psi}(t) = -\frac{\partial H(x(t),\psi(t),u(t),t)}{\partial x} \tag{143}$$

с граничным условием

$$\psi(t_1) = -\frac{\partial \varphi(x(t_1))}{\partial x}. \tag{144}$$

Система (143) дифференциальных уравнений называется системой, *сопряженной* к системе (134).

Сформулируем необходимое условие оптимальности для ЗТУ.

Теорема 20 (*принцип максимума Понтрягина для ЗТУ*). *Пусть* $u^0(t)$, $t \in T$, *– оптимальное управление в задаче (134)–(136), (138),* $x^0(t)$, $t \in T$, *– соответствующая ему оптимальная траектория (основной) системы (134);* $\psi^0(t)$, $t \in T$, *– решение сопряженной системы*

$$\dot{\psi}^0(t) = -\frac{\partial H(x^0(t), \psi^0(t), u^0(t), t)}{\partial x} \qquad (145)$$

с граничным условием $\psi^0(t_1) = -\dfrac{\partial \varphi(x^0(t_1))}{\partial x}$. *Тогда вдоль оптимальных решений* $x^0(\cdot), \psi^0(\cdot)$ *основной и сопряженной систем на оптимальном управлении* $u^0(\cdot)$ *выполняется* ***условие максимума***

$$H(x^0(t), \psi^0(t), u^0(t), t) = \max_{v \in U} H(x^0(t), \psi^0(t), v, t), \quad t \in T. \qquad (146)$$

Замечание 20. Теорема 20 дает *необходимое* условие оптимальности управления $u^0(t)$, $t \in T$, для задачи (134)–(136), (138). Отметим также, что принцип максимума как необходимое условие оптимальности может быть применен в задачах вариационного исчисления для получения необходимых условий сильного экстремума.

Рассмотрим *линейную систему* управления

$$\dot{x}(t) = A(t)x + B(t)u, \; x(t_0) = x_0, \qquad (147)$$

где $A(\cdot)$, $B(\cdot)$ – матрицы размеров $(n \times n)$ и $(n \times r)$ соответственно $(A(t) \in \mathbf{R}^{n \times n}, B(t) \in \mathbf{R}^{n \times r})$, элементы которых суть непрерывные функции, $u(t) \in U \subseteq \mathbf{R}^r$, с *общим критерием качества переходного процесса*

$$I(u) = \varphi\big(x(t_1)\big) + \int_{t_0}^{t_1} f_0(x, u, t)\,dt \to \min. \qquad (148)$$

Здесь $f_0(x, u, t) = f_1(x, t) + b(u, t)$, где $\varphi(x)$, $f_1(x, t)$ – выпуклые по x функции. Тогда для такой задачи принцип максимума (теорема 20) является не только *необходимым*, но и *достаточным условием* оптимальности. А именно, справедлива следующая теорема.

Теорема 21. *Для оптимальности допустимого управления* $u^0(t)$, $t \in T$, *необходимо и достаточно, чтобы вдоль этого управления и соответствующих ему траекторий* $x^0(t), \psi^0(t), t \in T$, *основной и сопряженной систем выполнялось условие максимума (146), где*

$$H(x,\psi,u,t)=\psi(t)\big(A(t)x+B(t)u\big)-f_0\big(x,u,t\big),\ t\in T,$$

$\psi(t)$ *удовлетворяет дифференциальному уравнению (143) с граничным условием (144).*

Пример 26. На отрезке времени $t\in T=\big[0;2\big]$ рассмотрим задачу терминального управления системой

$$\begin{cases}\dot{x}_1=x_2,\\ \dot{x}_2=u\end{cases}\tag{149}$$

с начальными условиями $x_1 = 0$, $x_2 = 0$ и терминальным критерием качества

$$I(u) = x_1(2) - x_2(2) \to \min,\tag{150}$$

где $u(\cdot)$ – ограниченное управление: $|u(t)|\le 1$, $t\in T=[0;2]$.

Решение. Сведем задачу (149), (150) к (147), (148) путем введения векторов $x(t)=\begin{pmatrix}x_1(t)\\ x_2(t)\end{pmatrix}$, $\psi(t)=\begin{pmatrix}\psi_1(t)\\ \psi_2(t)\end{pmatrix}$. Тогда $\dot{x}(t)=\begin{pmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{pmatrix}$ и система управления (149) запишется в виде

$$\begin{pmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{pmatrix}=\underbrace{\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}}_{A}\begin{pmatrix}x_1(t)\\ x_2(t)\end{pmatrix}+\underbrace{\begin{pmatrix}0\\ 1\end{pmatrix}}_{B}u(t),\tag{151}$$

где $A=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}$, $B=\begin{pmatrix}0\\ 1\end{pmatrix}$, $x_0=\begin{pmatrix}x_1(0)\\ x_2(0)\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$, $t_0 = 0$, $t_1 = 2$, $U = \{u{:}|u(t)| \le 1\}$, $\varphi(x(\cdot)) = x_1(\cdot) - x_2(\cdot)$, $f_0(x, u, t) \equiv 0$, $t\in T$. Функция $\varphi(x)$ выпукла по x (линейна по x). Тогда система управления (149) может быть переписана как $\dot{x}(t)=Ax(t)+Bu(t)$. Все условия теоремы 20 для нее выполняются. Поэтому для рассматриваемой задачи принцип максимума является необходимым и достаточным условием оптимальности. Определим оптимальное управление $u^0(t)$, $t\in T$, из условия максимума (146). Рассмотрим гамильтониан $H(x, \psi, u) = \psi'(t)(Ax(t) + Bu(t))$ системы (149), который не зависит явно от времени t, так как основная система (149) *стационарная* (с постоянными коэффициентами). Определим решение $\psi(t)$, $t\in T$, сопряженной системы согласно формулам (143), (144):

$$\dot{\psi}(t)\equiv\begin{pmatrix}\dot{\psi}_1(t)\\ \dot{\psi}_2(t)\end{pmatrix}=-\frac{\partial H(x,\psi,u)}{\partial x}=-A'\psi(t)=$$

$$=-\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}\begin{pmatrix}\psi_1(t)\\ \psi_2(t)\end{pmatrix}=\begin{pmatrix}0\\ -\psi_1(t)\end{pmatrix},\tag{152}$$

$$\psi(2) \equiv \begin{pmatrix} \psi_1(2) \\ \psi_2(2) \end{pmatrix} = -\frac{\partial \varphi(x(2))}{\partial x} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \qquad (153)$$

Заметим, что $\frac{\partial \varphi}{\partial x} \equiv \begin{pmatrix} \frac{\partial \varphi}{\partial x_1} \\ \frac{\partial \varphi}{\partial x_2} \end{pmatrix} = -\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, и так как $H(x, \psi, u) = \psi'(t)(Ax(t) +$ $+ Bu(t)) = x'A'\psi(t) + \psi'(t)Bu(t)$, то $\frac{\partial H(x,\psi,u)}{\partial x} = A'\psi(t)$. Следовательно, из (152), (153) получаем

$$\begin{aligned} &\dot{\psi}_1(t) = 0, \ \psi_1(2) = -1, \\ &\dot{\psi}_2(t) = -\psi_1(t), \ \psi_2(2) = 1. \end{aligned} \qquad (154)$$

Решая сопряженную систему (154), из ее первого уравнения находим $\psi_1(t) = C_1$; тогда $\psi_1(2) = C_1 = -1$. Следовательно, $\psi_1^0(t) = -1$. Из второго уравнения системы (154) следует $\dot{\psi}_2(t) = 1$, откуда $\psi_2(t) = t + C_2$, что в силу граничного условия (153) для $\psi_2(t)$ дает: $\psi_2(2) = 2 + C_2 = 1$, т. е. $C_2 = -1$. Таким образом, $\psi_2^0(t) = t - 1,\ t \in T$. Из условия максимума (146) для задачи (149) будем иметь

$$\psi^{0\prime}(t)(Ax^0(t) + bu^0(t)) = \max_{|u(t)| \le 1} \psi^{0\prime}(t)(Ax^0(t) + bu), \quad t \in T = [0;\ 2],$$

или в силу линейности системы с использованием ее представления в виде (151)

$$\psi^{0\prime}(t)bu^0(t) = \max_{|u(t)| \le 1} \psi^{0\prime}(t)bu, \quad t \in T = [0;\ 2]. \qquad (155)$$

Из (155) видно, что если $\psi^{0\prime}(t)b > 0,\ t \in T$, то max $\psi 0'(t)bu$ при $|u| \le 1$ достигается при $u = 1$; если же $\psi^{0\prime}(t)b < 0,\ t \in T$, то max $\psi^{0\prime}(t)bu$, $|u| \le 1$, достигается при $u = -1$. Иными словами, ограниченное управление $|u| \le 1$, доставляющее максимум правой части (155), имеет вид $u^0(t) = \mathrm{sign}\psi^{0\prime}(t)b,\ t \in T$, где функция sign$t$ определяется по формуле

$$\mathrm{sign}\, t = \begin{cases} 1, & t \ge 0, \\ -1, & t < 0. \end{cases}$$

Подставляя в полученное оптимальное управление $u^0(t)$ найденное значение сопряженного вектора $\psi^{0\prime}(t) = (\psi_1^0(t), \psi_2^0(t)) = (-1;\ t-1)$, видно, что

$$u^0(t) = \mathrm{sign}(-1;\ t-1)\begin{pmatrix}0\\1\end{pmatrix} = \mathrm{sign}(t-1), \quad t \in T = [0;\ 2].$$

Значит, допустимое управление $u^0(t) = \mathrm{sign}(t-1), \quad t \in T = [0;\ 2]$, решающее задачу (149), (150), т. е. *оптимальное управление*, является кусочно-постоянным вида

$$u^0(t) = \begin{cases} -1, & t \in [0; 1), \\ 1, & t \in (1; 2]. \end{cases} \tag{156}$$

Момент времени $t = 1$, в котором оптимальное управление $u^0(t),\ t \in [0; 2]$, меняет свое значение с –1 на 1, называется *точкой переключения оптимального управления* (рис. 22).

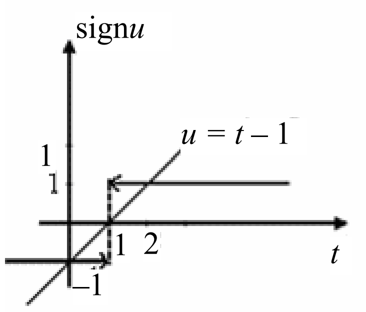


Рис. 22. Вид оптимального управления в примере 26

Для вычисления значения критерия $I(u^0)$ найдем оптимальную траекторию $x^0(t),\ t \in T$. Сначала в исходную систему управления (149) подставим управление $u^0(t) \equiv -1$, $t \in [0;\ 1)$, из (156). Получим систему

$$\begin{cases} \dot{x}_1 = x_2, & x_1(0) = 0, \\ \dot{x}_2 = -1, & x_2(0) = 0. \end{cases}$$

Решением $x_2(t)$ второго уравнения является

$$x_2(t) = -t + c_1 \ \Rightarrow\ x_2(0) = c_1 = 0 \ \Rightarrow x_2(t) = -t,$$

следовательно, из первого уравнения системы получаем

$$\dot{x}_1(t) = -t \ \Rightarrow x_1(t) = -\frac{t^2}{2} + c_2 \Rightarrow\ x_1(0) = c_2 = 0 \ \Rightarrow x_1(t) = -\frac{t^2}{2}.$$

Тогда при $t = 1$ имеем $x_1^0(1) = -\dfrac{1}{2}, \quad x_2^0(1) = -1.$

Для управления $u^0(t) \equiv 1,\ t \in (1; 2]$ из (154) получаем систему

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = 1, \end{cases} \tag{157}$$

для которой начальным условием при $t = 1$ в силу найденных на [0; 1] решений системы (149) является: $x_1(1) = -\dfrac{1}{2}$, $x_2(1) = -1$. Тогда находим решение $x_2(t)$ второго уравнения системы (157):

$$\dot{x}_2(t)=1 \Rightarrow x_2(t)=t+c_1,\ x_2(1)=1+c_1=-1 \Rightarrow c_1=-2 \ \Rightarrow\ x_2(t)=t-2.$$

Подставив найденное $x_2(t) = t - 2$ в первое уравнение системы (157), получаем

$$\dot{x}_1 = x_2 \Rightarrow \dot{x}_1(t) = t-2 \Rightarrow x_1(t) = \frac{t^2}{2} - 2t + c_2 \Rightarrow$$

$$\Rightarrow x_1(1) = \frac{1}{2} - 2 + c_2 = -\frac{1}{2} \Rightarrow c_2 = 1 \Rightarrow x_1(t) = \frac{t^2}{2} - 2t + 1.$$

Таким образом, решением системы (157) является $x_1(t) = \frac{t^2}{2} - 2t + 1$, $x_2(t) = t - 2$, откуда следует: $x_1^0(2) = -1,\quad x_2^0(2) = 0.$ Минимальное значение критерия качества (150) на оптимальном управлении (156) равно $I(u^0) = x_1^0(2) - x_2^0(2) = -1.$ ▲

18.2. Задача оптимального быстродействия

Рассмотрим задачу оптимального быстродействия

$$\dot{x} = f(x,u,t),\ x(t_0) = x_0,\ x(t_1) = 0,$$

$$u(t) \in U,\ t \ge t_0,\ I(u) = t_1 - t_0 \to \min. \qquad (158)$$

Для задачи (158) сохраняются предположения задачи (134)–(136), (138).

Задача оптимального быстродействия (ЗОБ) – одна из ***основных задач оптимального управления:*** среди допустимых управлений $u(\cdot)$, $u(t) \in U$, найти такое $u^0(t)$, $t \ge t_0$, которое переводит порожденную им траекторию $x^0(t)$, $t > t_0$, системы (158) из точки x_0 в точку x_1 (обычно начало координат) за минимально возможное время $t_1 - t_0$.

Теорема 22 (***принцип максимума Понтрягина в ЗОБ***). *Пусть $u^0(t)$, $t \in [t_0; t_1^0]$, – оптимальное управления в задаче (158), $t_1^0 - t_0$ – время быстродействия, $x^0(t)$, $t \in [t_0; t_1^0]$, – оптимальная траектория основной системы. Тогда вдоль $u^0(t)$, $x^0(t)$, $t \in [t_0; t_1^0]$, существует такое нетривиальное решение $\psi^0(t)$, $t \in [t_0; t_1^0]$, сопряженной системы (143), (144), что на оптимальном управлении $u^0(t)$, $t \in [t_0; t_1^0]$, вдоль решений $x^0(t)$, $t \in [t_0; t_1^0]$, основной и $\psi^0(t)$, $t \in [t_0; t_1^0]$, сопряженной систем выполняются условия:*

1) $H\left(x^0(t),\ \psi^0(t),\ u^0(t),\ t\right) = \max\limits_{u \in U} H\left(x^0(t),\ \psi^0(t), u, t\right), \quad t \in [t_0; t_1^0],$

2) $H\left(x^0(t_1^0), \psi^0(t_1^0), u^0(t_1^0), t_1^0\right) \ge 0.$

Условие 2 Теоремы 22 является *условием оптимальности момента* t_1^0.

• **Задача 28*. Простейшая задача о быстродействии [6].** Пусть имеется тележка, движущаяся прямолинейно без трения по горизонтальным рельсам. Тележка управляется внешней силой, ограниченной по абсолютной величине. Требуется остановить ее в определенном положении в кратчайшее время.

18.3. Управляемость линейных систем

При нахождении оптимального управления, например при решении ЗОБ, необходимо прежде всего выяснить вопрос, существует ли хотя бы одно допустимое управление, переводящее траекторию $x(t)$ исследуемой системы из начального состояния $x_0 \in \mathbf{R}^n$ в конечное состояние $x_1 \in \mathbf{R}^n$ (в частности, в начало координат).

Рассмотрим объект управления, поведение которого на отрезке времени $T = [0; t_1]$ описывается системой дифференциальных уравнений

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{159}$$

с начальными условиями

$$x(0) = x_0, \tag{160}$$

где $x(t)$ – n-вектор состояния объекта, $x(t) \in \mathbf{R}^n$; $u(t)$ – вектор управляющих воздействий; A и B – постоянные матрицы соответствующих размеров, $A \in \mathbf{R}^{n\times n}$, $B \in \mathbf{R}^{n\times r}$. *Допустимыми управлениями* будем считать кусочно-непрерывные r-вектор-функции $u(\cdot)$.

Определение 20. Система (159), (160) называется ***полностью управляемой*** на отрезке $T = [0; t_1]$, если для любых n-векторов x_0, x_1 существует допустимое управление $u(\cdot)$, такое, что соответствующая ему в силу (159), (160) траектория $x(\cdot)$ удовлетворяет условиям $x(0) = x_0$, $x(t_1) = x_1$.

Теорема 22 (*критерий управляемости Калмана*) [5, 6]. *Система управления (159) полностью управляема тогда и только тогда, когда*

$$\operatorname{rank}\{B, AB, \ldots, A^{n-1}B\} = n. \tag{161}$$

Матрица $[B, AB, \ldots, A^{n-1}B]$ размерности $n \times rn$ называется *матрицей управляемости* системы (159).

Пример 27. Исследовать на полную управляемость систему (159), где

$$A = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}. \tag{162}$$

Решение. В данном примере $n = 3$, $r = 1$. Поэтому критерий (161) полной управляемости принимает вид

$$\mathrm{rank}\{B, AB, A^2B\} = 3. \tag{163}$$

Прежде чем вычислить ранг входящих в критерий (163) матриц, найдем вид этих матриц:

$$B = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad AB = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \quad A^2B = A \cdot AB = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Тогда

$$\mathrm{rank}\{B, AB, A^2B\} = \mathrm{rank}\left\{ \begin{matrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{matrix} \right\}.$$

Поскольку существует определитель $\begin{vmatrix} -1 & 2 \\ 1 & 1 \end{vmatrix}$ второго порядка матрицы управляемости B, AB, A^2B, отличный от нуля $\left(\begin{vmatrix} -1 & 2 \\ 1 & 1 \end{vmatrix} = -3 \neq 0 \right)$, а определитель третьего порядка $\begin{vmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{vmatrix} = 0,$ то ранг исследуемой матрицы равен двум. Следовательно,

$$\mathrm{rank}\{B, AB, A^2B\} = 2 \neq 3$$

и исследуемая система (159) с параметрами (162) не является полностью управляемой. ▲

Пример 28. Исследовать на полную управляемость систему (159), где

$$A = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 0 \\ 0 & 1 \\ 1 & -2 \end{pmatrix}. \tag{164}$$

Решение. В данном примере $n = 3$, $r = 2$. Поэтому критерий полной управляемости имеет вид (163). Аналогично примеру 27 вычислим B, $AB = \begin{pmatrix} 2 & -2 \\ 0 & 2 \\ 1 & -1 \end{pmatrix}$, $A^2B = \begin{pmatrix} -1 & 0 \\ 0 & 4 \\ 1 & -2 \end{pmatrix}$ и рассмотрим $\mathrm{rank}\{B, AB, A^2B\} =$

$$= \operatorname{rank}\begin{Bmatrix} -1 & 0 & 2 & -2 & -1 & 0 \\ 0 & 1 & 0 & 2 & 0 & 4 \\ 1 & -2 & 1 & -2 & 1 & -2 \end{Bmatrix}.$$ Поскольку существует определитель третьего порядка $\begin{vmatrix} -1 & 0 & 2 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \end{vmatrix}$, отличный от нуля $\left(\begin{vmatrix} -1 & 0 & 2 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \end{vmatrix} = -3 \neq 0\right)$, то ранг матрицы управляемости системы (159), (164) равен трем. Следовательно, исследуемая система полностью управляема. ▲

• **Задача 29.** Исследуйте на полную управляемость систему (159), где $A = \begin{pmatrix} -0 & -1 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ 0 & -1 \\ -1 & 1 \end{pmatrix}$.

ЛИТЕРАТУРА

1. Айвазян, С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – М.: ЮНИТИ, 1998. – 1022 с.

2. Альсевич, В. В. Оптимизация линейных экономических систем. Статические задачи / В. В. Альсевич, Р. Габасов, В. С. Глушенков. – Минск: БГУ, 2000. – 324 с.

3. Блинова, Е. И. Теория вероятностей / Е. И. Блинова, В. М. Марченко, Н. П. Можей. – Минск: БГТУ, 2005. – 125 с.

4. Блисс, Г. А. Лекции по вариационному исчислению / Г. А. Блисс. – М.: ИЛ, 1950. – 244 с.

5. Габасов, Р. Оптимизация линейных систем / Р. Габасов, Ф. М. Кириллова. – Минск: Изд-во БГУ, 1973. – 356 с.

6. Габасов, Р. Методы оптимизации / Р. Габасов, Ф. М. Кириллова. – Минск: Изд-во БГУ, 1975. – 386 с.

7. Галлеев, Э. М. Оптимизация: теория, примеры, задачи / Э. М. Галлеев, В. М. Тихомиров. – М.: Эдиториал УРСС, 2000. – 320 с.

8. Гельфанд, И. М. Вариационное исчисление / И. М. Гельфанд, С. В. Фомин. – М.: Госиздат. физ.-мат. литературы, 1961. – 354 с.

9. Герасимович, А. И. Математическая статистика / А. И. Герасимович. – Минск: Выш. школа, 1983. – 279 с.

10. Гмурман, В. Е. Теория вероятностей и математическая статистика: Учеб. пособие для вузов / В. Е. Гмурман – М.: Высш. школа, 2004. – 480 с.

11. Гусак, А. А. Высшая математика: в 2 т. / А. А. Гусак. – Минск: БГУ, 1978. – Т. 2. – 342 с.

12. Игнатенко, В. В. Высшая математика. Математические методы и модели в расчетах на ЭВМ: Лаб. практикум / В. В. Игнатенко, О. Н. Пыжкова, Л. Д. Яроцкая. – Минск: БГТУ, 2006. – 126 с.

13. Карманов, В. Г. Математическое программирование / В. Г. Карманов. – М.: Наука, 1975. – 272 с.

14. Краснов, М. П. Вариационное исчисление, задачи и упражнения / М. П. Краснов, Г. И. Макаренко, А. И. Киселев. – М.: Наука, 1973. – 244 с.

15. Кремер, Н. Ш. Теория вероятностей и математическая статистика: учебник для вузов / Н. Ш. Кремер. – М.: ЮНИТИ-ДАНА, 2001. – 543 с.

16. Кузнецов, А. В. Высшая математика. Математическое программирование / А. В. Кузнецов, В. А. Сакович, Н. И. Холод. – Минск: Выш. школа, 2001. – 351 с.

17. Кузнецов, А. В. Руководство к решению задач по математическому программированию / А. В. Кузнецов, Н. И. Холод, Л. С. Костевич. – Минск: Выш. школа, 2001. – 448 с.

18. Методическое пособие по разделу «Математическое программирование» курса «Прикладная математика» / сост. В. М. Марченко, В. И. Янович. – Минск: БТИ, 1987. – 62 с.

19. Микулик, Н. А. Решение технических задач по теории вероятностей и математической статистике: Справ. пособие / Н. А. Микулик, Г. Н. Рейзина. – Минск: Выш. школа, 1991. – 164 с.

20. Никольский, С. М. Курс математического анализа: в 2 т. / С. М. Никольский. – М.: Наука, 1973. – Т. 1. – 431 с.

21. Пантелеев, А. В. Вариационное исчисление в примерах и задачах / А. В. Пантелеев. – М.: Высш. школа, 2006. – 272 с.

22. Пантелеев, А. В. Методы оптимизации в примерах и задачах: учеб. пособие / А. В. Пантелеев, Т. А. Летова. – М.: Высш. школа, 2002. – 544 с.

23. Пискунов, Н. С. Дифференциальное и интегральное исчисление: в 2 т. / Н. С. Пискунов. – М.: Наука, 1970. – Т. 1. – 456 с.

24. Эльсгольц, Л. Э. Вариационное исчисление / Л. Э. Эльсгольц. – М.: КомКнига, 2006. – 205 с.

ОГЛАВЛЕНИЕ