Тема 1. Линейное программирование

1.1 Задача линейного программирования (ЗЛП)

Найти $x_j \geq 0, \quad j = \overline{1,n}$, удовлетворяющие *m*-линейным ограничениям

$$\sum_{j=1}^{n} a_{ij} x_j < \cdot > b_i, \; i = \overline{1,m}$$

и оптимизирующих линейную форму (целевую функцию)

$$Y = \sum_{j=1}^{n} c_j x_j \text{ (max/min), где } a_{ij},\, b_i,\, c_j - \text{ известны.}$$

Различают 3 формы ЗЛП:

Стандартная форма

$$\left.\begin{array}{c} \sum\limits_{j=1}^{n} a_{ij} x_j \leq b_i, \; i = \overline{1,m} \\ \forall x_j \geq 0 \\ Y = \sum\limits_{j=1}^{n} c_j x_j \to \max \end{array}\right| \begin{array}{l} AX \leq B, \; (C, X) \Rightarrow \max \\ A = \left\| a_{ij} \right\|, \; X = \left(x_1, \ldots, x_n \right)^T, \; x_i \geq 0 \\ B = \left(b_1, \ldots, b_m \right)^T, \; C = \left(c_1, \ldots, c_n \right) \end{array}$$

Каноническая форма

$$AX = B, \; \max(C, X), \; x_j \geq 0, \; j = \overline{1, n+m}.$$

Общая форма

$$\begin{array}{l} \sum\limits_{j=1}^{n} a_{ij} x_j \leq b_i, \; i = \overline{1,k}; \\ \sum\limits_{j=1}^{n} a_{ij} x_j = b_i, \; i = \overline{k+1,m}; \end{array} \qquad \max(C, X), \; x_j \geq 0, \; j = \overline{1,n}$$

Важным моментом является овладение методами перехода от одной формы ЗЛП к другой:

1) $\max Y \Rightarrow \min Y$: целевая функция исходной задачи берется со знаком минус.
2) Неравенства типа (≥) заменяются на (≤) путем умножения обоих частей неравенства на (-1).

3) Переход от неравенства (≤) на равенство (=) осуществляется добавлением дополнительной (неотрицательной (свободной)) переменной $x_j,\ j > n$, с коэффициентом $c_j = 0$ в целевой функции.

4) Переход от (≥) $\Rightarrow$ (=) осуществляется по следующей цепочке $(\geq) \Rightarrow (\leq) \Rightarrow (=)$ (пункты 2, 3).

5) Если есть переменная, на которую не наложено условие неотрицательности, то эту переменную заменяют на две неотрицательные: $x_j = x_j' - x_j''$.

6) Если переменная x_j ограничена снизу не нулем, а некоторой константой *d*, то заменой переменных $x_j = d + x_j'$ можно перейти к неотрицательной переменной x_j'.

Итак, стандартная форма превращается в каноническую в матричной виде следующим образом.

$$\begin{cases} \tilde{A} \cdot \tilde{X} = B,\ x_j \geq 0,\ \ j = \overline{1,n} \\ \max\left(\tilde{C}, \tilde{X}\right) \end{cases} \text{, где} \qquad (*)$$

$$\tilde{X} = \left(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m}\right)^T$$

$$\tilde{C} = \left(c_1, \ldots, c_n, \underbrace{0, \ldots, 0}_{m}\right)$$

$$\tilde{A} = \begin{pmatrix} a_{11} & \cdots & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ \vdots & & & \vdots & 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & & & \\ a_{m1} & \cdots & \cdots & a_{mn} & 0 & 0 & \cdots & 1 \end{pmatrix}$$

<u>Векторная форма записи</u>:

$$A_1 x_1 + A_2 x_2 + \ldots + A_n x_n = B,\ \ Y = \tilde{C} \cdot \tilde{X} \Rightarrow \max$$

$A_i = \left(a_{1i}, a_{2i}, \ldots, a_{mi}\right)^T -$ векторы-столбцы матрицы $\tilde{A}$.

<u>Определение 1</u>. Планом или допустимым решением ЗЛП называется вектор *X*, удовлетворяющий условиям (*).

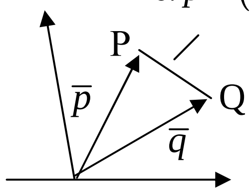
<u>Определение 2</u>. План X называется опорным, если векторы $A_i, i=\overline{1,m}$ являются линейно независимыми.

<u>Определение 3</u>. Опорный план называется невырожденным, если он содержит m положительных компонент.

<u>Определение 4</u>. Оптимальный план – доставляет min (max) целевой функции.

<u>Введем некоторые геометрические понятия</u>

$\alpha \overline{p} + (1-\alpha)\overline{q}$

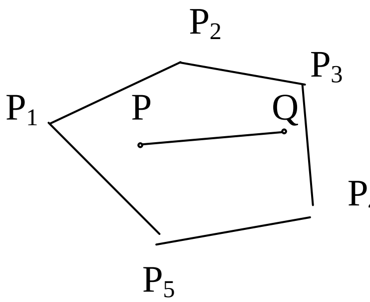


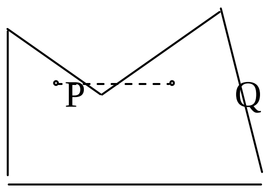
Отрезок PQ – множество точек, определенное соотношением $\alpha\vec{p}+(1-\alpha)\vec{q},\ 0\le\alpha\le 1$

$\vec{p},\vec{q}$ – вектора точек P и Q.

<u>Выпуклое множество S</u>.

$(\forall P \text{ и } \forall Q \in S) \Rightarrow PQ \in S$





(невыпуклое множество)

Выпуклая оболочка точек $P_1, \ldots, P_k$, представленных векторами $\vec{P}_1,\ldots,\vec{P}_k$, называется множество точек вида

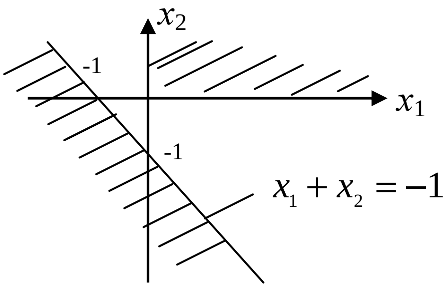
$$\alpha_1\vec{p}_1+\ldots+\alpha_k\vec{p}_k,\quad \alpha_i\ge 0,\quad i=\overline{1,k}$$

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

Выпуклым многоугольником называется замкнутое ограниченное множество на плоскости, имеющее конечное число угловых точек и является выпуклой линейной комбинацией своих угловых точек.

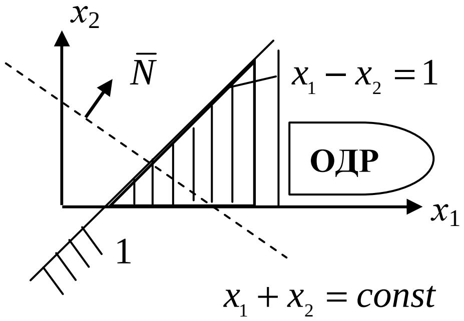
<u>1.2 Геометрическая интерпретация ЗЛП</u>

<u>Пример 1</u>. $\max(x_1+x_2),\quad x_1\ge 0,\quad x_2\ge 0,\quad x_1+x_2\le -1$



Задача не имеет решения, т.к. множество планов пусто.

Пример 2. $\max(x_1 + x_2)$, $x_1 \geq 0$, $x_2 \geq 0$, $x_1 - x_2 \leq 1$



Множество планов не пусто, но не ограничено.

Для того чтобы определить экстремум функции, удобно изобразить ее линии уровня, т.е. линии, на которых значения функции постоянны.

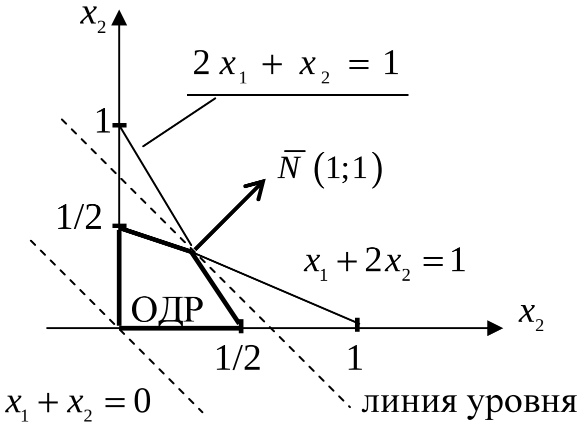
Направление наискорейшего роста или убывания функции определяется перпендикуляром к линии уровня.

В примере 2: $x_1 + x_2 = const$ – линия уровня

$\bar{N}(1;1)$ – нормальный вектор к линии уровня.

Если бы искали $\min(x_1 + x_2)$, то задача имела бы решение $x_1^\circ = x_2^\circ = 0$.

Пример 3. $\max(x_1 + x_2)$, $x_1 \geq 0$, $x_2 \geq 0$, $\begin{cases} x_1 + 2x_2 \leq 1 \\ 2x_1 + x_2 \leq 1 \end{cases}$



Множество ОДР замкнуто и ограничено (является компактом)

Целевая функция непрерывна, то по теореме Вейерштрасса задача имеет решение.

Линия наивысшего уровня $x_1 + x_2 = \frac{2}{3}$ и проходит через вершину (1/3; 1/3).

PS. В многомерном случае ОДР ограничена гиперплоскостями, соответствующими линейным ограничениям.

Пример 4. Фирма выпускает два изделия А и В, в неделю модно использовать 160 часов станочного времени.

запас сырья (1700 ед.) на неделю $\begin{cases} \text{сырье 3 ед./изд. А} \to \text{12 мин.} \to 2\$ \\ \text{сырье 4 ед./изд. В} \to \text{30 мин.} \to 3\$ \end{cases}$ Y

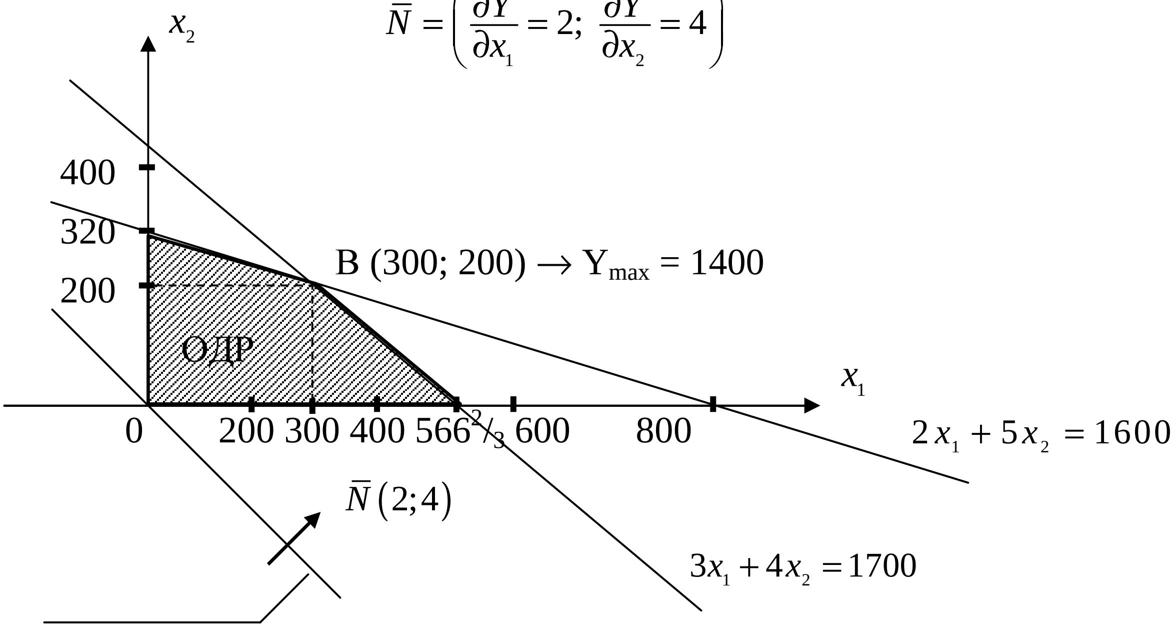
Пусть $x_1 -$ число изделий А за неделю

$x_2 -$ число изделий В за неделю

Тогда $\begin{cases} 3x_1 + 4x_2 \le 1700 \\ 12/60 x_1 + 30/60 x_2 \le 160 \end{cases} \Bigg| \to \begin{cases} 3x_1 + 4x_2 \le 1700 \\ 2x_1 + 5x_2 \le 1600 \end{cases}$

$$\max Y = 2x_1 + 4x_2$$

$$\bar{N} = \left(\frac{\partial Y}{\partial x_1} = 2;\ \frac{\partial Y}{\partial x_2} = 4 \right)$$



$Y_0 = 2x_1 + 4x_2 = 0 -$ линия уровня со значением = 0;

Запишем этот пример в каноническом виде:

$$\left. \begin{aligned} 3x_1 + 4x_2 + x_3 \quad &= 1700 \\ 2x_1 + 5x_2 \quad + x_4 &= 1600 \end{aligned} \right| \quad x_i \ge 0$$

$$Y = 2x_1 + 4x_2 + 0 \cdot x_3 + 0 \cdot x_4 \Rightarrow \max$$

В матричном виде:

$$\begin{pmatrix} 3 & 4 & 1 & 0 \\ 2 & 5 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1700 \\ 1600 \end{pmatrix}$$

Имеем 2 уравнения с 4 неизвестными. Любое неотрицательное решение является допустимым. Если 2 неизвестных задать произвольно, также получаем допустимое решение.

Если 2 неизвестных равны нулю, то получаем базисное решение. Если оно допустимо, то это решение называется **базисное допустимое решение**.

Переменные не равные нулю называют **небазисными**. При $m > n$, $(n-m)-$ переменных C_n^{n-m} способами приравниваем нулю. Базисные решения соответствуют вершинам ОДР.

x_1	x_2	x_3	x_4	
0	0	1700	1600	т. O, Y=0
0	320	420	0	т. A, Y=1280
300	200	0	0	т. B, Y=1400 - !
566⅔	0	0	466⅔	т. C, Y=1133
0	425	0	–525	не допустимые
800	0	–700	0	базисы

1.3 Основные утверждения ЗЛП.

1. Матрица A имеет ранг $= m(m < n)$.
2. Если ограничения имеют допустимое решение, то они имеют и базисное решение.
3. ОДР является выпуклым множеством.
4. Оптимальное решение находится только на границе ОДР.
5. Если решение единственное, то ого достигается в одной из вершин многогранника ОДР и называется **базисным допустимым**.
6. Если решений множество, то они достигаются на одной из сторон ОДР.
7. Если Y имеет конечный экстремум, то, по крайней мере, одно оптимальное решение является базисным.

8. Чтобы найти оптимальное решение, достаточно перебрать все вершины ОДР и выбрать ту из них, где целевая функция достигается $\max(\min)$.

Переход к смежной угловой точке происходит последовательно. Выбор точки зависит от коэффициентов целевой функции.

Так, если ищется $\max Y$, то требуется направление перехода, которое соответствует увеличению x_i, у которого коэффициент c_i в Y больше.

1.4 Практикум.

Задача 1.

Составить математическую модель линейной производственной задачи, где А – матрица удельных затрат, В – вектор объемов ресурсов и С – вектор удельной прибыли при возможном выпуске двух видов продукции с использованием трех видов сырья:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{13} \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \quad C = \begin{pmatrix} c_1 & c_2 \end{pmatrix}$$

или компактно записываются в виде:

$$\begin{array}{ccc} c_1 & c_2 & \\ \hline \multicolumn{1}{|c}{a_{11}} & a_{12} & \multicolumn{1}{c|}{b_1} \\ \multicolumn{1}{|c}{a_{21}} & a_{22} & \multicolumn{1}{c|}{b_2} \\ \multicolumn{1}{|c}{a_{31}} & a_{32} & \multicolumn{1}{c|}{b_3} \\ \hline \end{array}$$

<u>Задание 1.</u> Решить задачу графическим методом. Варианты даны в таблице.

Номера задач	Вариант			Номера задач	Вариант			Номера задач	Вариант		
	1	2			1	2			2	2	
1	1	1	8	11	2	1	8	21	1	3	7
	1	4	20		1	3	22		5	4	24
	1	0	5		1	0	5		1	0	5

2	3	2		12	1	2		22	1	2	
	1	1	7		1	1	8		1	2	9
	4	7	28		5	4	22		3	4	21
	1	0	5		2	0	5		1	0	5
3	3	2		13	3	2		23	3	2	
	1	1	8		1	1	7		2	3	7
	1	4	21		4	7	27		4	7	26
	8	0	5		1	0	6		1	0	5
4	3	2		14	3	2		24	3	2	
	2	1	5		1	1	8		3	3	8
	4	7	28		1	4	21		1	4	27
	1	0	4		5	0	5		7	0	5
5	1	2		15	3	2		25	3	2	
	2	3	8		1	1	6		2	1	4
	1	3	21		4	7	30		4	7	27
	1	0	4		1	0	4		1	0	4
6	1	2		16	1	2		26	1	2	
	1	1	9		2	1	8		2	1	8
	4	6	22		1	3	22		1	3	22
	2	0	5		1	0	5		1	0	5
7	3	2		17	1	2		27	1	2	
	1	1	7		1	1	8		1	1	8
	4	8	29		5	4	32		5	4	22
	1	0	6		1	0	5		2	0	5
8	3	2		18	3	2		28	3	2	
	2	1	8		1	1	7		1	1	7
	1	4	21		4	8	25		4	7	29
	5	0	7		1	0	7		1	0	6
9	3	2		19	3	2		29	3	2	
	3	1	5		1	1	8		1	2	8
	4	7	27		1	4	21		1	4	25
	1	0	7		5	0	5		5	0	5
10	3	2		20	3	2		30	3	2	
	1	1	7		2	3	5		3	1	5
	4	5	27		3	7	27		4	7	23
	2	0	3		1	0	4		1	0	3

Пример решения

На изготовление двух видов продукции требуется три вида сырья, запасы которого и количество, необходимое для изготовления единиц продукции каждого даны в таблице

Продукция		Запасы сырья
P_1	P_2	
1	5	35

2	1	16
1	0	6

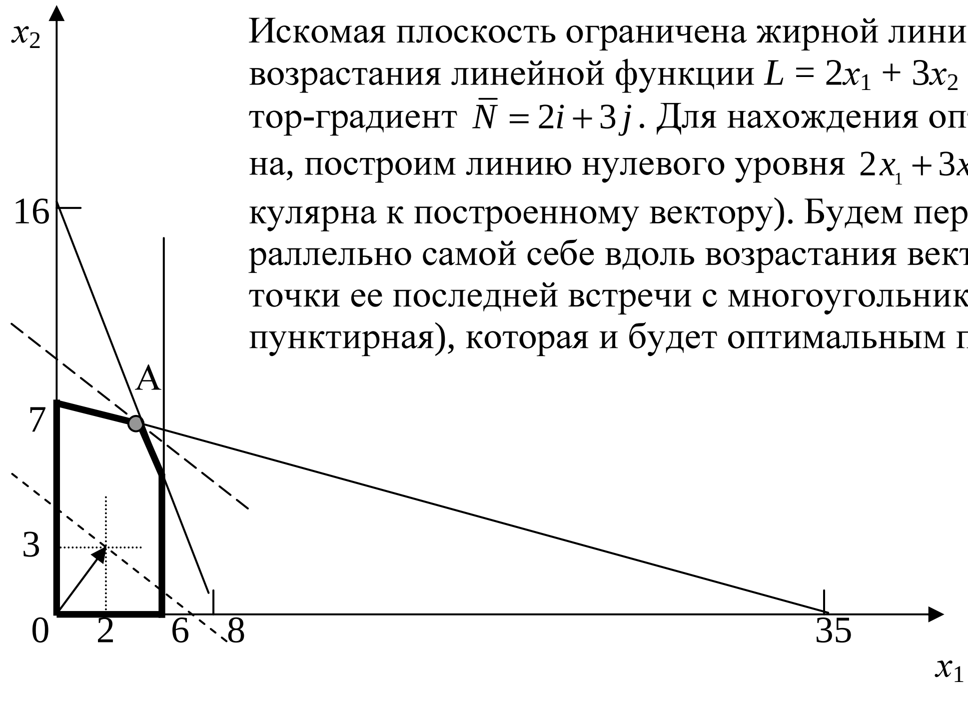
Прибыль от реализации первой продукции равна 2 условным денежным единицам, второй – 3.

Требуется составить план производства, при котором прибыль от реализации всей продукции была бы максимальной.

Обозначим через x_1 и x_2 количество единиц продукции первого и второго вида. Прибыль при таком плане выпуска продукции будет $L = 2x_1 + 3x_2$. Ограничения по запасам сырья выразятся в виде системы неравенств:

$$\begin{cases} x_1 + 5x_2 \le 35 \\ 2x_1 + x_2 \le 16 \\ x_1 \le 6 \\ x_1 \ge 0,\ x_2 \ge 0 \end{cases}$$

Найдем графически часть плоскости, координаты точек которой удовлетворяют полученной системы неравенств.



Искомая плоскость ограничена жирной линией. Направление возрастания линейной функции $L = 2x_1 + 3x_2$ показывает вектор-градиент $\overline{N} = 2i + 3j$. Для нахождения оптимального плана, построим линию нулевого уровня $2x_1 + 3x_2 = 0$ (перпендикулярна к построенному вектору). Будем передвигать ее параллельно самой себе вдоль возрастания вектора до точки **A**, точки ее последней встречи с многоугольником (жирная пунктирная), которая и будет оптимальным планом.

Точка **A** имеет координаты удовлетворяющих системе уравнений:

Решение этой системы: $\begin{cases} x_1 = 5 \\ x_2 = 6 \end{cases}$, тогда $L_{\max} = 2 \cdot 5 + 3 \cdot 6 = 28$

<u>1.5 Симплекс – метод</u>

Пример 1. Для изготовления изделий B_1, B_2, B_3, B_4 потребуется оборудование A_1, A_2, A_3.

	B_1	B_2	B_3	B_4	b_i
A_1	3	5	2	7	15
A_2	4	3	3	5	9
A_3	5	6	4	8	30
c_j	40	50	30	20	

a_{ij} – время изготовления j изделия на i оборудовании;

b_i – ресурсы оборудования;

c_j – прибыль от j изделия;

x_j – количество изделий j, которое можно выпустить

Критерий – суммарная величина прибыли

$$Y = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 \Rightarrow \max.$$

Ограничения имеют вид:

$$\begin{cases} 3x_1 + 5x_2 + 2x_3 + 7x_4 \le 15 \\ 4x_1 + 3x_2 + 3x_3 + 5x_4 \le 9 \\ 5x_1 + 6x_2 + 4x_3 + 8x_4 \le 30 \end{cases} \Bigg| \, x_j \ge 0, \qquad j = \overline{1,4}$$

Перейдем к канонической форме, введя фиктивные изделия x_5, x_6, x_7, <u>характеризующие остатки ресурсов</u>.

Тогда система ограничений в виде равенства будет иметь вид

$$\begin{cases} 3x_1 + 5x_2 + 2x_3 + 7x_4 + 1 \cdot x_5 + 0 \cdot x_6 + 0 \cdot x_7 = 15 \\ 4x_1 + 3x_2 + 3x_3 + 5x_4 + 0 \cdot x_5 + 1 \cdot x_6 + 0 \cdot x_7 = 9 \\ 5x_1 + 6x_2 + 4x_3 + 8x_4 + 0 \cdot x_5 + 0 \cdot x_6 + 1 \cdot x_7 = 30 \end{cases}$$

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

$$Y = 40x_1 + 50x_2 + 30x_3 + 20x_4 + 0 \cdot x_5 + 0 \cdot x_6 + 0 \cdot x_7 \to \max$$

Переменных стало n = 7, ограничений m = 3, отсюда (n – m) = 4 неизвестных должны равняться нулю.

I этап <u>Выбор опорного решения (базиса)</u>

В качестве базисных переменных возьмем x_5, x_6, x_7.

Остальные переменные (небазисные) приравняем к нулю.

Опорное решение: X = {0,0,0,0,15,9,30}.

Строим первую симплекс – таблицу:

табл. 1

базис	свобод. члены	коэффициенты при x_i						
		x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_5	15	3	$\boxed{5}$	2	7	1	0	0
x_6	9	4	3	3	5	0	1	0
x_7	30	5	6	4	8	0	0	1
Y	0	40	50	30	20	0	0	0

II этап Последовательное улучшение решения.

1) Проверка на оптимальность

$$\left(\begin{array}{c}\forall c_i<0 \text{ при небазисных переменных}\\ \forall c_i>0 \text{ при базисных переменных}\end{array}\right)\Rightarrow\left(\begin{array}{c}\text{Небазисные переменные обращаем в}\\ \text{нуль, а значения базисных найдем из}\\ \text{системы ограничений}\end{array}\right)$$

Если решение не оптимальное, переходим к следующему этапу.

2) Проверка на наличие решения

Если столбец коэффициентов при небазисной переменной в системе ограничений состоит только из отрицательных чисел, то задача решения не имеет.

3) Введение в базис новой переменной, способной увеличить значение Y.

Для этого выбирают переменную с наибольшим коэффициентом в целевой функции. В нашем примере это x_2.

4) Выведение из базиса переменной

Для всех положительных коэффициентов столбца при x_2 в системе ограничений находим отношение свободного члена к соответствующему коэффициенту.

$$\min\frac{15}{5}=3;\quad \frac{9}{3}=3;\quad \frac{30}{6}=5.$$

Минимальное (любое) отношение указывает на выводимую из базиса переменную (пусть это будет x_5). Элемент, стоящий на отмеченных строке и столбце, называют разрешающим (его следует пометить $\boxed{5}$)

5) Выражение вводимой переменной x_2 через выводимую x_5 и небазисные переменные.

Для этого составляется таблица 2 по следующим правилам:

а) строку с разрешающим элементом делим на его значение, б) полученная строка умножается на числа, позволяющие, чтобы после сложения с другими строками первой таблицы в <u>отмеченном столбце</u> появились нули.

- первую строку умножаем на (-3) и складываем со 2-й;
- первую строку умножаем на (-6) и складываем с 3-й;
- первую строку умножаем на (-50) и складываем с 4-й;
- и т.д.

табл. 2

базис	свобод. члены	коэффициенты при x_i						
		x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_2	3	3/5	1	2/5	7/5	1/5	0	0
x_6	0	11/5	0	9/5	4/5	-3/5	1	0
x_7	12	7/5	0	8/5	-2/5	-6/5	0	1
Y	-150	10	0	10	-50	-10	0	0

× (-3)

+

Решение имеется, но неоптимальное.

Включаем переменную x_3, т.к. коэффициент c_3 = 10 наибольший.

Выводим переменную x_6, т.к. отношения свободного члена к коэффициенту при x_3 наименьший (0 / (9/5)) = 0.

$\tilde{B}$

табл. 3

базис	свобод. члены	коэффициенты при x_i						
		x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_2	3	1/9	1	0	11/9	1/3	-2/9	0
x_3	0	11/9	0	1	4/9	-3/9	5/9	0
x_7	12	-5/9	0	0	-10/9	-2/3	-8/9	1
Y	-150	$-2^2/_9$	0	0	$-54^4/_9$	$-6^2/_3$	$-5^5/_9$	0

Q^{-1}

В последней строке нет положительных элементов.

Оптимальное решение: $\boxed{x_2 = 3;\ x_3 = 0;\ x_7 = 12.}$

Небазисные переменные приравниваем нулю:

$$x_1 = x_4 = x_5 = x_6 = 0$$

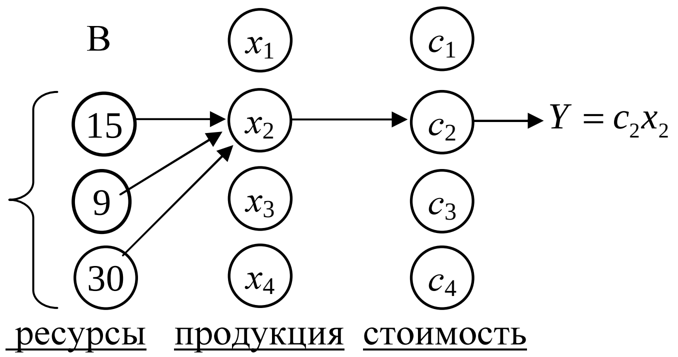
Выгодно изготовлять только изделие B_2, при этом ресурсы оборудования A_1, A_2 полностью использованы:

$$\begin{cases} a_{12} \cdot x_2 = 5 \cdot 3 = 15 = b_1 \\ a_{22} \cdot x_2 = 3 \cdot 3 = 9 = b_2 \end{cases}.$$

Ресурс оборудования A_3 не использован полностью:

$$a_{32} \cdot x_2 = 6 \cdot 3 = 18 \le b_3 = 30.$$

Значение целевой функции $Y = c_2 x_2 = 50 \cdot 3 = 150.$



Ресурсы b_1, b_2 использованы полностью

Итак, третий ресурс является избыточным. Из последней таблицы найдем обращенный базис Q^{-1}, отвечающий оптимальной производственной программе.

$$Q^{-1} = \begin{pmatrix} 1/3 & -2/9 & 0 \\ -3/9 & 5/9 & 0 \\ -2/3 & -8/9 & 1 \end{pmatrix}; \quad \tilde{B} = \begin{pmatrix} 3 \\ 0 \\ 12 \end{pmatrix}; \quad B = \begin{pmatrix} 15 \\ 9 \\ 30 \end{pmatrix};$$

Проверка: $\boxed{\tilde{B} = Q^{-1} \cdot B.}$

$$\begin{pmatrix} 1/3 & -2/9 & 0 \\ -3/9 & 5/9 & 0 \\ -2/3 & -8/9 & 1 \end{pmatrix} \cdot \begin{pmatrix} 15 \\ 9 \\ 30 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 12 \end{pmatrix}. \quad *$$

Значения c_5, c_6, из последней таблицы являются двойственными оценками ресурсов. Например, $c_5 \approx 6$ показывает, что добавление одной единицы первого ресурса обеспечивает прирост прибыли на 6 единиц, а второго на 5 единиц.

Добавлять третий ресурс не надо, т.к. он является избыточным, при этом ресурсы 1, 2 образуют «узкие места» производства.

Надо их заказать дополнительно.

Пусть $T(t_1, t_2, t_3)$ – дополнительные объемы ресурсов.

В нашем примере $T(t_1, t_2, 0)$.

Прирост прибыли будет $W = 6t_1 + 5t_2 \Rightarrow$ max

Используя двойственные оценки ресурсов, мы должны добиться условия:

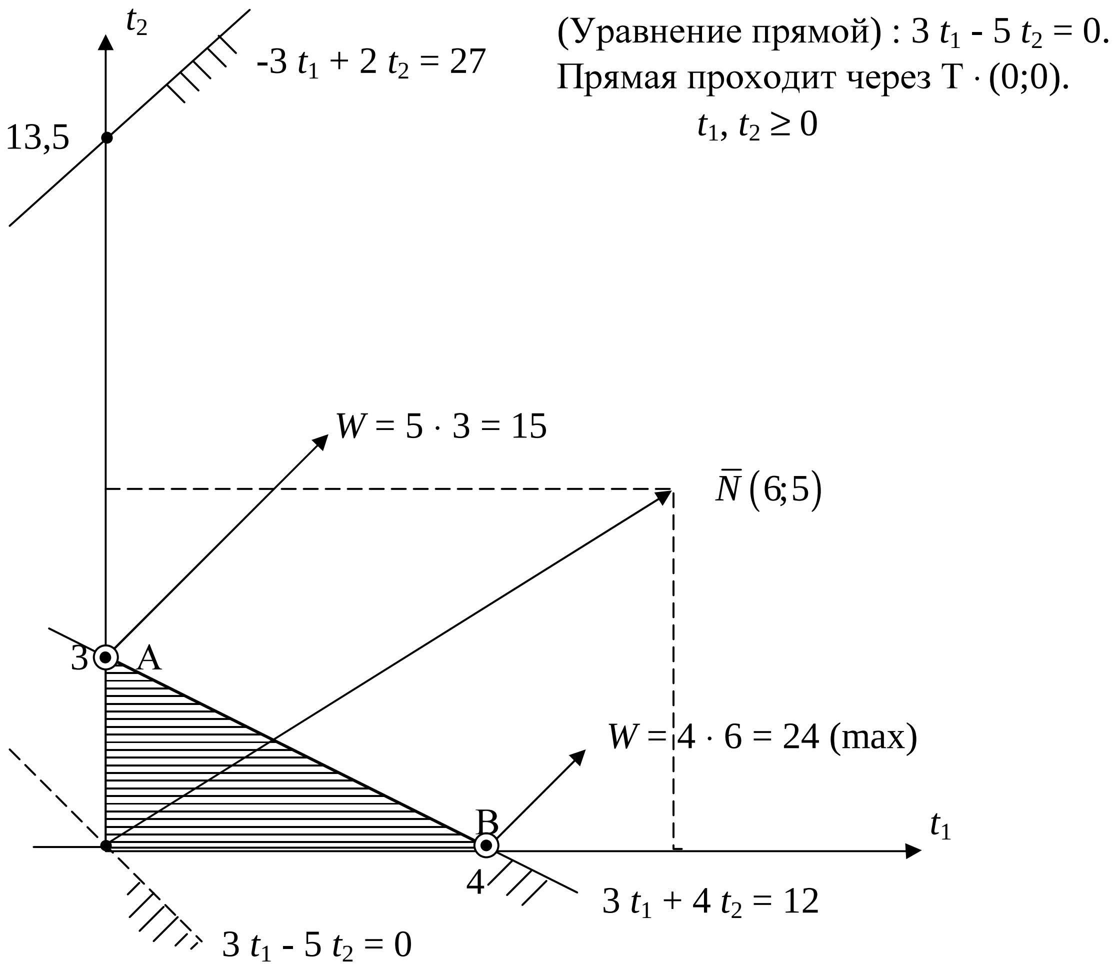
$$\boxed{\tilde{B} + Q^{-1} \cdot T \ge 0}$$

$$\begin{pmatrix} 3 \\ 0 \\ 12 \end{pmatrix} + \begin{pmatrix} 1/3 & -2/9 & 0 \\ -3/9 & 5/9 & 0 \\ -2/3 & -8/9 & 1 \end{pmatrix} \cdot \begin{pmatrix} t_1 \\ t_2 \\ 0 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\left\{\begin{array}{l} 3+\dfrac{t_1}{3}-\dfrac{2}{9}t_2 \geq 0 \\ 0+\left(-\dfrac{3}{9}\right)t_1+\dfrac{5}{9}t_2 \geq 0 \\ 12-\dfrac{2}{3}t_1-\dfrac{8}{9}t_2 \geq 0 \end{array}\right. \Bigg| \Rightarrow \left\{\begin{array}{l} 3t_1-2t_2 \geq -27 \\ -3t_1+5t_2 \geq 0 \\ -6t_1-8t_2 \geq -12 \end{array}\right. \Bigg\| \Rightarrow$$

$$\boxed{\left\{\begin{array}{l} -3t_1+2t_2 \leq 27 \\ 3t_1-5t_2 \leq 0 \\ 3t_1+4t_2 \leq 12 \end{array}\right. \Bigg| W = 6t_1+5t_2 \Rightarrow \max}$$

стандартная форма ЗЛП



$$(t_1 = 4, \quad t_2 = 0) \rightarrow W_{\max} = 24.$$

Если первый ресурс увеличим на 4 единицы, то доход увеличится на 24 единицы.

Пример 2. <u>Линейная производственная задача.</u>

Предприятие выпускает n = 4 вида продукции, используя m = 3 видов ресурсов.

$A = \left\| a_{ij} \right\|$ – матрица удельных затрат;

$B = (b_1, b_2, b_3)^T$ – вектор-столбец объемов ресурсов;

$C = (c_1, c_2, c_3, c_4)$ – вектор удельной прибыли;

a_{ij} – расход *i* ресурса на единицу *j* продукции; $i = \overline{1,3}$;

x_j – искомое количество единиц *j* продукции; $j = \overline{1,4}$;

$$\underbrace{\begin{pmatrix} a_{11} & \cdots & a_{14} \\ \vdots & & \vdots \\ a_{31} & \cdots & a_{34} \end{pmatrix}}_{A} \times \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_4 \end{pmatrix}}_{X} \le \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}}_{B}; \quad AX \le B;$$

Найти программу $(x_1, \dots, x_4)$, максимизирующую прибыль:

$$Y = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 \to \max$$

Исходные данные представляются в следующем виде:

c_1	c_2	c_3	c_4	
a_{11}	…	…	a_{14}	b_1
⋮			⋮	b_2
a_{31}	…	…	34	b_3

Решить симплексным методом, найти оптимальную программу производства, максимальную прибыль, указать остатки ресурсов различных видов и «узкие места» производства.

Сформулировать и составить математическую модель задачи о «расшивке узких мест» производства. При этом от поставщиков можно получить не более трети первоначально выделенного объема ресурсов любого вида. Пусть задание имеет вид:

26	35	18	30	
2	5	1	4	126
3	0	7	2	84
2	1	4	0	75

Система линейных неравенств:

$$\begin{cases} 2x_1 + 5x_2 + x_3 + 4x_4 \le 126; \\ 3x_1 + \ 0 + 7x_3 + 2x_4 \le 84; \\ 2x_1 + \ x_2 + 4x_3 + \ 0 \ \le 75; \end{cases}$$

Введем дополнительные неотрицательные переменные x_5, x_6, x_7, характеризующие остатки ресурсов.

В матричном виде ЗЛП имеет вид:

$$\underbrace{\begin{pmatrix} 2 & 5 & 1 & 4 & 1 & 0 & 0 \\ 3 & 0 & 7 & 2 & 0 & 1 & 0 \\ 2 & 17 & 4 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\tilde{A}} \times \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_7 \end{pmatrix}}_{\tilde{X}} = \underbrace{\begin{pmatrix} 126 \\ 84 \\ 75 \end{pmatrix}}_{B}; \qquad \tilde{A} \cdot \tilde{X} = B$$

$$Y = 26x_1 + 35x_2 + 18x_3 + 30x_4 + 0x_5 + 0x_6 + 0x_7 \to \max$$

I этап. **Нахождение опорного базиса.**

Пусть базис (x_5, x_6, x_7), небазисные переменные приравниваем к нулю $x_1 = x_2 = x_3 = x_4 = 0$.

Первое решение: $X_0 = (0,0,0,0,126,84,75)$;

$$Y_0 = 0; \qquad C = (0,0,0,0,0,0,0);$$

табл. 1

базис	Правые части B_1	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_5	126	2	$\boxed{5}$	1	4	1	0	0
x_6	84	3	0	7	2	0	1	0
x_7	75	2	1	4	0	0	0	1
Y	0	26	35	18	30	0	0	0

Улучшаем Y путем введения в базис новой переменной. Вводим x_2, т.к. продукция второго вида имеет наибольшую прибыль ($C_2 = 35$ при x_2 в Y).

Какую переменную вывести из базиса?

Находим отношение правых частей B_1 к <u>положительным коэффициентам</u> при x_2 и выбираем наименьшее.

$$\min\left(\frac{b_i}{a_{i2} > 0}\right) = \min\left(\frac{126}{5}; \frac{75}{1}; \cancel{\frac{84}{0}}\right) = \frac{126}{5}$$

Это указывает на то, что выводится переменная x_5.

Первое уравнение (строка) – разрешающая, а коэффициент $\boxed{a_{12} = 5}$ – разрешающий.

Проверка на оптимальность: в Y все коэффициенты при небазисных переменных должны быть меньше нуля.

табл. 2

базис	B_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_2	126/5	2/5	1	1/5	4/5	1/5	0	0
x_6	84	$\boxed{3}$	0	7	2	0	1	0
x_7	249/5	8/5	0	19/5	-4/5	-1/5	0	1
Y	-882	12	0	11	2	-7	0	0

Первая строка – результат деления первой строки табл.1 на разрешающий коэффициент $\boxed{5}$.

Во всех позициях столбца при x_2 необходимо получить нули. Для этого эту разрешающую строку умножим на такое число, чтобы после сложения с оставшимися строками первой таблицы в колонке при x_2 был ноль.

В нашем случае вторая строка остается неизменной, т.к. там уже ноль.

Разрешающую строку умножаем на (-1) и складываем с третьей строкой первой таблицы.

Аналогично, разрешающую строку умножаем на (-35) и складываем с последней строкой таблицы 1.

Решение не оптимальное.

Строим таблицу 3.

Наибольший коэффициент в Y при x_1 равен 12. Эту переменную вводим в базис.

Определяем: кого выводить из базиса?

$$\min\left(\frac{126}{5}\cdot\frac{5}{2}; \frac{84}{3}, \frac{249}{5}\cdot\frac{5}{8}\right) = 28$$

Это соответствует второму уравнению, следовательно, выводится переменная x_6.

табл. 3

базис	B$_3$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_2	14	0	1	-11/15	8/15	1/5	-2/5	0
x_1	28	1	0	7/3	2/3	0	1/3	0
x_7	5	0	0	1/15	-28/15	-1/5	-8/15	1
Y	-1218	0	0	-17	-6	-7	-4	0

Q^{-1}

Все коэффициенты при Y отрицательны, следовательно, имеем оптимальное решение.

Производственная программа: $x_1 = 28; \quad x_2 = 14;;$

Прибыль: $Y \max = 1218$.

Если эти значения подставить в исходную систему неравенств:

$$\begin{cases} 2x_1 + 5x_2 = 126 & (126) \\ 3x_1 \qquad = 84 & (84) \\ 2x_1 + x_2 = 70 & (75) \end{cases}.$$

Третий ресурс недоиспользован на 5 единиц, т.е. является избыточным.

Проверка: $\boxed{B_3 = Q^{-1} \cdot B_1}$ (Q^{-1} – обращенный базис)

$$\begin{pmatrix} 1/5 & -2/15 & 0 \\ 0 & 1/3 & 0 \\ -1/5 & -8/15 & 1 \end{pmatrix} \begin{pmatrix} 126 \\ 84 \\ 75 \end{pmatrix} = \begin{pmatrix} 14 \\ 28 \\ 5 \end{pmatrix}.$$

Произведем «расшифровку узких мест производства». Из таблицы 3 строка

-7	-4	0

– двойственные оценки ресурсов (например, 7 показывает, что добавление одной единицы первого ресурса обеспечивает прирост прибыли на 7 единиц, а второго – на 4 единицы.

Добавлять третий ресурс не надо, т.к. он является избыточным.

«Узкие места» образуют ресурсы 1 и 2. Надо их заказать дополнительно.

Пусть $T(t_1, t_2, t_3)$ – дополнительные объемы ресурсов.

В нашем примере: $T(t_1, t_2, 0)$.

Прирост прибыли очевиден: $W = 7t_1 + 4t_2 \Rightarrow \max$.

Используя двойственные оценки ресурсов, мы должны добиться условия:

$$\boxed{B_3 + Q^{-1} \cdot T \geq 0}$$

$$\begin{pmatrix}14\\28\\5\end{pmatrix}+\begin{pmatrix}1/5 & -2/15 & 0\\0 & 1/3 & 0\\-1/5 & -8/15 & 1\end{pmatrix}\cdot\begin{pmatrix}t_1\\t_2\\0\end{pmatrix}\ge 0$$

$$\left\{\begin{aligned}3t_1-2t_2&\ge -210\\ t_2&\ge -\ 84\\ -3t_1-8t_2&\ge -\ 75\end{aligned}\right| \longrightarrow \forall t_2\ge 0$$

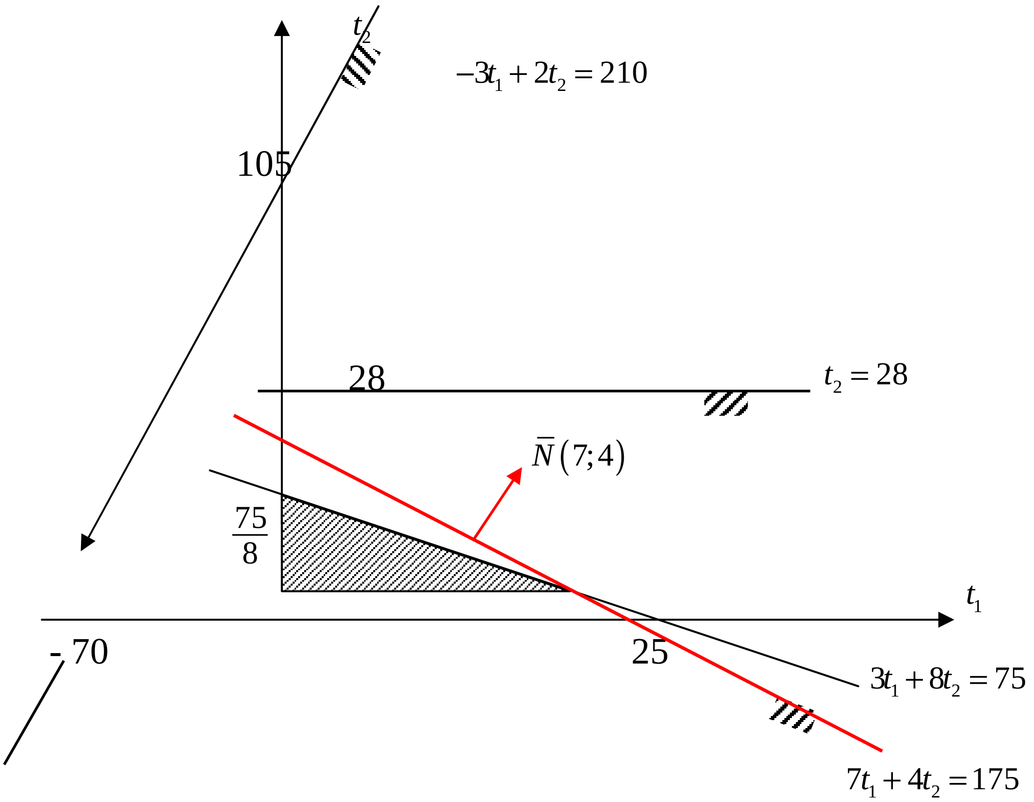
Найдем такое t_2, чтобы удовлетворялось 1-ое и 3-е уравнение:

$$-10t_2\ge -285\Rightarrow t_2\le 28{,}5$$

Вместо значения 28,5 возьмем 28,0

$$\left\{\begin{aligned}-3t_1+2t_2&\le 210\\ 3t_1+8t_2&\le\ 75\\ t_2&\le\ 28\end{aligned}\right| W=7t_1+4t_2\Rightarrow \max$$

Новую задачу ЛП решим геометрически.



Первый ресурс увеличили на 25 единиц. Доход увеличился на 175 единиц.

$$W=3t_1=3\cdot 25=75.$$

Для закрепления методики решения задачи о «расшивки узких мест» рассмотрим следующий пример.

<u>Пример 3</u>.

$$\begin{pmatrix} 4 & 3 & 4 & 5 \\ 2 & 5 & 0 & 2 \\ 3 & 1 & 2 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_4 \end{pmatrix} \le \begin{pmatrix} 208 \\ 107 \\ 181 \end{pmatrix}; \qquad \max\left(36,14,20,50\right)\cdot X$$

Оптимальная программа: $x_1 = 27,\ x_2 = 0,\ x_3 = 0,\ x_4 = 20;$

$$Y_{\max} = 1972.$$

Второй ресурс имеет остаток $x_6 = 13$. Первый и третий ресурсы использованы полностью, т.е. образуют «узкие места производства».

Введем $T\left(t_1, 0, t_3\right) -$ вектор дополнительных объемов ресурсов,

Используя двойственные оценки (6;4) ресурсов, добьемся условия

$$W = 6t_1 + 4t_3 \Rightarrow \max$$

$$\tilde{B} + Q^{-1}\cdot T \ge 0$$

$$\begin{pmatrix} 27 \\ 13 \\ 20 \end{pmatrix} + \begin{pmatrix} 1 & 0 & -1 \\ -4/5 & 1 & 2/5 \\ -3/5 & 0 & 4/5 \end{pmatrix} \begin{pmatrix} t_1 \\ 0 \\ t_3 \end{pmatrix} \ge \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

<u>Предположим, что мы хотим получить дополнительно не более 1/3 первоначального объема ресурса</u>

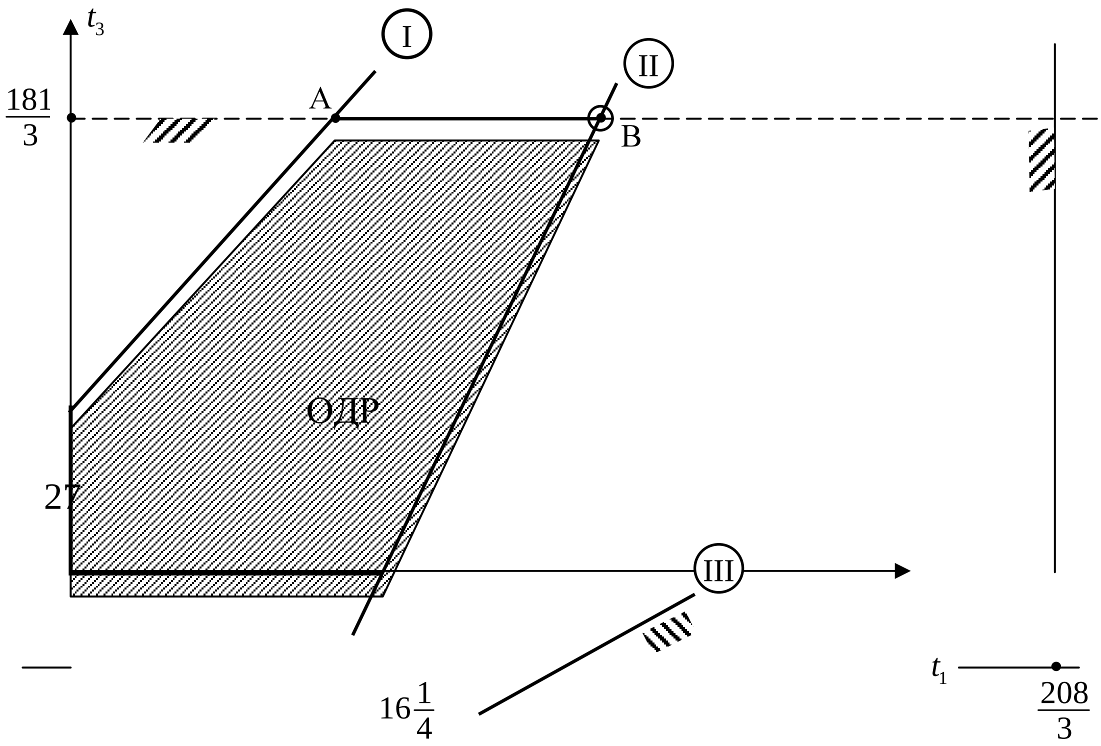
$$\begin{pmatrix} t_1 \\ 0 \\ t_3 \end{pmatrix} \le \frac{1}{3}\begin{pmatrix} 208 \\ 107 \\ 181 \end{pmatrix} \qquad \begin{matrix} t_1 \le \dfrac{208}{3} \\ t_3 \le \dfrac{181}{3} \end{matrix} \qquad\qquad (*)$$

Тогда ЗЛП будет иметь вид:

$$\begin{cases} -t_1 + t_3 \le 27; \\ \dfrac{4}{5}t_1 - \dfrac{2}{5}t_3 \le 13; \\ \dfrac{3}{5}t_1 - \dfrac{4}{3}t_3 \le 20; \end{cases}$$

Решение: $t_1 \approx 46; \quad t_3 \approx 60; \quad W_{\max} \approx 520.$

PS: Ограничения (*) могут позволить «обрезать» ОДР ЗЛП, сделав ее компактом.



$$
\text{т.A} \quad \begin{cases} -t_1 + t_3 = 27 \\ t_3 \approx 60 \end{cases} \Bigg| \to A(33;60) \to W = 6 \cdot 33 + 4 \cdot 60 = 438
$$

$$
\text{т.B} \quad \begin{cases} t_3 \approx 60 \\ \frac{4}{5}t_1 - \frac{2}{5}t_3 = 13 \end{cases} \Bigg| \to t_1 = 46 \to B(46;60)
$$

$$
W = 6 \cdot 46 + 4 \cdot 60 \approx 520 \quad \max
$$

1.6 Анализ моделей линейного программирования после нахождения оптимального решения (на примитивном уровне)

В рамках такого анализа выявляется чувствительность оптимального решения к определенным изменениям исходной модели.

<u>Задача 1</u>. Как отражаются на оптимальном решении изменения запасов ресурсов?

Этот вид анализа идентифицируется как анализ модели <u>на чувствительность к правой части В ограничений</u>. <u>Это анализ стоимости единицы ресурса</u>. При изменении количества допустимых ресурсов на единицу значение целевой функции в оптимальном решении изменится на стоимость единицы ресурса.

<u>Дефицитные ресурсы</u> называют <u>связывающими (активными)</u>, т.к. они используются полностью.

<u>Недефицитные ресурсы</u> называют не <u>связывающими (не активными)</u>, т.к. они имеются в некотором избытке.

<u>Во-первых</u>, – на сколько можно увеличить запас некоторого ресурса для улучшения оптимального значения целевой функции *Y*.

<u>Во-вторых</u> – на сколько можно снизить запас некоторого ресурса при сохранении полученного оптимального значения *Y*.

Таким образом, при анализе ЗЛП на чувствительность к правым частям определяется:

- предельно допустимое увеличение запаса дефицитного ресурса, позволяющее улучшить найденное оптимальное решение;
- предельно допустимое снижение запаса недефицитного ресурса, не изменяющее найденное ранее оптимальное решение.

<u>Замечания:</u>

1) Уменьшение дефицитного ресурса никогда не улучшает значения *Y*.

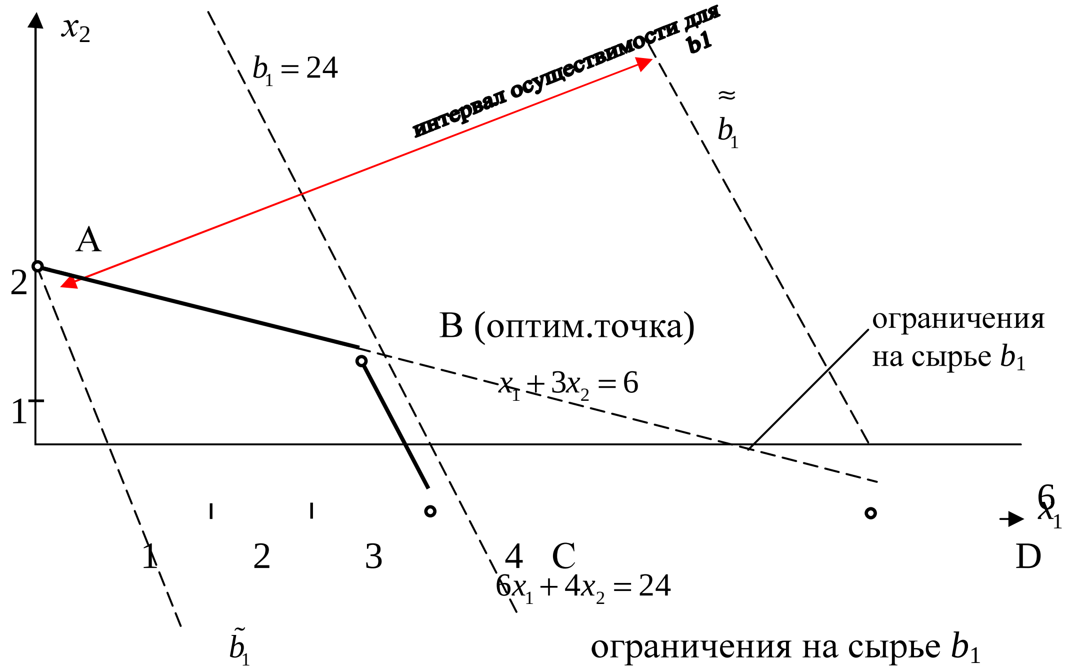
Исключение избыточного (недефицитного) ограничения не влияет ни на ОДР, ни на Y_{max}, тогда как исключение исходного ограничения, соответствующего недефицитному ресурсу, всегда влияет на ОДР, но не всегда на Y_{max}.

<u>Задача 2</u>. Увеличение какого ресурса выгодно? Какому ресурсу следует отдать предпочтение при вложении дополнительных средств. Для этого вводится характеристика ценности y_j каждой дополнительной единицы дефицитного ресурса *j*.

$$y_j = \frac{\max\ Y}{\max\ b_j} = \frac{(\text{максимальное приращение } Y)}{(\text{максимально допустимый прирост ресурса})}$$

Дополнительные вложения в первую очередь направляются на увеличение ресурса *j*, у которого y_j больше. При этом объем недефицитных ресурсов увеличивать не следует.

Пусть имеется графическое решение некоторой задачи:



Концевые точки A и D определяют интервал осуществимости для ресурса b_1. Количество сырья соответствующее точке A равно 0+3·2=6.

Таким образом, интервал осуществимости для $b_1 : \tilde{b}_1 \le b_1 \le \tilde{\tilde{b}}_1$. Вычислим стоимость единицы материала b_1. Изменения b_1 от $\tilde{b}_1$, до $\tilde{\tilde{b}}_1$, значения целевой функции будут соответствовать положению точки B на отрезке AD. Стоимость y_1 единицы ресурса b_1:

$$y_1 = \frac{\text{изменение Y при перемещении точки B от A до D}}{\text{изменение } b_1 \text{ при перемещении точки B от A до D}}$$

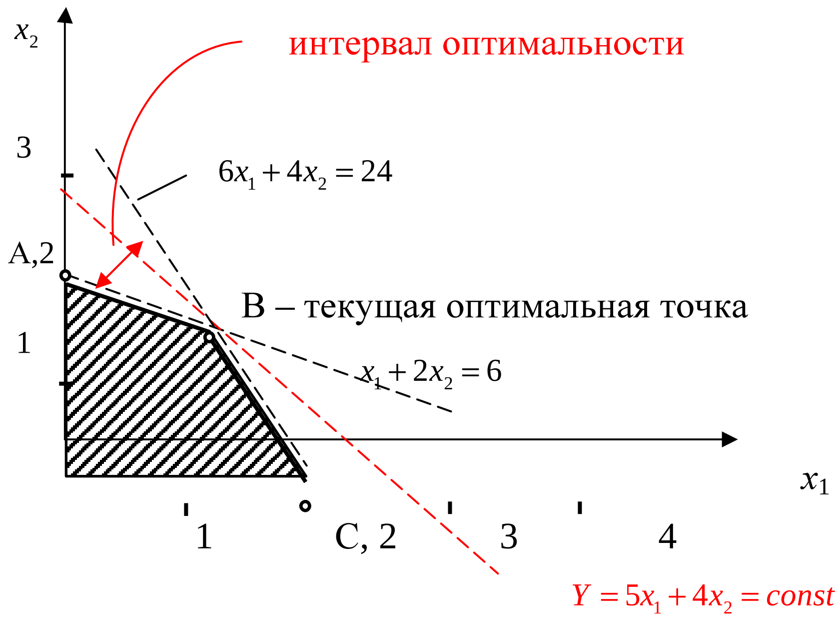
<u>**Задача 3**</u>. В каких пределах допустимо изменение коэффициентов целевой функции *Y*.

Изменение коэффициентов c_i оказывает влияние на наклон прямой (гиперплоскости), представляющей эту функцию. Это может привести к изменению совокупности связывающих ограничений и, следовательно, статуса того или иного ресурса, т.е. сделать недефицитный ресурс дефицитным и наоборот. При этом решаются следующие вопросы:

1. Каков диапазон изменения того или иного коэффициента c_i, при котором не происходит изменения оптимального решения.
2. На сколько следует изменить тот или иной коэффициент, чтобы сделать некоторый недефицитный ресурс дефицитным и наоборот.

Вместе с тем, очевидно, существуют интервалы изменения коэффициентов c_i и c_j, когда текущее значение $Y_{\max}$ сохраняется. В частности, для $Y = c_1x_1 + c_2x_2$ существует <u>интервал оптимальности</u> c_1/c_2 (c_2/c_1). Если значение c_1/c_2 не выходит за пределы этого интервала, то оптимальное решение в данной модели сохраняется.

Пусть имеется графическое решение некоторой задачи:



Если прямая $Y = c_1x_1 + c_2x_2$ совпадает с прямой $x_1 + 2x_2 = 6$ $\left(\dfrac{c_1}{c_2} = \dfrac{1}{2}\right)$, то оптимальным решением будет любая точка отрезка AB, включая и точку B.

Если Y совпадает с прямой $6x_1 + 4x_2 = 24$ $\left(\dfrac{c_1}{c_2} = \dfrac{6}{4}\right)$, то оптимальное решение в любой точке отрезка BC.

Имеем двойное неравенство:

$$(*)\ \boxed{\frac{1}{2} \le \frac{c_1}{c_2} \le \frac{6}{4}}$$

Можно определить интервал оптимальности для одного коэффициента c_i, считая другой неизменным.

Пусть $c_2 = 4$, тогда интервал оптимальности для c_1 получаем из неравенства (*) путем подстановки туда $c_2 = 4$. Получаем

$$2 \le c_1 \le 6$$

<u>Задание 2.</u> Составить математическую модель линейной производственной задачи.

A – матрица удельных затрат,

B – вектор объемов ресурсов,

C – вектор удельной прибыли при возможном выпуске четырех видов продукции с использованием четырех видов ресурсов

$$A = \begin{pmatrix} a_{11} & \cdots & a_{14} \\ \vdots & & \vdots \\ a_{31} & \cdots & a_{34} \end{pmatrix}; \qquad B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}; \quad C = \left(c_1, c_2, c_3, c_4\right).$$

Варианты заданий записаны в виде

c_1	c_2	c_3	c_4	
a_{11}	$\dots$	$\dots$	a_{14}	b_1
$\vdots$			$\vdots$	b_2
a_{31}	$\dots$	$\dots$	a_{34}	b_3

<u>1.7 Практикум</u>

<u>Линейная производственная задача</u>

№ 1

45	60	21	14	
3	6	3	0	180

№ 2

36	32	10	13	
2	3	4	1	103

№ 3

48	15	11	32	
4	2	3	1	116

6	2	0	6	210
2	3	5	7	112

4	2	0	2	148
2	8	7	0	158

2	0	3	2	94
4	1	0	5	196

№ 4

30	11	45	6	
3	2	6	0	150
4	2	3	5	130
4	3	2	4	124

№ 5

48	30	29	10	
3	2	4	3	198
2	3	1	2	96
6	5	1	0	228

№ 6

28	14	11	20	
4	2	2	4	112
2	3	1	0	32
1	4	0	2	46

№ 7

35	41	22	12	
2	2	3	4	151
3	1	0	2	156
1	4	4	0	162

№ 8

38	12	28	21	
3	0	3	3	186
2	3	1	1	102
4	3	2	2	196

№ 9

60	12	44	17	
4	2	4	1	180
4	0	2	2	160
2	4	3	0	109

№ 10

59	27	20	35	
1	3	2	2	102
3	2	0	3	204
4	2	3	1	188

№ 11

30	28	9	23	
1	0	2	5	110
3	6	0	4	126
2	4	1	3	114

№ 12

16	18	14	12	
4	3	0	6	192
0	1	5	0	24
1	2	4	3	90

№ 13

34	32	28	36	
2	4	5	3	128
3	0	4	1	130
3	5	0	2	142

№ 14

27	39	18	20	
2	1	6	5	140
0	3	0	4	90
3	2	4	0	198

№ 15

24	20	31	10	
3	0	2	5	162
3	6	0	3	134
2	4	3	1	148

№ 16

27	10	9	8	
3	5	0	6	144
2	0	1	0	130
1	4	2	3	140

№ 17

31	10	14	20	
1	4	3	4	120
3	0	2	2	168
2	5	0	3	80

№ 18

34	20	8	23	
2	0	2	3	142
1	5	4	2	100
3	4	0	1	122

№ 19

30	25	14	12	
2	3	0	4	148
4	1	5	0	116
0	2	4	3	90

№ 20

18	19	8	5	
3	2	0	3	168
0	1	4	2	60
1	3	5	0	112

№ 21

50	27	34	54	
5	4	6	7	275
2	0	4	2	100
3	2	0	1	85

№ 22

16	30	18	26

№ 23

44	28	78	23

№ 24

42	28	17	19

2	5	1	4	126
3	0	7	2	84
2	1	4	0	75

4	1	6	3	288
7	3	1	2	240
2	4	5	1	200

5	2	4	1	132
3	4	0	6	124
4	2	5	4	117

№ 25

31	10	41	29	
4	0	8	7	316
3	2	5	1	216
5	6	3	2	199

№ 26

36	14	25	50	
4	3	4	5	208
2	5	2	2	99
3	1	2	5	181

№ 27

48	33	16	22	
6	3	1	4	252
2	4	5	1	144
1	2	4	3	80

№ 28

36	30	16	12	
4	5	2	3	180
6	0	4	1	150
0	7	6	5	140

№ 29

45	33	30	42	
4	9	8	1	220
5	2	3	0	200
0	3	1	6	216

№ 30

35	10	54	40	
9	8	4	2	176
3	1	6	0	180
2	1	0	8	200

№ 31

8	21	17	36	
8	5	6	2	100
1	0	1	4	80
2	7	3	0	70

№ 32

21	16	32	18	
2	6	1	8	220
3	1	0	2	240
0	2	4	1	200

№ 33

13	24	17	25	
9	2	8	1	70
1	4	1	0	96
2	0	3	5	80

№ 34

31	16	18	8	
5	7	1	8	140
8	3	0	1	60
0	4	6	2	100

№ 35

15	16	52	13	
2	2	2	1	250
1	0	4	3	220
7	3	0	5	240

№ 36

32	8	10	18	
4	6	4	5	100
8	3	1	0	72
0	2	7	9	63

№ 37

18	32	5	6	
4	2	3	1	80
0	8	1	2	96
6	0	1	1	84

№ 38

40	15	27	55	
5	3	2	5	175
2	1	0	6	140
1	0	3	4	80

№ 39

6	32	25	68	
3	4	1	5	142
5	1	0	2	63
0	2	3	6	106

№ 40

1	20	12	54	
2	3	5	4	105
1	0	1	3	24
0	4	2	6	108

№ 41

47	28	15	9	
5	4	0	1	181
3	5	2	6	165
2	0	3	1	50

№ 42

17	80	48	33	
1	6	2	4	186
0	5	6	1	259
2	4	0	3	68

№ 43

32	43	11	16	

№ 44

6	30	106	54	

№ 45

54	20	21	45	

5	3	0	2	253	1	3	5	4	231	3	2	1	4	247
1	4	2	1	105	2	4	0	3	84	2	1	0	5	215
3	0	4	6	140	0	1	6	3	224	6	3	4	0	210

№ 46

17	59	71	4	
1	3	5	1	266
2	4	6	0	326
3	0	2	4	110

№ 47

25	26	28	80	
1	3	4	2	202
2	1	0	6	318
3	0	5	1	211

№ 48

50	27	34	54	
5	4	6	7	275
2	0	4	2	100
3	2	0	1	85

№ 49

26	35	18	30	
2	5	1	4	126
3	0	7	2	84
2	1	4	0	75

№ 50

44	28	78	23	
4	1	6	3	288
7	3	1	2	240
2	4	5	1	200

№ 51

42	28	17	19	
5	2	4	1	132
3	4	0	6	124
4	2	5	4	117

№ 52

31	10	41	29	
4	0	8	7	316
3	2	5	1	216
5	6	3	2	199

№ 53

36	14	25	50	
4	3	4	5	208
2	5	0	2	99
3	1	2	5	181

№ 54

48	33	16	22	
6	3	1	4	252
2	4	5	1	144
1	2	4	3	80

1.8 Транспортная задача ЛП

Постановка задачи. Однородный продукт, сосредоточенный в m пунктах производства (хранения) в количестве a_i единиц $\left(i=\overline{1,m}\right)$, необходимо распределить между n пунктами потребления, которым необходимо $b_1,\ldots,b_n$ единиц.

c_{ij} – стоимость перевозки единицы продукции из i в j (известная величина).

x_{ij} – количество груза, планируемое к перевозке от i пункта поставки к j пункту потребителя.

$Y=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}$ – стоимость всех перевозок.

$$A=\left\{\begin{matrix}a_1\\ \vdots\\ a_i\\ \vdots\\ a_m\end{matrix}\right\} \xrightarrow[x_{ij}]{c_{ij}} \left\{\begin{matrix}b_1\\ \vdots\\ b_j\\ \vdots\\ b_n\end{matrix}\right\}=B$$

Составить план перевозок из пунктов A_i, $i=\overline{1,m}$ в пункты B_j, $j=\overline{1,n}$ при <u>минимальных</u> транспортных расходах.

<u>Решение ТЗ всегда существует.</u>

<u>Математическая постановка</u>

Найти $Y=\min\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}$ <u>при ограничениях:</u>

$$(*)\qquad \left.\begin{aligned} &\sum_{i=1}^{m} x_{ij}=b_j,\quad j=\overline{1,n} \longrightarrow \\ &\sum_{j=1}^{n} x_{ij}=a_i,\quad i=\overline{1,m} \longrightarrow \end{aligned}\right|\; x_{ij}\ge 0$$

$\forall$ потребителю доставляется требуемое колитество груза

из $\forall$ пункта производства вывозится весь груз

<u>Здесь Т задача сбалансирована</u> (<u>замкнутая (закрытая) модель</u>)

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

В случае отсутствия баланса производства и потребления система (*) заменяется системой неравенств.

<u>Например</u>, $a_1+\ldots+a_m>b_1+\ldots+b_n$, то вместо $\sum_j x_{ij}=a_i$ будем иметь

$$x_{i1}+\ldots+x_{in}<a_i,\quad i=\overline{1,m}$$

Получилась открытая модель ТЗ.

При нарушении баланса вводят «фиктивного» потребителя и «фиктивный» пункт назначения.

Для решения ТЗ чаще всего применяется <u>метод потенциалов</u>.

Первое $\left|\dfrac{\text{базисное допустимое решение}}{\text{(или опорный план)}}\right|$ можно построить различными способами.

I <u>Метод «северо-западного» угла</u>. Рассмотрим пример сбалансированной задачи.

	B_1	B_2	B_3	B_4	
B_j / A_i	$\cancel{30}$ 16	$\cancel{22}$ 18	$\cancel{15}$ 7	34	$\Sigma = 101$
A_1: 14	$\boxed{70}$ 14	$\boxed{38}$ -	$\boxed{24}$ -	$\boxed{92}$ -	
A_2: $\cancel{20}$ 4	$\boxed{58}$ 16	$\boxed{18}$ 4	$\boxed{56}$ -	$\boxed{72}$ -	
A_3: $\cancel{26}$ 8	$\boxed{19}$ -	$\boxed{10}$ 18	$\boxed{100}$ 8	$\boxed{30}$ -	
A_4: 41	$\boxed{3}$ -	$\boxed{36}$ -	$\boxed{121}$ 7	$\boxed{8}$ 34	

$\Sigma = 101$

$X = \left\| x_{ij} \right\|$ – количество перевозимого груза из i в j.
В уголках клеток x_{ij} стоят стоимости перевозки единицы груза из i в j.

<u>Алгоритм</u>:

1) выбираем x_{11} и сравниваем a_1, b_1.
2) если $a_1 \le b_1$, тогда $x_{11} = \min(a_1, b_1) = 14$; переходим вниз по столбцу к клетке x_{21}; выбираем $\min(a_2, b_1 - a_1) = 16$; в клетке $b_1 = 30 - 14 = 16$ и т.д.
3) если $a_1 > b_1$, тогда $x_{11} = b_1$ и переходим по строке вправо; находим $\min\{a_1 - b_1, b_2\}$ и т.д.

$Y = 14 \cdot 70 + 16 \cdot 58 + 4 \cdot 18 + 18 \cdot 10 + 8 \cdot 100 + 7 \cdot 121 + 34 \cdot 8 = 4079$

Решение не оптимальное, т.к. не учитывалась стоимость перевозок.

II <u>Правило: «самая дешевая продукция по стоимости перевозки» реализуется первой.</u>

1) Самая «дешевая клетка» (4, 1). Тогда $A_4 \xrightarrow{30} B_1$. Первый столбец отсекается. В A_4 осталось $41 - 30 = 11$.
2) Самая «дешевая клетка» из оставшихся (4, 4). Тогда $A_{40} \xrightarrow{11} B_4$. Последняя стока отсекается. В B_4 стало $34 - 11 = 23$.

A_i \ B_j	30	22	~~15~~ 1	~~34~~ ~~23~~ 19
14	70 / -	38 / -	**24** / 14	92 / -
20	58 / -	18 / -	56 / 1	72 / 19
~~26~~ 4	19 / -	**10** / 22	100 / -	30 / 4
~~41~~ 11	**3** / 30	36 / -	121 / -	**8** / 11

3) Из оставшихся клеток «самая дешевая» (3, 2). Тогда $A_3 \xrightarrow{22} B_2$. Второй столбец исчезает.

4) (1, 3): $A_1 \xrightarrow{14} B_3$. Первая строка исчезает.

5) (3, 4): $A_3 \xrightarrow{4} B_4$. Третья строка исчезает.

6) (2, 3): $A_2 \xrightarrow{1} B_3$

7) (2, 4): $A_2 \xrightarrow{19} B_4$;

$Y = 14 \cdot 24 + 1 \cdot 56 + 19 \cdot 72 + 22 \cdot 10 + 4 \cdot 30 + 30 \cdot 3 + 11 \cdot 8 = 2278.$

III <u>Способ аппроксимации Фогеля</u> (концентрация штрафов за выбор не самого оптимального маршрута)

A_i \ B_j	30	~~22~~ 2	~~15~~ 1	34	1	2	3	4	5
14	70 / -	38 / -	24 / 14	92 / -	14	14	-	-	-
20	58 / -	18 / 20	56 / -	72 / -	38	-	-	-	-
~~26~~ 24	19 / 23	10 / 2	100 / 1	30 / -	9	9	9	11	81
~~41~~ 7	3 / 7	36 / -	121 / -	8 / 34	5	5	5	5	118
1	16	8	32	22					
2	16	26	76	22					
3	16	26	21	22					
4	16	-	21	22					
5	16	-	21						

$Y = 14{\cdot}24 + 20{\cdot}18 + 23{\cdot}19 + 2{\cdot}10 + 1{\cdot}100 + 7{\cdot}3 + 34{\cdot}8 = 1546.$

а) <u>Первая окаймляющая полоса</u>: разность в каждом столбце и каждой строке между двумя наименьшими стоимостями перевозок.

б) Выбор на этой полосе наибольшего значения – 38.

в) Помещение в клетку (2, 2) с наименьшей стоимостью (18) максимального количества вывозимого товара $A_2 \xrightarrow{20} B_2$. Строка 2 исчезает.

г) Строится вторая полоса без учета рядов с выделенными полностью ресурсами. Наибольшее значение 76. Тогда $A_1 \xrightarrow{14} B_3$ и т.д.

(IV) <u>Максимально полное удовлетворение спроса при минимуме стоимости перевозок (полный минимальный вывоз).</u> (max B min C)

A_i \ B_j	~~30~~ 23	~~22~~ 19	15	34
14	$\boxed{70}$ -	$\boxed{38}$ -	$\boxed{24}$ 14	$\boxed{92}$ -
20	$\boxed{58}$ -	$\boxed{18}$ 19	$\boxed{56}$ 1	$\boxed{72}$ -
~~26~~ 7	$\boxed{19}$ 23	$\boxed{10}$ 3	$\boxed{100}$ -	$\boxed{30}$ -
~~41~~ 7	$\boxed{3}$ 7	$\boxed{36}$ -	$\boxed{121}$ -	$\boxed{8}$ 34

1) min (4, 1) $\Rightarrow$ полностью не удовлетворяет потребителя

2) min (4, 4) $\Rightarrow$ $A_4 \xrightarrow{34} B_4$

3) min (4 ,1) $\Rightarrow$ $A_4 \xrightarrow{7} B_1$

4) min (3, 2) $\Rightarrow$ $A_3 \xrightarrow{22} B_2$

…

и т.д.

$Y = 1494.$

1) $\max B_j = 34, \quad \min c_{44} = 8, \quad A_4 \xrightarrow{34} B_4$

2) $\max \tilde{B}_j = 30, \quad \min c_{41} = 3, \quad A_4 \xrightarrow{7} B_1$

3) $\max \tilde{\tilde{B}}_j = 23, \quad \min c_{31} = 19, \quad A_3 \xrightarrow{23} B_1$

4) $\max \tilde{\tilde{\tilde{B}}}_j = 22, \quad \min c_{32} = 10, \quad A_3 \xrightarrow{3} B_2$

и т.д.

$\underbrace{Y = 1494}$ план сразу оптимальный

Для улучшения опорного плана существует несколько способов:

- метод потенциалов (двойственный симплекс-метод)
- метод последовательного улучшения

- метод stepping-stone.

Основой алгоритма этих методов является определение критерия оптимальности

$\delta_{ij} = c_{ij} - z_{ij}$, где z_{ij} – затраты с доставкой единицы ресурса из i в j, где эти ресурсы еще не использованы.

План оптимальный $\Rightarrow \delta_{ij} \geq 0$.

1.9 Метод потенциалов

Улучшим опорное решение, полученное по методу Фогеля.

$A_i \backslash B_j$	30	22	15	34
14			$\boxed{24}$ 14	
20		$\boxed{18}$ 20		
26	$\boxed{19}$ 23	$\boxed{10}$ 2	$\boxed{100}$ 1	
41	$\boxed{3}$ 7			$\boxed{8}$ 34

Ненулевые клетки называются базисными, их ($m + n - 1$).

Нулевые клетки называются свободными, их ($n - 1$) · ($m - 1$).

Свободные клетки имеют потенциал.

Пусть α_i – потенциал пункта A_i,

β_j – потенциал пункта B_j.

$c_{ij}^* = \alpha_i + \beta_j$ – косвенная стоимость (тариф).

Базисные клетки имеют свойство: $c_{ij} = c_{ij}^*$ и для них составляется совместная система линейных уравнений. Число неизвестных этой системы на одно больше числа уравнений, поэтому одно из неизвестных приравниваем нулю.

Например, наиболее часто встречающееся.

$$\left.\begin{aligned}
&(1,3):\alpha_1+\beta_3=24;\\
&(2,2):\alpha_2+\beta_2=18;\\
&(3,1):\alpha_3+\beta_1=19;\\
&(3,2):\alpha_3+\beta_2=10;\\
&(3,3):\alpha_3+\beta_3=100;\\
&(4,1):\alpha_4+\beta_1=3;\\
&(4,4):\alpha_4+\beta_4=8;
\end{aligned}\right\}$$

Пусть $\alpha_3=0,$ тогда $\beta_1=19;\ \ \beta_2=10;\ \ \beta_3=100;$ $\beta_4=24;\ \ \alpha_1=-76;\ \ \alpha_2=8;\ \ \alpha_3=0;\ \ \alpha_4=-16.$

Вычисляем косвенные тарифы для свободных клеток

$c^*_{11}=\alpha_1+\beta_1=-57;$ $c^*_{12}=\alpha_1+\beta_2=-66;$

$c^*_{14}=-49;$ $c^*_{21}=27;$ $\boxed{c^*_{23}=108;\ >56}$

$c^*_{24}=32;$ $c^*_{34}=24;$ $c^*_{42}=-6;$ $c^*_{43}=-84;$

<u>Проводим проверку на оптимальность</u>: для всех свободных клеток имеем

1) $c^*_{ij}\le c_{ij}\Rightarrow$ план оптимальный,

2) $c^*_{ij}< c_{ij}\Rightarrow$ план оптимальный и единственный,

3) $c^*_{ij}= c_{ij}\Rightarrow$ план оптимальный, но НЕ единственный.

В нашем случае $c^*_{23}=108>c_{23}=56$, следовательно, план неоптимальный.

<u>Проводим улучшение плана</u>: пересчет по циклу для клетки, в которой $c^*_{ij}>c_{ij}$.

<u>Цикл</u> – замкнутый многоугольник с одной свободной вершиной, для которой производится пересчет.

Вершины соединены ломаной замкнутой линией, которая в каждой точке излома совершает поворот на 90°.

Каждый цикл имеет <u>четное число</u> вершин. Вершины помечаются знаками ⊕ и ⊖. Свободная клетка имеет знак ⊕, остальные знаки чередуются.

<u>Пересчет по циклу</u>:

1) Находят min из чисел в отрицательных вершинах (пусть это Δ).

2) К числам с положительными вершинами прибавляется Δ, а из чисел в отрицательных вершинах вычитается Δ. В этом случае, свободная клетка становится базисной, а базисная клетка освобождается. Равновесие между заявками и ресурсами сохраняется.

Осуществим пересчет по клетке (2, 3)

⊖ 20	⊕ (2,3)	Δ = min (20; 1) = 1	19	1

⊕ 2	⊖ 1		3	

Для базисных клеток снова составляем систему уравнений:

B_j / A_i	30	22	15	34
14	$\boxed{70}$	$\boxed{38}$	$\boxed{24}$ 14	$\boxed{90}$
20	$\boxed{58}$	$\boxed{18}$ 19	$\boxed{56}$	$\boxed{72}$
26	$\boxed{19}$ 23	3	$\boxed{100}$	$\boxed{30}$
41	$\boxed{3}$ 7	$\boxed{36}$	$\boxed{121}$	$\boxed{8}$ 34

$$(1;3):\alpha_1+\beta_3=24;$$
$$(2;2):\alpha_2+\beta_2=18;$$
$$(2;3):\alpha_2+\beta_3=56;$$
$$(3;1):\alpha_3+\beta_1=19;$$
$$(3;2):\alpha_3+\beta_2=10;$$
$$(4;1):\alpha_4+\beta_1=3;$$
$$(4;4):\alpha_4+\beta_4=8.$$

Пусть $\alpha_3=0$, тогда

$$\beta_1=19;\quad \beta_2=10;\quad \beta_3=48;\quad \beta_4=24;$$
$$\alpha_1=-24;\quad \alpha_2=8;\quad \alpha_3=0;\quad \alpha_4=-16;$$

Вычисляем косвенные тарифы:

$c_{11}^*=\alpha_1+\beta_1=-24+19=-5<70;$ $c_{12}^*=-14<38;$

$c_{14}^*=0<90;$ $c_{21}^*=27<58;$ $c_{24}^*=32<72;$

$c_{33}^*=48<100;$ $c_{34}^*=24<30;$ $c_{42}^*=-6<36;$

$c_{43}^*=32<121.$

$\Rightarrow$ план оптимальный и единственный

Затраты: Y = 14·24 + 19·18 + 1·56 + 23·19 + 3·10 + 7·3 + 34·8 = 1494.

Замечание. Вырождение ↔ число базисных клеток < (m + n – 1). В этом случае в недостающее число свободных клеток вводят нулевые поставки и рассматривают эти клетки, как заполненные (базисные).

Экономический смысл потенциалов: это локальные (поясные) цены или наценки к единой цене, создающие заинтересованность в правильном направлении перевозок.

Критерий оптимальности выражает следующее: если какая-то перевозка осуществляется, то цена в пункте B_j равна цене в пункте A_i плюс транспортные расходы, т.е. $\delta_{ij}=0$ и $\beta_j=\alpha_i+c_{ij}$.

В остальных случаях цена β_j не может быть больше, чем $\alpha_i + c_{ij}$, т.к. ресурс в B_j по такой цене можно было бы получить, приведя его с затратами c_{ij} из пункта A_i. Следовательно, $\beta_j \le \alpha_i + c_{ij}$, т.е. в обоих указанных случаях разность цен не превышает затрат по перевозке и критерий оптимальности $\delta_{ij} \ge 0$.

ТЗ с избытком запасов

$A_i \backslash B_j$	B_1	B_2	B_3	запасы a_i
A_1	5	7	6	50
A_2	6	6	5	40
A_3	8	4	5	20
заявки b_j	18	21	33	

$\Sigma = 100$

$\Sigma\ 72$

$$\underbrace{\sum_{j=1} b_j < \sum_i a_i}$$

разность = 38
Введем фиктивную заявку (потребит.) B_f с f_f = 38.
с нулевыми стоимостями

$A_i \backslash B_j$	B_1	B_2	B_3	B_f	запасы a_i
A_1	5 5 18	7 7 21	6 6 11 −	1 0 +	50
A_2	4 6	6 6	5 5 + 22	0 0 − 18	40
A_3	4 8	6 4	5 5	0 0 20	20
заявки b_j	18	21	33	38	110

Далее находим опорный план (например, методом северо-западного угла, как не оптимального).

В левых верхних углах клеток поставим косвенные тарифы.

В клетке (1, 4) $c^*_{14} > c_{14}$ – по ней осуществим пересчет.

- 11	+
+ 22	- 18

$\Delta = \min (11)$

0	11
33	7

	B_1	B_2	B_3	B_f
A_1	5 5 18	7 7 21 -	5 6	0 0 11 +
A_2	5 6	7 6	5 5 33	0 0 7
A_3	5 8	7 4 +	5 5	0 0 20 -

A = min (20)

7 > 4
$c^*_{32} > c_{32}$

	B_1	B_2	B_3	B_f
A_1	5 5 18	7 7 1 -	5 6	0 0 + 31
A_2	5 6	7 6 +	5 5 33	0 0 7 -
A_3	2 8	4 4 20	2 5	-3 0

Клетка (2, 2): $c^*_{22} > c_{22}$ – по ней пересчет

И, окончательно, получили оптимальный план, т.к. $\forall c^*_{ij} \le \forall c_{ij}$

	B_1	B_2	B_3	B_f	
A_1	5 5 18	7 7 21 -	5 6	0 0 11 +	50
A_2	5 6	7 6	5 5 33	0 0 7	40
A_3	5 8	7 4 +	5 5	0 0 20 -	20
	18	21	33	38	

Из 50 единиц груза в пункте A_1 не перевозятся 32.

Из 40 единиц груза в пункте A_2 не перевозятся 6.

1.10 Практикум

Транспортная задача линейного программирования

№ 1

	45	60	21	24
50	3	6	3	1
70	6	2	1	6
40	2	3	5	7

№ 2

	36	32	40	53
40	2	3	4	1
60	4	2	1	2
70	2	8	7	1

№ 3

	48	75	41	32
90	4	2	3	1
75	2	1	3	2
40	4	1	2	5

№ 4

	30	11	45	36
50	3	2	6	1
70	4	2	3	5
30	4	3	2	4

№ 5

	48	30	29	40
40	3	2	4	3
45	2	3	1	2
70	6	5	6	40

№ 6

	28	44	31	20
50	4	2	2	4
40	2	3	1	5
42	1	4	3	2

№ 7

	35	41	52	32
70	2	2	3	4
80	3	1	5	2
47	1	4	4	3

№ 8

	38	42	28	41
60	3	2	4	3
50	2	3	1	4
48	4	3	6	2

№ 9

	60	32	44	57
50	3	2	4	1
90	4	6	5	2
60	2	4	3	6

№ 10

	59	27	40	35
45	1	3	2	2
55	3	2	4	3

№ 11

	30	58	32	43
65	1	3	2	5
40	3	6	5	4

№ 12

	46	48	44	42
70	4	3	2	6
90	3	1	5	4

70	4	2	3	1

70	2	4	1	3

33	1	2	4	3

№ 13

	34	32	4	36
60	2	4	5	3
50	3	7	4	1
48	3	5	6	2

№ 14

	37	39	48	40
70	2	1	6	5
40	5	3	7	4
60	3	2	4	2

№ 15

	24	20	31	40
30	3	7	2	5
45	3	6	5	3
52	2	4	3	1

№ 16

	27	20	39	42
35	3	5	3	6
60	2	6	1	7
40	1	4	2	3

№ 17

	31	40	44	20
45	1	4	3	4
50	3	6	2	2
53	2	5	6	3

№ 18

	34	40	38	53
80	2	7	2	3
60	1	5	4	2
30	3	4	6	1

№ 19

	30	55	44	42
35	2	3	6	4
55	4	1	5	7
80	5	2	4	3

№ 20

	48	59	68	75
90	3	2	6	3
80	5	1	4	2
92	1	3	5	4

№ 21

	50	27	34	54
70	5	4	6	7
50	2	3	4	2
54	3	2	5	1

№ 22

	56	35	48	30
60	2	5	1	4
50	3	4	7	2
70	2	1	4	3

№ 23

	44	28	78	23
40	4	1	6	3
60	7	3	1	2
80	2	4	5	1

№ 24

	42	28	47	9
50	5	2	4	1
90	3	4	7	6
35	4	2	5	4

№ 25

	31	40	41	49
45	4	5	8	7
60	3	2	5	1
65	5	6	3	2

№ 26

	36	44	25	50
55	4	3	4	5
60	2	5	3	2
48	3	1	2	5

№ 27

	48	33	56	22
60	6	3	1	4
70	2	4	5	1
35	1	2	4	3

№ 28

	42	35	27	28
58	3	5	3	2
40	4	2	41	3
45	3	1	62	4

№ 29

	59	33	34	20
70	2	5	4	1
30	7	3	2	3
50	1	4	2	5

№ 30

	44	58	62	24
90	3	4	5	3
60	4	3	1	2
50	3	1	2	2

№ 31

	26	48	25	22

№ 32

	23	41	22	20

№ 33

	21	43	46	59

40	3	6	1	5	30	4	2	5	2	80	2	1	6	8
45	2	4	3	1	40	1	4	1	3	50	9	3	5	4
50	1	7	5	3	45	5	3	7	6	45	3	4	2	7

Тема 2. Теория игр

Введение

Очень часто в задачах бизнес-анализа два или более лиц принимают решения независимо друг от друга, и решение, принятое каждым конкретным лицом, влияет на результаты остальных. **ТЕОРИЯ ИГР** термин представляет собой русский эквивалент английского *theory of games* и используется для обозначения комплекса математич. моделей конфликтных ситуаций и способов их разрешения, основы к-рого разработаны математиком Дж. фон Нейманом. Формализованное описание игры задается списком ее участников (игроков) и множества стратегий для каждого из них. В рез-те выбора стратегий игроками образуется ситуация (состояние) игры. Интересы игроков характеризуются функциями выигрыша или отношениями предпочтения на множестве допустимых ситуаций. Т.обр., в понятии игры моделируются два основных факта: а) каждый участник конфликта лишь частично контролирует ситуацию; б) каждый участник имеет свои интересы. Нормативное направление в Т.и. занимается исследованием вопросов, какие состояния игры считать справедливыми, равновесными, оптимальными, а также анализом свойств и способов достижения таких состояний. Дескриптивное направление изучает различные способы поведения игроков и свойства результирующих состояний.

Наибольшие успехи достигнуты в Т.и. двух игроков с противоположными интересами (антагонистич. игры), где нормативный и дескриптивный аспекты конфликтной ситуации хорошо совмещаются в понятии седловой точки (максимина) состояния, в к-ром каждый игрок получает максимум выигрыша по контролируемым им переменным в условиях, когда этот выигрыш минимален по переменные, контролируемым др. игроком. В частности, для случая, когда множества стратегий обоих игроков конечны (матричная игра), Дж. фон Нейман установил, что седловая точка существует, если разрешить игрокам использовать смешанные стратегии вероятностный механизм выбора стратегий (теорема о минимаксе).

Теория антагонистич. игр находит применение в военных приложениях: в вопросах стратегии и тактики. Оказалось также, что антагонистич. игры во многих аспектах эквивалентны задачам *математического программирования*. Игровая методология является основой перспективного направления математич. статистики, трактующего статистич. задачи как игры исследователя с природой.

Анализ игр многих лиц существенно затруднен из-за сложности вопроса о механизмах формирования и действия коалиций. Моделирование коалиционных взаимодействий как антагонистич. игр привело к т.н. теории кооперативных игр, к-рая представляет интерес лишь с математич. т.зр. В теории бескоалиционных игр многих лиц имеются два направления, имеющие нетривиальное приложение к социально-экономич. проблематике. Одно из них игры с непротивоположными интересами и фиксированной последовательностью ходов, моделирование принятия решений в организационных системах на основе принципа гарантированного рез-та. Согласно этому принципу, каждый игрок при своем ходе выбирает стратегию, исходя из предположения, что следующие за ним участники будут максимизировать свои выигрыши в условиях, определенных всеми предыдущими выборами.

Др. направление связано с понятием равновесия (Нейман-Нэш) ситуации, устойчивой в том смысле, что никакой игрок не может увеличить свой выигрыш за счет только собственных действий. Это понятие, в частности, лежит в основе концепции социально-экономич. равновесия, согласно к-рой в равновесии все социальные в и экономич. агенты добиваются максимально возможного удовлетворения своих интересов в рамках определенных ограничений, причем предложение соответствует спросу по всем видам рассматриваемых благ и труда. Данная концепция используется для анализа ряда социально-экономич. процессов: поведение в условиях дефицита, распределение доходов, семейное поведение, межрегиональные взаимодействия и др.

В целом идеи Т.и. имеют несомненное стимулирующее значение как для внутриматематич., так и для социально-экономич. исследований, но в последнем случае собственные ее концепции слишком абстрактны и должны дополняться более конкретными конструкциями в каждом приложении.

2.1. Основные понятия теории игр

Теория игр— это математическая дисциплина, исследующая ситуации, в которых принятие решений зависит от нескольких участников. Интересы участников могут быть как антагонистические (полностью противоположные) так и неантагонистические. В последнем случае может исследоваться вопрос о наиболее эффективных совместных действиях (кооперативные игры).

Игра — это упрощенная формализованная модель реальной ситуации, описывающая действия двух или более участников. Предполагается, что известны варианты действий сторон (стратегии), исход игры для каждого участника в случае выбора конкретных действий всеми участниками, степень и порядок информированности каждого участника игры о поведении всех других участников.

Приведем классификацию игр в зависимости от различных параметров.

Количество игроков. Различаются игры двух лиц (2 участника игры) и игры N лиц (число участников более 2).

Количество стратегий. Если каждый из игроков имеет конечное число возможных стратегий в игре, то игра называется конечной. Если число стратегий хотя бы одного из участников игры бесконечно, то игра называется бесконечной.

Соотношение интересов участников. Игра с нулевой суммой - сумма выигрышей участников всегда равна нулю (антагонистические интересы — антагонистические игры). Игры с ненулевой суммой.

Возможности взаимодействия участников. С этой точки зрения можно рассматривать коалиционные (допускается образование коалиций между участниками), бескоалиционные (коалиции не допускаются) и кооперативные игры (коалиции определены заранее).

Тип функции выигрыша. По данному критерию традиционно рассматриваются такие классы игр, как матричные (игра 2-х лиц, выигрыш одного из игроков (соответственно проигрыш другого) задается в виде матрицы), би-матричные (игра 2-х лиц, выигрыш каждого из игроков задается своей матрицей), непрерывные (функция выигрышей является непрерывной функцией на множестве стратегий каждого из игроков), выпуклые (функция выигрышей есть выпуклая функция на множестве стратегий) и так далее.

Количество ходов. Если после одного хода каждого игрока игра заканчивается и происходит распределение выигрышей, то игра называется одношаговой. В противном случае игра называется многошаговой (позиционной, например, шахматы).

Кроме этого выделяются различные классы игр по иным признакам (статистические, дифференциальные и многие другие). В частности рассматриваются так называемые «игры с природой», т.е. игры, когда в качестве второго игрока выступает не игрок с противоположными интересами, а некоторая сторона с «неопределенными» интересами (природа). В этом случае для поиска оптимальных стратегий используются наряду с принципом гарантированного результата и другие критерии например Максимакса, Вальда, Сэвиджа, Гурвица, которые рассмотрим далее.

2.2. Игры двух лиц с нулевой суммой

2.2.1. Основные предположения для игр двух лиц с нулевой суммой.

В игре с нулевой суммой сумма выигрышей игроков всегда равна нулю. Платильщиком выигрыша первого игрока является второй игрок. Таким образом, какая-либо кооперация между ними не возможна.

Игра двух лиц с нулевой суммой задается следующими условиями:

имеются два игрока, стратегии одного из которых расположены по строкам (первый игрок), а другого по столбцам (второй игрок);

каждый игрок выбирает одну из своих стратегий независимо от другого: первый одну из ***m*** стратегий, второй одну из **n**;

если первый игрок выбирает стратегию i, а второй стратегию *j*, то первый игрок получает выигрыш a_{ij}, который интерпретируется как платеж от второго игрока.

Такая игра называется игрой двух лиц с нулевой суммой и представляется в виде матрицы игры (табл. 1), которая содержит выигрыши первого игрока (или, как отмечалось, проигрыши второго игрока).

Таблица 1

	Стратегия 1	**Стратегия 2**	**….**	**Стратегия n**
Стратегия 1	a_{11}	a_{12}		a_{1n}
Стратегия 2 Игрока	a_{21}	a_{22}		a_{2n}
….				
Стратегия m Игрока 1	a_{m1}	a_{m2}		a_{mn}

В табл. 2 приведена некоторая конкретная матрица игры, согласно которой выигрыш первого игрока составит 2 единицы, если первый игрок выберет свою вторую стратегию, а второй игрок свою первую стратегию.

Таблица

	Стратегия	**Стратегия**	**Стратегия**	**Стратегия**
Страте-	**1**	**2**	**3**	**-1**
Страте-	**2**	**1**	**-2**	**0**

В игре с нулевой суммой сумма выигрышей игроков всегда равна нулю. Как уже отмечалось, плательщиком выигрыша первого игрока является второй игрок. Таким образом, какая-либо кооперация между ними не возможна.

2.2.2. Верхнее и нижнее значение игры, условие седловой точки.

Предполагается, что каждый из игроков знает стратегию своего противника и платежную матрицу игры. Рассмотрим с этой точки зрения некоторую конкретную игру (табл. 3).

Таблица 3

	Стратегия 1	Стратегия 2	Стратегия 3	Минимум по строкам
Стратегия 1	4	4	**10**	**4**
Стратегия 2	2	3	1	**1**
Стратегия 3	6	5	7	**5**
Максимум по столбцам	**6**	**5**	**10**	

Как должен играть первый игрок? Если первый игрок выберет свою первую стратегию, то второй игрок, очевидно, выберет первую или вторую, поскольку в этом случае его потери будут минимальными - 4 единицы. Значение «4» является минимальным в первой строке. Рассуждая аналогично, легко видеть, что если первый игрок выбирает свою вторую стратегию, то второй выбирает 3-ю, проигрывая при этом 1. В крайнем правом столбце таблицы записаны минимумы по строкам. Логично предположить, что первый игрок будет выбирать стратегию, обеспечивающую ему выигрыш максимального из этих значений.

Мы показали, что первый игрок может гарантированно выиграть по крайней мере 5 единиц. Он понимает, что на большее он рассчитывать не может, так как, выбирая стратегию 2, второй игрок обеспечивает выигрыш первого не более 5.

Матрица, которую мы рассматриваем, удовлетворяет условию **седловой точки**:

max (минимумы по строкам) = min (максимум по столбцам).

Говорят, что если выполнено это условие, то игра имеет седловую точку.

Нижнее значение любой матричной игры не превосходит верхнего значения.

Если игра имеет седловую точку, то первый игрок может выбирать любую стратегию, для которой реализуется максимум в левой части соотношения (максиминная стратегия), а второй игрок может выбрать любую стратегию, на которой реализуется минимум в правой части соотношения (минимаксная стратегия). Если игра имеет седловую точку, то общее значение выигрыша (проигрыша), которое достигается слева и справа в соотношении, называется **ценой игры**.

Седловая точка может рассматриваться как точка равновесия в том смысле, что отклонение от нее для каждого из игроков невыгодно. Действительно, в нашем примере если первый игрок сменит свою оптимальную стратегию 2 на 1 или 3, то выигрыш первого (соответственно проигрыш второго) увеличится.

Игра двух лиц с нулевой суммой далеко не всегда имеет седловую точку. Например, рассмотрим игру со следующей платежной матрицей (табл. 4).

Таблица 4

	Стратегия	Стратегия	Минимум
Стратегия	**-1**	**+ 1**	**-1**
Страте-	**+ 1**	**-1**	**-1**
Максимум по столбцам	**+1**	**+1**	

Как легко видеть, условие существования седловой точки для данной платежной матрицы не выполняется в силу того что справедливо условие

$$\max\min a_{ij} = -1 < \min\max a_{ij} = +1$$

Рассмотрим другое пояснение матричных игр.
Матричная игра двух игроков с нулевой суммой может рассматриваться как следующая абстрактная игра двух игроков.
Первый игрок имеет m стратегий $i = 1,2,...,m$, второй имеет n стратегий $j = 1,2,...,n$. Каждой паре стратегий (i,j) поставлено в соответствие число a_{ij}, выражающее выигрыш игрока 1 за счёт игрока 2, если первый игрок примет свою i-ю стратегию, а 2 – свою j-ю стратегию.
Каждый из игроков делает один ход: игрок 1 выбирает свою i-ю стратегию ($i=\overline{1,m}$), 2 – свою j-ю стратегию ($j=\overline{1,n}$), после чего игрок 1 получает выигрыш a_{ij} за счёт игрока 2 (если $a_{ij} < 0$, то это значит, что игрок 1 платит второму сумму $|a_{ij}|$). На этом игра заканчивается.
Каждая стратегия игрока $i=\overline{1,m}$; $j = \overline{1,n}$ часто называется чистой стратегией.
Если рассмотреть матрицу

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ \multicolumn{6}{c}{\dotfill} \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \multicolumn{6}{c}{\dotfill} \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

то проведение каждой партии матричной игры с матрицей A сводится к выбору игроком 1 i-й строки, а игроком 2 j-го столбца и получения игроком 1 (за счёт игрока 2) выигрыша a_{ij}.
Главным в исследовании игр является понятие оптимальных стратегий игроков. В это понятие интуитивно вкладывается такой смысл: стратегия игрока является оптимальной, если применение этой стратегии обеспечивает ему наибольший гарантированный выигрыш при всевозможных стратегиях другого игрока.
Исходя из этих позиций, игрок 1 исследует матрицу выигрышей A следующим образом: для каждого значения i ($i=\overline{1,m}$) определяется минимальное значение выигрыша в зависимости от применяемых стратегий игрока 2

$$\min_j a_{ij} \qquad (i = \overline{1,m})$$

т.е. определяется минимальный выигрыш для игрока 1 при условии, что он примет свою i-ю чистую стратегию, затем из этих минимальных выигрышей отыскивается такая стратегия $i = i_o$, при которой этот минимальный выигрыш будет максимальным, т.е. находится

$$\max_i \min_j a_{ij} = a_{i_o j_o} = \underline{\alpha} \qquad (1).$$

Определение. Число $\underline{\alpha}$, определённое по формуле (1) называется *нижней чистой ценой игры* и показывает, какой минимальный выигрыш может гарантировать себе игрок 1, применяя свои чистые стратегии при всевозможных действиях игрока 2.

Игрок 2 при оптимальном своём поведении должен стремится по возможности за счёт своих стратегий максимально уменьшить выигрыш игрока 1. Поэтому для игрока 2 отыскивается

$$\max_i a_{ij}$$

т.е. определяется max выигрыш игрока 1, при условии, что игрок 2 применит свою j-ю чистую стратегию, затем игрок 2 отыскивает такую свою $j = j_1$ стратегию, при которой игрок 1 получит min выигрыш, т.е. находит

$$\min_j \max_i a_{ij} = a_{i_1 j_1} = \overline{\alpha} \qquad (2).$$

<u>*Определение*</u>. Число $\overline{\alpha}$, определяемое по формуле (2), называется <u>*чистой верхней ценой игры*</u> и показывает, какой максимальный выигрыш за счёт своих стратегий может себе гарантировать игрок 1.

Другими словами, применяя свои чистые стратегии игрок 1 может обеспечить себе выигрыш не меньше $\underline{\alpha}$, а игрок 2 за счёт применения своих чистых стратегий может не допустить выигрыш игрока 1 больше, чем $\overline{\alpha}$.

<u>*Определение*</u>. Если в игре с матрицей А $\underline{\alpha} = \overline{\alpha}$, то говорят, что эта игра имеет <u>*седловую точку*</u> в чистых стратегиях и <u>*чистую цену*</u> игры

$$\upsilon = \underline{\alpha} = \overline{\alpha}.$$

<u>*Седловая точка*</u> – это пара чистых стратегий (i_o, j_o) соответственно игроков 1 и 2, при которых достигается равенство $\underline{\alpha} = \overline{\alpha}$. В это понятие вложен следующий смысл: если один из игроков придерживается стратегии, соответствующей седловой точке, то другой игрок не сможет поступить лучше, чем придерживаться стратегии, соответствующей седловой точке. Математически это можно записать и иначе:

$$a_{ij_o} \le a_{i_o j_o} \le a_{i_o j} \qquad (3)$$

где i, j – любые чистые стратегии соответственно игроков 1 и 2; (i_o, j_o) – стратегии, образующие седловую точку.

Таким образом, исходя из (3), седловой элемент $a_{i_o j_o}$ является минимальным в i_o-й строке и максимальным в j_o-м столбце в матрице А. Отыскание седловой точки матрицы А происходит следующим образом: в матрице А последовательно в каждой <u>строке</u> находят минимальный элемент и проверяют, является ли этот элемент максимальным в своём <u>столбце</u>. Если да, то он и есть седловой элемент, а пара стратегий, ему соответствующая, образует седловую точку. Пара чистых стратегий (i_o, j_o) игроков 1 и 2, образующая седловую точку и седловой элемент $a_{i_o j_o}$, называется <u>*решением игры*</u>. При этом i_o и j_o называются <u>*оптимальными чистыми стратегиями*</u> соответственно игроков 1 и 2.

<u>Пример 1</u>

$$
\begin{array}{rcl}
 & & \min\limits_{j} a_{ij} \\
 & & \parallel \\
A = & \begin{pmatrix} 1 & -3 & -2 \\ 0 & 5 & 4 \\ 2 & 3 & 2 \end{pmatrix} & \left.\begin{matrix} -3 \\ 0 \\ 2 \end{matrix}\right\} \max\limits_{i} \min\limits_{j} a_{ij} = 2 \\
\max\limits_{i} a_{ij} = & \underbrace{\begin{matrix} 2 & 5 & 4 \end{matrix}} & \\
 & \min\limits_{j} \max\limits_{i} a_{ij} = 2 &
\end{array}
$$

Седловой точкой является пара ($i_o = 3$; $j_o = 1$), при которой $\upsilon = \underline{\alpha} = \overline{\alpha} = 2$. Заметим, что хотя выигрыш в ситуации (3;3) также равен $2 = \underline{\alpha} = \overline{\alpha}$, она не является седловой точкой, т.к. этот выигрыш не является максимальным среди выигрышей третьего столбца.

<u>Пример 2</u>

$$
\begin{array}{rcccl}
 & & & \min\limits_{j} a_{ij} & \\
H = & \begin{pmatrix} 10 & 30 \\ 40 & 20 \end{pmatrix} & \begin{matrix} \to \\ \to \end{matrix} & \left.\begin{matrix} 10 \\ 20 \end{matrix}\right\} & \max\limits_{i} \min\limits_{j} a_{ij} = 20 \\
\max\limits_{i} a_{ij} & \begin{matrix} \downarrow & \downarrow \end{matrix} & & & \\
 & \underbrace{\begin{matrix} 40 & 30 \end{matrix}} & & & \\
 & \min\limits_{j} \max\limits_{i} a_{ij} = 30 & & &
\end{array}
$$

Из анализа матрицы выигрышей видно, что $\underline{\alpha} < \overline{\alpha}$, т.е. данная матрица не имеет седловой точки. Если игрок 1 выбирает свою чистую максиминную стратегию $i = 2$, то игрок 2, выбрав свою минимаксную $j = 2$, проиграет только 20. В этом случае игроку 1 выгодно выбрать стратегию i = 1, т.е. отклониться от своей чистой максиминной стратегии и выиграть 30. Тогда игроку 2 будет выгодно выбрать стратегию j = 1, т.е. отклониться от своей чистой минимаксной стратегии и проиграть 10. В свою очередь игрок 1 должен выбрать свою 2-ю стратегию, чтобы выиграть 40, а игрок 2 ответит выбором 2-й стратегии и т.д.

Другая особенность применения методов теории игр заключается в выборе решений, получаемых на основе анализа конфликтной ситуации. В теории игр доказывается теорема о том, что оптимальная стратегия для каждого из игроков является оптимальной и для другого. Так, если решение игры получено в чистых стратегиях (имеется седловая точка), то выбор решения однозначен. Например, если для парной антагонистической игры 3x4 составить матрицу, где элементами u_{ij} будут выигрыши (проигрыши) игроков, то седловая точка находится на пересечении максимина строк и минимакса столбцов

Стратегии	Стратегии B				min
A	1	2	3	4	строк
1	8	2	9	5	2

2	6	5	7	18	***5***
3	7	3	-4	10	-4
max столбцов	8	***5***	9	18	

Оптимальными стратегиями будут для А - 2, для В - 2. Цена игры равна 5. Отметим, что в случае наличия седловой точки ни один из игроков не может улучшить стратегию и стратегии называются *чистыми*. Отметим, что игра с чистыми стратегиями может существовать только при наличии полной информации о действиях противника.

Если же решение игры получено в смешанных стратегиях, то это эквивалентно созданию множества вариантов проектируемого компонента и использованию их с оптимальными частотам, соответствующими оптимальной смешанной стратегии. В случаях, когда не имеется полной информации о действиях противника, вводятся вероятности применения той или иной стратегии в виде векторов

$P_{<n>}=<p_1, p_2, ..., p_n>$ - для игрока А, где $\sum_{i=1}^{n} p_i = 1$;

$Q_{<m>}=<q_1, q_2, ..., q_n>$ - для игрока В, где $\sum_{i=1}^{m} q_i = 1$.

При этом игрок А выбирает стратегию в соответствии с принципом максимина по выражению:

$$\max_{p_i}\{\min(\sum_{i=1}^{n} a_{i1} p_i, \sum_{i=1}^{n} a_{i2} \dots \sum_{i=1}^{n} a_{im} p_i)\},$$

а игра В по принципу минимакса

$$\min_{p_i}\{\max(\sum_{j=1}^{m} a_{j1} q_i, \sum_{j=1}^{m} a_{j2} q_j \dots \sum_{j=1}^{m} a_{jn} q_j)\}.$$

Рассмотрим пример: пусть рассматривается принятие решения в игре 2х2, где игрок А знает вероятность стратегии 1, то есть p_1, тогда очевидно вероятность стратегии 2 будет 1-р, соответственно стратегии игрока В будут q_1 и $1-q_1$. Платежная матрица будет иметь вид:

		B	
		q_1	$1-q_1$
A	p_1	a_{11}	a_{12}
	$1-p_1$	a_{21}	a_{22}

На основании матрицы и приведенных выше выражений составляется таблица:

Чистые стратегии игрока В	Ожидаемые выигрыши игрока А
1	$(a_{11}-a_{21})p_1 + a_{21}$
2	$(a_{12}-a_{22})p_1 + a_{22}$

Из таблицы видно, что ожидаемый выигрыш игрока А линейно зависит от вероятности p_1 (в данном случае задача может быть решена графоаналитически). Тогда смешанная стратегия игрока А будет иметь вид

$$<p^*_1, p^*_2>,$$

то есть игроку А выгодно применять стратегию 1 с частотой (вероятностью) - p_1, а стратегию 2 с частотой p_2.

Очевидно, что разработка нескольких вариантов изделия сопряжена с большими затратами, не всегда реализуема и затрудняет использование системы. Поэтому при получении решения в смешанных стратегиях рекомендуются следующие случаи принятия окончательного решения [20,26,28]:

- для дальнейшего проектирования выбирается тот вариант, который гарантирует максимальное качество (выбор по максиминной стратегии аналогично критерию Вальда);
- выбирается тот вариант, который в смешанной стратегии должен использоваться с максимальной вероятностью;
- реализуется несколько вариантов изделия с частотами, соответствующими смешанной стратегии (создание адаптивно-модульных конструкций).

Важное значение в задачах исследования качества адаптивных систем имеет не только решение игры, но и анализ платежной матрицы. Это особенно важно в тех случаях, когда решение в смешанных стратегиях не реализуется. Этот анализ может проводиться на основе: оценки возможных потерь эффективности в случае реализации чистой стратегии; определения дополнительных затрат на их компенсацию с помощью "гибких" конструкторских решений; оценки достоверности рассмотренных стратегий противодействия; определения возможности реализации компромиссных вариантов и т.д.

Для анализа конфликтной ситуации требуется на основе математической модели операции построить платежную матрицу $[W_{mn}] =[W_{ij}]$, где W_{ij} характеризует качество изделия при выборе i-го варианта проектируемого изделия и при j-м варианте противодействия противника.

Решение может быть получено в чистых стратегиях, когда есть седловая точка. Условие седловой точки имеет вид

$$\max_i \min_j W_{ij} = \min_j \max_i W_{ij}, \quad (6.21)$$

где левая часть выражения - нижняя цена игры, правая - верхняя цена игры. Если условие (6.8) не выполняется, то седловая точка отсутствует и требуется реализация смешанной стратегии.

Решение в смешанных стратегиях состоит в реализации чистых стратегий с различными вероятностями, задаваемыми распределением:

для проектируемого изделия в виде вектора-столбца

$$G = \{g_i\}, \text{ где } i = 1,2 \ldots m; \ \sum_{i=1}^{m} g_i = 1;$$

для противодействия в виде вектора-строки

$$F = \{f_j\}, \text{ где } j = 1,2 \ldots n; \ \sum_{j=1}^{n} f_j = 1,$$

где

g_i - вероятность выбора стратегии u_i;

f_j - вероятность выбора стратегии v_j.

Платежную функцию запишем в следующем виде:

$$W(G,F) = G^T W_{ij} F^T = \sum_{i=1}^{m} \sum_{j=1}^{n} W_{ij} g_i f_j . \qquad (6.22)$$

где индексом "т" обозначена процедура транспонирования.

Платежная функция W(G,F) всегда имеет седловую точку, т.е. всегда существует решение матричной игры. Это утверждение соответствует основной теореме теории матричных игр [26]: каждая матричная игра с нулевой суммой имеет, по крайней мере, одно решение в чистых или смешанных стратегиях.

Последовательность решения игры следующая:

1. Анализируется платежная матрица на предмет исключения заведомо невыгодных и дублирующих стратегий.
2. Проверяется наличие седловой точки по условию (6.21).
3. Если решение в чистых стратегиях отсутствует, то ищется решение в смешанных стратегиях с помощью методов линейного программирования или методом Монте-Карло.

замена	2	6	5	17	18	7	18
(стра-	3	7	3	14	10	8	14
тегии	4	4	6	16	9	19	19
эксплуа-	5	12	4	15	8	10	15
тации)	min столбца	6	2	***9***	5	6	

Далее исследуем вопрос о нахождении значения игры и оптимальных стратегий игроков в случае отсутствия седловой точки.

2.2.3. Смешанные стратегии

Как мы отмечали, не все игры двух лиц с нулевой суммой имеют седловую точку. Рассмотрим, каким образом строятся оптимальные стратегии в этом случае. В качестве примера приведем следующую игру.

Пример (игра «четное-нечетное»). Два игрока одновременно выбрасывают один или два пальца. Если сумма выброшенных пальцев оказывается четной, то первый игрок получает от второго 1$. Если же количество выброшенных пальцев будет нечетным, то первый игрок платит второму 1$. Рассмотрим платежную матрицу данной игры (табл. 5).

Таблица 5

	Стратегия	Стратегия 2	Минимум
Стратегия 1	**-1**	**+ 1**	**-1**
Стратегия 2	**+ 1**	**-1**	**-1**
Максимум по	**+1**	**+1**	

Как было показано, данная игра не имеет седловой точки, поскольку не выполняется ранее рассмотренное соотношение. В этом случае неясно, как находить значение игры и как определять оптимальные стратегии.

Заметим, что для любой пары стратегий одному из игроков будет выгодно отклоняться от своей стратегии при условии, что другой сохраняет свою стратегию.

Для того чтобы ввести понятие оптимальных стратегий в случае игры двух лиц без седловой точки, необходимо обобщить понятие стратегии, введя понятие случайной стратегии. До этого под оптимальной стратегией мы понимали некоторую стратегию, которую игрок применяет каждый раз, когда реализуется данная игра. Почему бы не предоставить игрокам возможность выбирать свои стратегии некоторым случайным образом в соответствии с некоторыми вероятностями. Пусть: x_1 — вероятность, с которой первый игрок выбрасывает один палец; x_2 — вероятность, с которой первый игрок выбрасывает два пальца; y_1— вероятность, с которой второй игрок выбрасывает один палец; y_2 — вероятность, с которой второй игрок выбрасывает два пальца.

Если x_1 и x_2 таковы, что x_1 > 0, x_2 >: 0 и $x1+ x_2$ = 1, то $(x_1\ x_2)$ — случайная, или смешанная, стратегия первого игрока. Аналогично $(y_1 y_2)$ — случайная, или смешанная, стратегия второго игрока, если

$$y_1 > 0,\ y_2>0,\ \ y_1 + y_2=1.$$

В общем случае смешанные стратегии есть векторы $(x_1, x_2, ..., x_m)$ и $(\boldsymbol{У_1, У_2}, ..., \boldsymbol{y_n})$

Если смешанная стратегия $(x_1\ x_2, ..., x_m)$ такова, что одно из значений x,= 1, а следовательно, остальные равны нулю, то такая стратегия называется **чистой стратегией** игрока. Чистая стратегия уже не является случайной. Вспоминая игру, показанную ранее, можно записать оптимальную стратегию первого игрока, как x*= (0,0,1), а второго, как y* = (0,1,0).

Подход к определению решения игры при смешанных стратегиях также основывается на критерии минимакса. Единственная разница заключается в том, что первый игрок выбирает x_i- так, чтобы максимизировать наименьший ожидаемый выигрыш по столбцам, тогда как второй игрок выбирает y_i с целью минимизировать наибольший ожидаемый проигрыш по строкам.

Эти величины определяются соответственно как среднеожидаемые максиминные и среднеожидаемые минимаксные платежи.

Как и в случае чистых стратегий, выполняется соотношение:

минимаксный ожидаемый проигрыш > максиминный ожидаемый выигрыш.

Когда х,- и y,- соответствуют оптимальным решениям, выполняется

Справедлива следующая основная теорема теории матричных игр с нулевой суммой (теорема фон Неймана).

Теорема. Каждая конечная игра с нулевой суммой имеет по крайней мере одно оптимальное решение среди смешанных стратегий.

Определение. Если чистая стратегия входит в смешанную стратегию с отличной от нуля вероятностью, то она называется активной.

Справедлива теорема об активных стратегиях.

Теорема . Если один из игроков придерживается своей оптимальной смешанной стратегии, то выигрыш остается неизменным и равным цене игры v, если второй игрок не выходит за пределы своих активных стратегий.

Эта теорема имеет большое практическое значение — она дает конкретный способ нахождения оптимальных стратегий при отсутствии седловой точки.

Для игры 2x2 любая чистая стратегия противника является активной, если отсутствует седловая точка

Доминируемые стратегии могут быть исключены, при этом цена игры не изменяется.

Доминирование можно распространить и на смешанные стратегии.

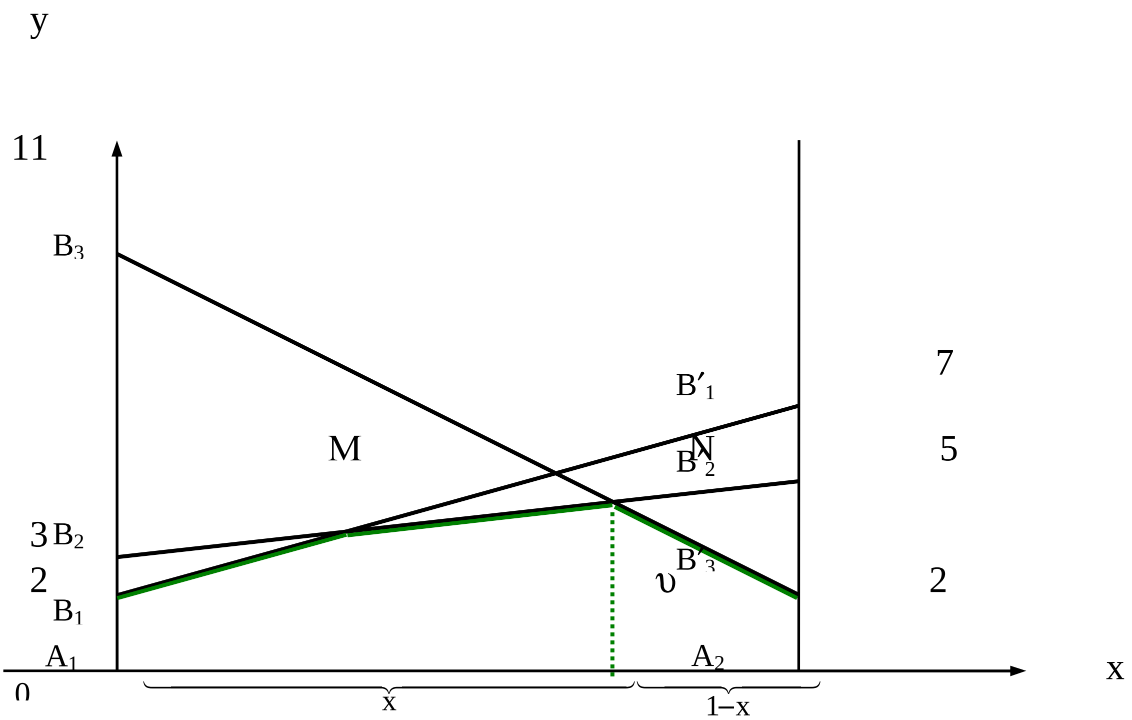
2.2.4. Графический метод решения игр 2 x n и m x 2.

Поясним метод на примерах.

<u>Пример 1</u>.
Рассмотрим игру, заданную платёжной матрицей.

$$\begin{array}{cc} & 2 \\ & \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ 1 \; \begin{array}{c} A_1 \\ A_2 \end{array} & \begin{pmatrix} 2 & 3 & 11 \\ 7 & 5 & 2 \end{pmatrix} \end{array}$$

На плоскости xOy введём систему координат и на оси Ox отложим отрезок единичной длины A_1, A_2, каждой точке которого поставим в соответствие некоторую смешанную стратегию игрока 1 (x, 1 – x). В частности, точке A_1 (0;0) отвечает стратегия A_1, точке A_2 (1;0) – стратегия A_2 и т.д.



В точках A_1 и A_2 восстановим перпендикуляр и на полученных прямых будем откладывать выигрыш игроков. На первом перпендикуляре (в данном случае он совпадает с осью *0y*) отложим выигрыш игрока 1 при стратегии A_1, а на втором – при стратегии A_2. Если игрок 1 применит стратегию A_1, то выиграет при стратегии B_1 игрока 2 – 2, при стратегии B_2 – 3, а при стратегии B_3 – 11. Числам 2, 3, 11 на оси *0x* соответствуют точки B_1, B_2 и B_3.
Если же игрок 1 применит стратегию A_2, то его выигрыш при стратегии B_1 равен 7, при B_2 – 5, а при B_3 – 2. Эти числа определяют точки B'_1, B_2', B_3' на перпендикуляре, восстановленном в точке A_2.Соединяя между собой точки B_1 и B'_1, B_2 и B'_2, B_3 и B'_3 получим три прямые, расстояние до которых от оси *0x* определяет средний выигрыш при любом сочетании соответствующих стратегий. Например, расстояние от любой точки отрезка $B_1B'_1$ до оси *0x* определяет сред-

ний выигрыш υ_1 при любом сочетании стратегий A_1 A_2 (с частотами x и $1–x$) и стратегией B_1 игрока 2. Это расстояние равно

$$2x_1 + 6(1 - x_2) = \upsilon_1$$

(Вспомните планиметрию и рассмотрите трапецию A_1 B_1 B'_1 A_2). Таким образом, ординаты точек, принадлежащих ломанной B_1 M N B'_3 определяют минимальный выигрыш игрока 1 при применении им любых смешанных стратегий. Эта минимальная величина является максимальной в точке N; следовательно этой точке соответствует оптимальная стратегия $X^* = (x, 1–x)$, а её ордината равна цене игры υ. Координаты точки N находим как точку пересечения прямых B_2 B'_2 и B_3 B'_3.

Соответствующие два уравнения имеют вид

$$\begin{cases} 3x + 5(1 - x) = \upsilon \\ 11x + 2(1 - x) = \upsilon \end{cases} \Rightarrow x = \frac{3}{11},\ \upsilon = \frac{49}{11}.$$

Следовательно $X = (\frac{3}{11}; \frac{9}{11})$, при цене игры $\upsilon = \frac{49}{11}$. Таким образом мы можем найти оптимальную стратегию при помощи матрицы

$$\begin{pmatrix} 3 & 11 \\ 5 & 2 \end{pmatrix}$$

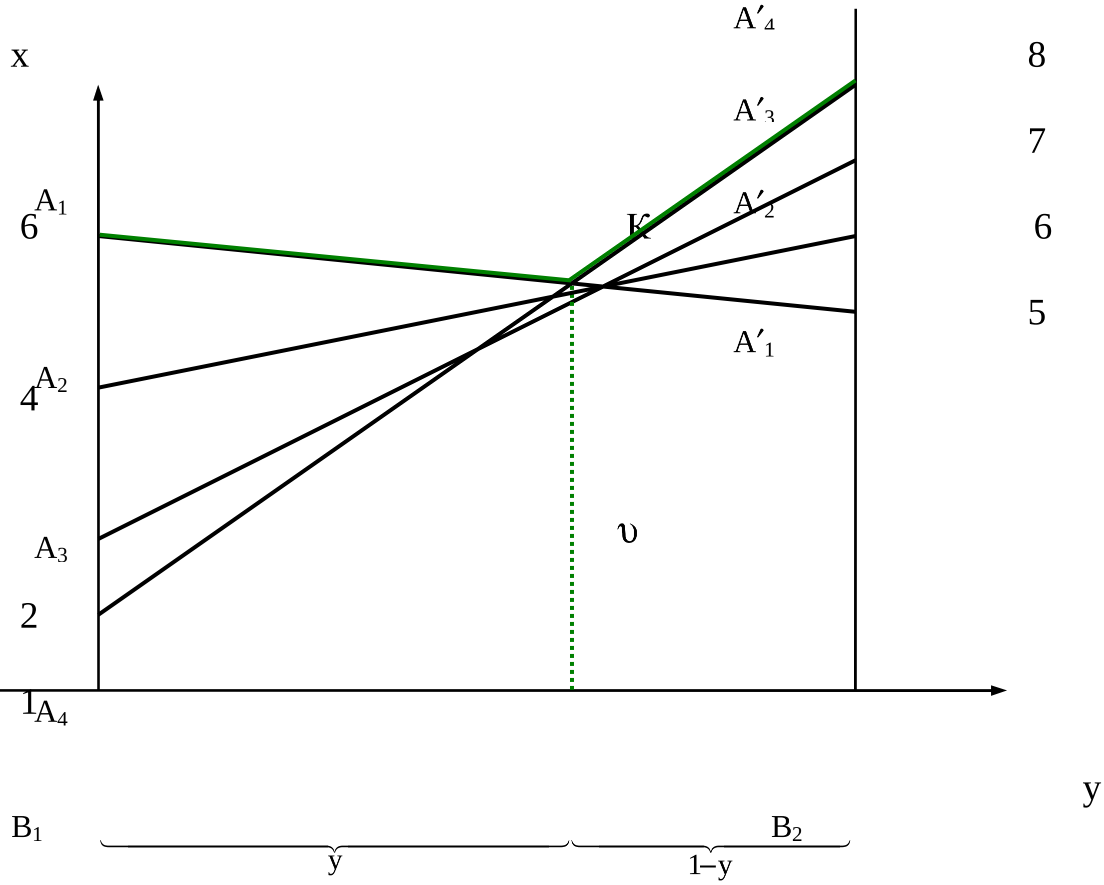
Оптимальные стратегии для игрока 2 можно найти из системы

$$\begin{cases} 3y + 5(1 - y) = \upsilon \\ 5y + 2(1 - y) = \upsilon \end{cases} \Rightarrow y = \frac{9}{11}$$

и, следовательно, $Y = (0; \frac{9}{11}; \frac{2}{11})$. (Из рисунка видно, что стратегия B_1 не войдёт в оптимальную стратегию.

Пример 2. Найти решение игры, заданной матрицей

$$\begin{array}{cc} & 2 \\ & \begin{array}{cc} B_1 & B_2 \end{array} \\ 1 & \begin{array}{c} A_1 \\ A_2 \\ A_3 \\ A_4 \end{array}\begin{pmatrix} 6 & 5 \\ 4 & 6 \\ 2 & 7 \\ 1 & 8 \end{pmatrix} \end{array}$$



Решение. Матрица имеет размерность 2 х 4. Строим прямые, соответствующие стратегиям игрока 1. Ломанная A_1 K A'_4 соответствует верхней границе выигрыша игрока 1, а отрезок N K –цене игры. Решение игры таково

$$Y = \left(\frac{3}{8};\ \frac{5}{8}\right); \quad X = \left(\frac{7}{8};\ 0;\ 0;\ \frac{1}{8}\right); \quad \upsilon = \frac{43}{8}.$$

2.3. Выбор оптимальной стратегии в условиях неопределенности (игры с природой)

2.3.1. Специфика ситуации полной неопределенности. Мы предполагали, что все участники игры имеют свои интересы, которые выражаются либо платежными матрицами (антагонистические игры, би-матричные игры), либо платежными функциями (игры n лиц). Однако так бывает далеко не всегда. Ситуации, при которой нам либо ничего не известно об интересах второй стороны (или сторон), либо эти интересы действительно отсутствуют (второй игрок — «природа»), характеризуются как ситуации принятия решений в условиях полной неопределенности (или игры с «природой»). Естественно, что термин «природа» употребляется здесь в некотором символическом смысле как обозначение некой действительности, мотивы проявления которой нам неизвестны.

Как мы отмечали, теория игр — это математическая дисциплина, исследующая ситуации, в которых принятие решений зависит от нескольких участников. Поэтому тот факт, что в рассматриваемой ситуации вторая сторона не имеет, с нашей точки, зрения каких-либо интересов, несколько меняет и наш подход к выбору своей оптимальной стратегии. То есть разумно рассмотреть несколько иные критерии, чем, например, принцип минимакса для антагонистической игры (игры с нулевой суммой) двух лиц.

Следует отметить и еще одно отличие. При применении принципа доминирования стратегий мы уже не можем производить исключение стратегий второго игрока («природы»), поскольку не имеем информации о его интересах, а следовательно, никаких разумных оснований для исключения его стратегий.

2.3.2. Критерии выбора оптимальной стратегии.

1. *Максиминный критерий Вальда.* Это тот самый критерий, который использовался при рассмотрении игр с нулевой суммой (антагонистических игр). Он отражает «принцип гарантированного результата», то есть мы откладываемся на самый неблагоприятный для нас случай и пытаемся выбрать такую стратегию, которая максимизировала бы наш выигрыш в самой неблагоприятной для нас ситуации. В качестве оптимальной выбирается стратегия, на которой достигается значение max. Иногда этот критерий называют критерием «крайнего пессимизма».

2. *Критерий максимакса.* Этот критерий является в определенном смысле противоположным по своему смыслу предыдущему критерию. А именно, он предполагает рассмотрение не самого для нас неблагоприятного случая (критерий Вальда), а наоборот наиболее благоприятного. Выбирается в качестве оптимальной такая стратегия, для которой этот самый благоприятный случай дает самый большой выигрыш. В качестве оптимальной стратегии выбирается стратегия, на которой достигается значение max. Иногда этот критерий называют критерием «крайнего оптимизма»

3. *Критерий Гурвица.* Этот критерий является своего рода обобщением двух предыдущих критериев. Он представляет из себя целое семейство критериев, зависящих от некоторого параметра a, смысл которого—в определении баланса между подходами «крайнего пессимизма» и «крайнего оптимизма». В качестве оптимальной стратегии выбирается стратегия, на которой достигается значение max. Значение параметра выбирается из интервала $0 < a < 1$. Критерий Вальда получается как частный случай при $a = 0$, а критерий максимакса при $a = 1$. Выбор

конкретного значения параметра определяется скорее субъективными факторами, например склонностью к риску ЛПР (лица принимающего решение). При отсутствии каких-либо явных предпочтений вполне логично, например, выбрать значение a = 0,5.

4. *Критерий Сэвиджа {критерий минимаксного риска)*. Применение данного критерия предполагает рассмотрение некоторой производной матрицы, смысл которой состоит в том, что для каждой стратегии второго игрока определяется выигрыш в наиболее благоприятном случае (при наиболее правильном выборе стратегии первым игроком для данной ситуации), а далее вычисляются величины «недополученных» выигрышей для всех остальных стратегий первого игрока при рассматриваемой стратегии второго игрока. Далее к матрице рисков применяется минимаксный подход.

В качестве оптимальной стратегии выбирается стратегия, на которой достигается min. Тем самым мы выбираем такую стратегию, для которой наибольшее значение «недополучения» будет иметь наименьшее значение.

5. *Критерий Лапласа*. Этот критерий исходит из следующего соображения. Поскольку нам ничего не известно о принципах или вероятностях применения вторым игроком своих стратегий, то мы предполагая эти вероятности все равными. Таким образом, смысл данного критерия — максимизация ожидаемого выигрыша в предположении о равновероятности применения вторым игроком своих стратегий.

2.4. Выбор стратегии при наличии вероятностной информации

В отличие от ситуации полной неопределенности весьма частой является ситуация, когда в распоряжении первого игрока есть информация о вероятностях применения стратегий второй стороной. Эти вероятности называются априорными вероятностями. Выбираем критерий Лапласа, с учетом вероятностей применения стратегий второй стороной.

Тема 3. Марковские процессы

Очень удобно описывать появление случайных событий в виде вероятностей переходов из одного состояния системы в другое, так как при этом считается, что, перейдя в одно из состояний, система не должна далее учитывать обстоятельства того, как она попала в это состояние.

Случайный процесс называется **марковским процессом** (или **процессом без последействия**), если для каждого момента времени t вероятность любого состояния системы в будущем зависит только от ее состояния в настоящем и не зависит от того, как система пришла в это состояние.

Итак, марковский процесс удобно задавать графом переходов из состояния в состояние. Мы рассмотрим два варианта описания марковских процессов — **с дискретным и непрерывным временем**.

В первом случае переход из одного состояния в другое происходит в заранее известные моменты времени — такты (1, 2, 3, 4, …). Переход осуществляется на каждом такте, то есть исследователя интересует только последовательность состояний, которую проходит случайный процесс в своем развитии, и не интересует, когда конкретно происходил каждый из переходов.

Во втором случае исследователя интересует и цепочка меняющих друг друга состояний, и моменты времени, в которые происходили такие переходы.

И еще. Если вероятность перехода не зависит от времени, то марковскую цепь называют **однородной**.

3.1 Марковский процесс с дискретным временем

Итак, модель марковского процесса представим в виде графа, в котором состояния (вершины) связаны между собой связями (переходами из i-го состояния в j-е состояние), см. **рис. 3.1**.

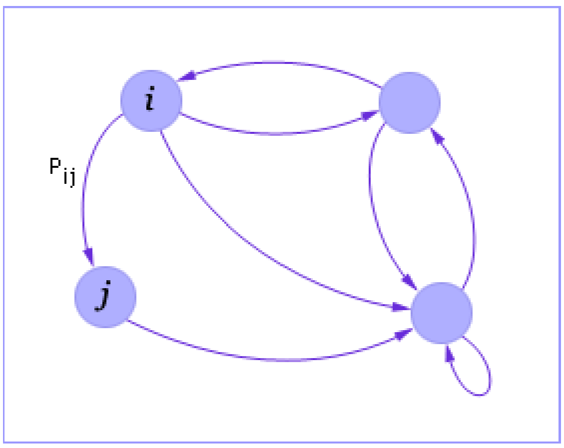


Рис. 3.1. Пример графа переходов

Каждый переход характеризуется **вероятностью перехода** P_{ij}. Вероятность P_{ij} показывает, как часто после попадания в i-е состояние осуществляется затем переход в j-е состояние. Конечно, такие переходы происходят случайно, но если измерить частоту переходов за достаточно большое время, то окажется, что эта частота будет совпадать с заданной вероятностью перехода.

Ясно, что у каждого состояния сумма вероятностей всех переходов (исходящих стрелок) из него в другие состояния должна быть всегда равна 1 (см. **рис. 3.2**).

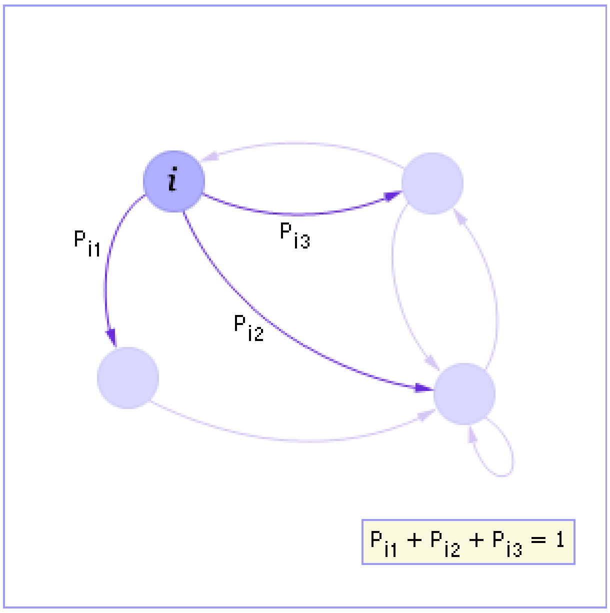


Рис. 3.2. Фрагмент графа переходов (переходы из i-го состояния являются полной группой случайных событий)

Например, полностью граф может выглядеть так, как показано на **рис. 33.3**.

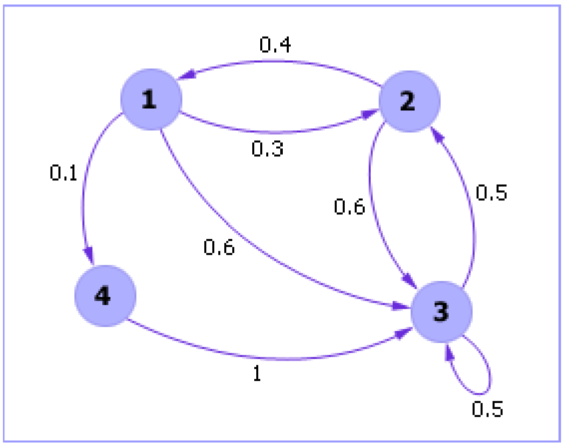


Рис. 3.3. Пример марковского графа переходов

Реализация марковского процесса (процесс его моделирования) представляет собой вычисление последовательности (цепи) переходов из состояния в состояние (см. **рис. 3.4**). Цепь на **рис. 3.4** является случайной последовательностью и может иметь также и другие варианты реализации.

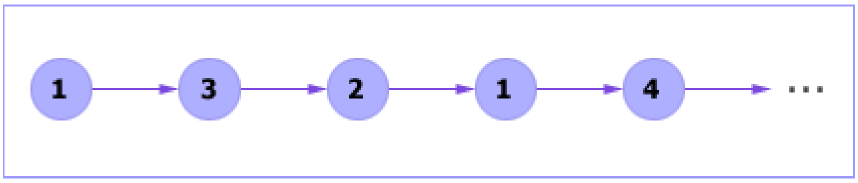


Рис. 3.4. Пример марковской цепи, смоделированной по марковскому графу, изображенному на рис. 33.3

Чтобы определить, в какое новое состояние перейдет процесс из текущего i-го состояния, достаточно разбить интервал $[0;\, 1]$ на подынтервалы величиной P_{i1}, P_{i2}, P_{i3}, … ($P_{i1} + P_{i2} + P_{i3} + \ldots = 1$), см. **рис. 33.5**. Далее с помощью ГСЧ надо получить очередное равномерно распределенное в интервале $[0;\, 1]$ случайное число r_{pp} и определить, в какой из интервалов оно попадает (см. лекцию 23).

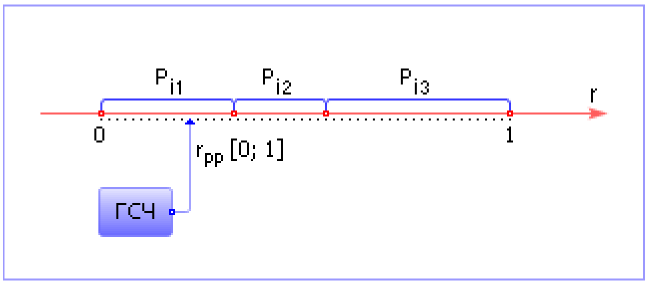


Рис. 3.5. Процесс моделирования перехода из i-го состояния марковской цепи в j-е с использованием генератора случайных чисел

После этого осуществляется переход в состояние, определенное ГСЧ, и повтор описанной процедуры для нового состояния. Результатом работы модели является марковская цепь (см. рис. 3.4).

Пример. Имитация стрельбы из пушки по цели. Для того, чтобы проимитировать стрельбу из пушки по цели, построим модель марковского случайного процесса.

Определим следующие три состояния: S_0 — цель не повреждена; S_1 — цель повреждена; S_2 — цель разрушена. Зададим вектор начальных вероятностей:

	S_0	S_1	S_2
P_0	0.8	0.2	0

Значение P_0 для каждого из состояний показывает, какова вероятность каждого из состояний объекта до начала стрельбы.

Зададим матрицу перехода состояний (см. табл. 33.1).

Таблица 33.1.
Матрица вероятностей перехода
дискретного марковского процесса

	В S_0	В S_1	В S_2	**Сумма вероятностей переходов**
Из $\mathbf{S_0}$	0.45	0.40	0.15	0.45 + 0.40 + 0.15 = 1
Из $\mathbf{S_1}$	0	0.45	0.55	0 + 0.45 + 0.55 = 1
Из $\mathbf{S_2}$	0	0	1	0 + 0 + 1 = 1

Матрица задает вероятность перехода из каждого состояния в каждое. Заметим, что вероятности заданы так, что сумма вероятностей перехода из некоторого состояния в остальные всегда равна единице (куда-то система должна перейти обязательно).

Наглядно модель марковского процесса можно представить себе в виде следующего графа (см. **рис. 3.6**).

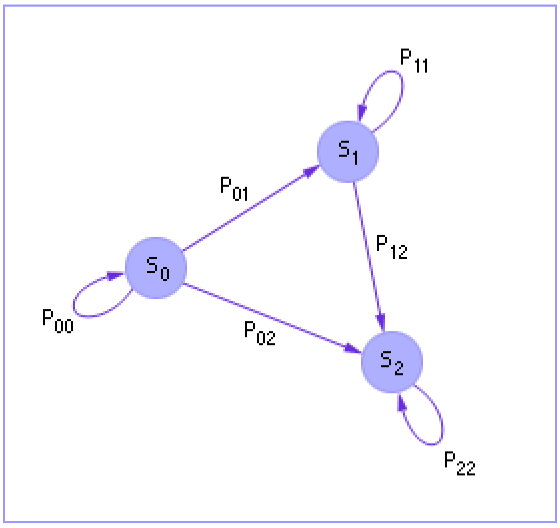


Рис. 3.6. Граф марковского процесса, моделирующий стрельбу из пушки по цели

Используя модель и метод статистического моделирования, попытаемся решить следующую задачу: определить среднее количество снарядов, необходимое для полного разрушения цели.

Проимитируем, используя таблицу случайных чисел, процесс стрельбы. Пусть начальное состояние будет S_0. Возьмем последовательность из таблицы случайных чисел: 0.31, 0.53, 0.23, 0.42, 0.63, 0.21, … (случайные числа можно взять, например, из этой таблицы).

0.31: цель находится в состоянии S_0 и остается в состоянии S_0, так как $0 < \mathbf{0.31} < 0.45$;
0.53: цель находится в состоянии S_0 и переходит в состояние S_1, так как $0.45 < \mathbf{0.53} < 0.45 + 0.40$;
0.23: цель находится в состоянии S_1 и остается в состоянии S_1, так как $0 < \mathbf{0.23} < 0.45$;
0.42: цель находится в состоянии S_1 и остается в состоянии S_1, так как $0 < \mathbf{0.42} < 0.45$;
0.63: цель находится в состоянии S_1 и переходит в состояние S_2, так как $0.45 < \mathbf{0.63} < 0.45 + 0.55$.

Так как достигнуто состояние S_2 (далее цель переходит из S_2 в состояние S_2 с вероятностью 1), то цель поражена. Для этого в данном эксперименте потребовалось 5 снарядов.

На **рис. 3.7** приведена временная диаграмма, которая получается во время описанного процесса моделирования. Диаграмма показывает, как во времени происходит процесс изменения состояний. Такт моделирования для данного случая имеет фиксированную величину. Нам важен сам факт перехода (в какое состояние переходит система) и не важно, когда это происходит.

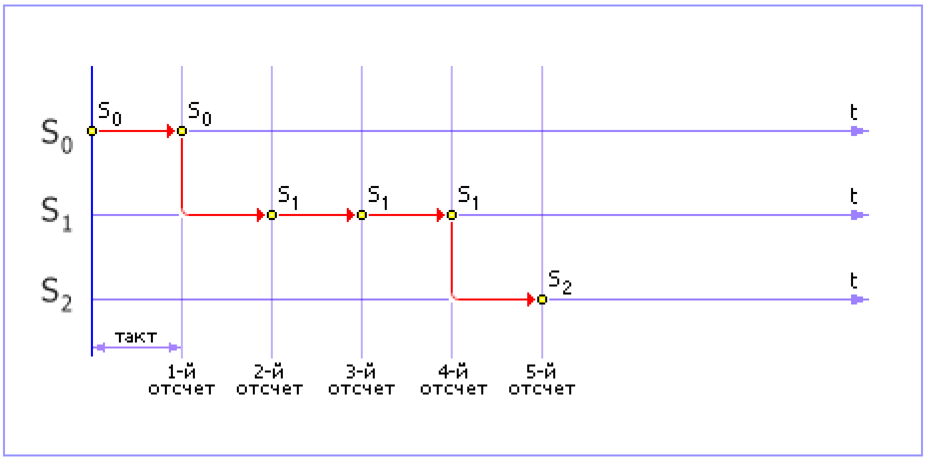


Рис. 3.7. Временная диаграмма переходов в марковском графе (пример имитации)

Процедура уничтожения цели совершена за 5 тактов, то есть марковская цепь этой реализации выглядит следующим образом: S_0—S_0—S_1—S_1—S_1—S_2. Конечно, ответом задачи это число быть не может, так как в разных реализациях получатся разные ответы. А ответ у задачи может быть только один.

Повторяя данную имитацию, можно получить, например, еще такие реализации (это зависит от того, какие конкретно случайные числа выпадут): 4 (S_0—S_0—

S_1—S_1—S_2); 11 (S_0—S_0—S_0—S_0—S_0—S_1—S_1—S_1—S_1—S_1—S_1—S_2); 5 (S_1—S_1—S_1—S_1—S_1—S_2); 6 (S_0—S_0—S_1—S_1—S_1—S_1—S_2); 4 (S_1—S_1—S_1—S_1—S_2); 6 (S_0—S_0—S_1—S_1—S_1—S_1—S_2); 5 (S_0—S_0—S_1—S_1—S_1—S_2). Всего уничтожено 8 целей. Среднее число циклов в процедуре стрельбы составило:
(5 + 4 + 11 + 5 + 6 + 4 + 6 + 5)/8 = 5.75 или, округляя, 6. Именно столько снарядов, в среднем, рекомендуется иметь в боевом запасе пушки для уничтожения цели при таких вероятностях попаданий.

Теперь следует определить точность. Именно точность может нам показать, насколько следует доверять данному ответу. Для этого проследим, как сходится последовательность случайных (приближенных) ответов к правильному (точному) результату. Напомним, что, согласно центральной предельной теореме (см. лекцию 25, лекцию 21), сумма случайных величин есть величина неслучайная, поэтому для получения статистически достоверного ответа необходимо следить за средним числом снарядов, получаемых в ряде случайных реализаций.

На первом этапе вычислений средний ответ составил 5 снарядов, на втором этапе средний ответ составил (5 + 4)/2 = 4.5 снаряда, на третьем —
(5 + 4 + 11)/3 = 6.7. Далее ряд средних величин, по мере накопления статистики, выглядит следующим образом: 6.3, 6.2, 5.8, 5.9, 5.8. Если изобразить этот ряд в виде графика средней величины выпущенных снарядов, необходимых для поражения цели, в зависимости от номера эксперимента, то обнаружится, что данный ряд сходится к некоторой величине, которая и является ответом (см. **рис. 3.8**).

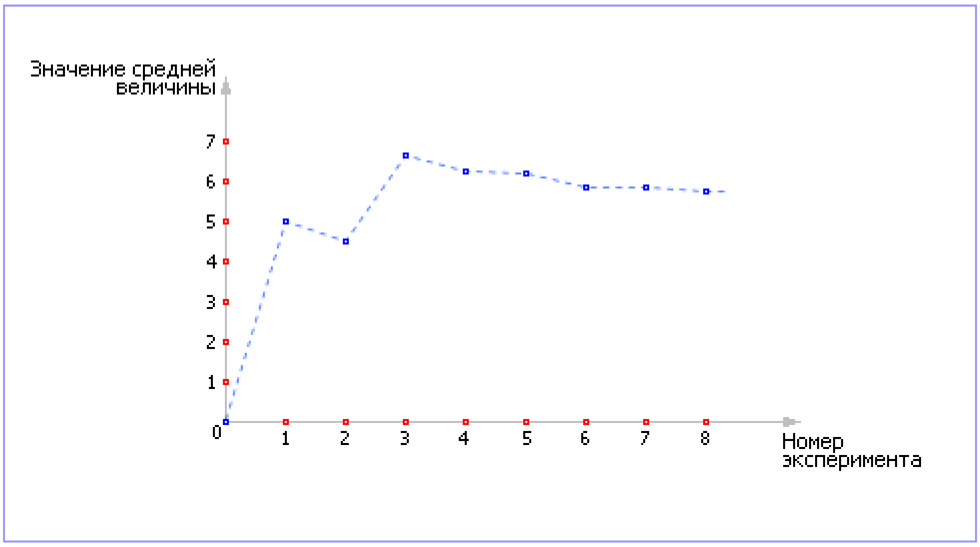


Рис. 3.8. Изменение средней величины в зависимости от номера эксперимента

Визуально мы можем наблюдать, что график «успокаивается», разброс между вычисляемой текущей величиной и ее теоретическим значением со временем уменьшается, стремясь к статистически точному результату. То есть в некоторый момент график входит в некоторую «трубку», размер которой и определяет точность ответа.

Еще раз заметим, что в вышерассмотренном случае нам безразлично, в какие моменты времени будет происходить переход. Переходы идут такт за тактом. Если важно указать, в какой именно момент времени произойдет переход, сколько времени система пробудет в каждом из состояний, требуется применить модель с непрерывным временем.

3.2 Марковские случайные процессы с непрерывным временем

Итак, снова модель марковского процесса представим в виде графа, в котором состояния (вершины) связаны между собой связями (переходами из i-го состояния в j-е состояние), см. **рис. 3.10**.

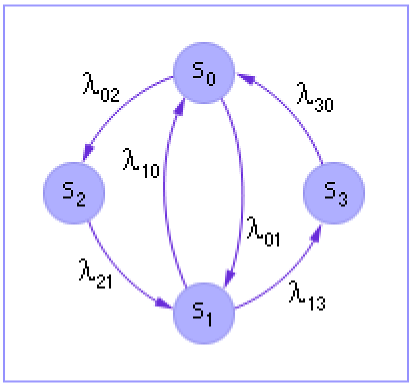


Рис. 3.10. Пример графа марковского процесса с непрерывным временем

Теперь каждый переход характеризуется плотностью вероятности перехода λ_{ij}. По определению:

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

При этом плотность понимают как распределение вероятности во времени.

Переход из *i*-го состояния в *j*-е происходит в случайные моменты времени, которые определяются интенсивностью перехода λ_{ij}.

К интенсивности переходов (здесь это понятие совпадает по смыслу с распределением плотности вероятности по времени t) переходят, когда процесс непрерывный, то есть, распределен во времени.

С интенсивностью потока (а переходы — это поток событий) мы уже научились работать в лекции 28. Зная интенсивность λ_{ij} появления событий, порождаемых потоком, можно сымитировать случайный интервал между двумя событиями в этом потоке.

$$\tau_{ij} = -\frac{1}{\lambda_{ij}} \cdot Ln(R)$$

где τ_{ij} — интервал времени между нахождением системы в i-ом и j-ом состоянии.

Далее, очевидно, система из любого i-го состояния может перейти в одно из нескольких состояний j, $j + 1$, $j + 2$, …, связанных с ним переходами λ_{ij}, λ_{ij+1}, λ_{ij+2}, ….

В j-е состояние она перейдет через τ_{ij}; в $(j + 1)$-е состояние она перейдет через τ_{ij+1}; в $(j + 2)$-е состояние она перейдет через τ_{ij+2} и т. д.

Ясно, что система может перейти из i-го состояния только в одно из этих состояний, причем в то, переход в которое наступит раньше.

Поэтому из последовательности времен: τ_{ij}, τ_{ij+1}, τ_{ij+2} и т. д. надо выбрать минимальное и определить индекс j, указывающий, в какое именно состояние произойдет переход.

Пример. Моделирование работы станка. Промоделируем работу станка (см. **рис. 3.10**), который может находиться в следующих состояниях: S_0 — станок исправен, свободен (простой); S_1 — станок исправен, занят (обработка); S_2 — станок исправен, замена инструмента (переналадка) $\lambda_{02} < \lambda_{21}$; S_3 — станок неисправен, идет ремонт $\lambda_{13} < \lambda_{30}$.

Зададим значения параметров λ, используя экспериментальные данные, получаемые в производственных условиях: λ_{01} — поток на обработку (без переналадки); λ_{10} — поток обслуживания; λ_{13} — поток отказов оборудования; λ_{30} — поток восстановлений.

Реализация будет иметь следующий вид (см. **рис. 3.11**).

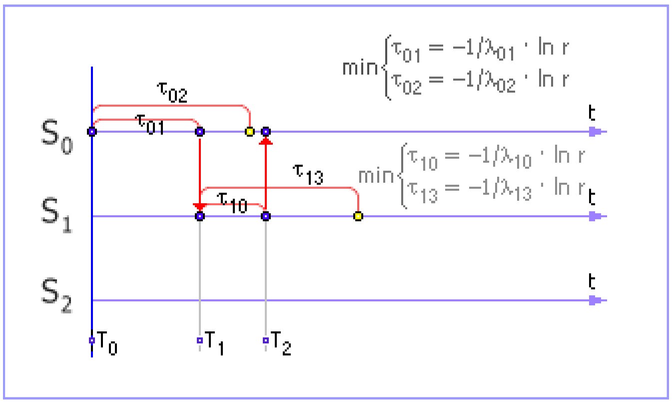


Рис. 3.11. Пример моделирования непрерывного марковского процесса с визуализацией на временной диаграмме (желтым цветом указаны запрещенные,

синим — реализовавшиеся состояния)

В частности, из **рис. 3.11** видно, что реализовавшаяся цепь выглядит так: S_0—S_1—S_0—… Переходы произошли в следующие моменты времени: T_0—T_1—T_2—T_3—…, где $T_0 = 0$, $T_1 = \tau_{01}$, $T_2 = \tau_{01} + \tau_{10}$.

Задача. Поскольку модель строят для того, чтобы на ней можно было решить задачу, ответ которой до этого был для нас совсем не очевиден (см. лекцию 01), то сформулируем такую задачу к данному примеру. Определить долю времени в течение суток, которую занимает простой станка (посчитать по рисунку) $T_{ср} = (T + T + T + T\ )/N$.

Алгоритм имитации будет иметь следующий вид (см. **рис. 3.12**).

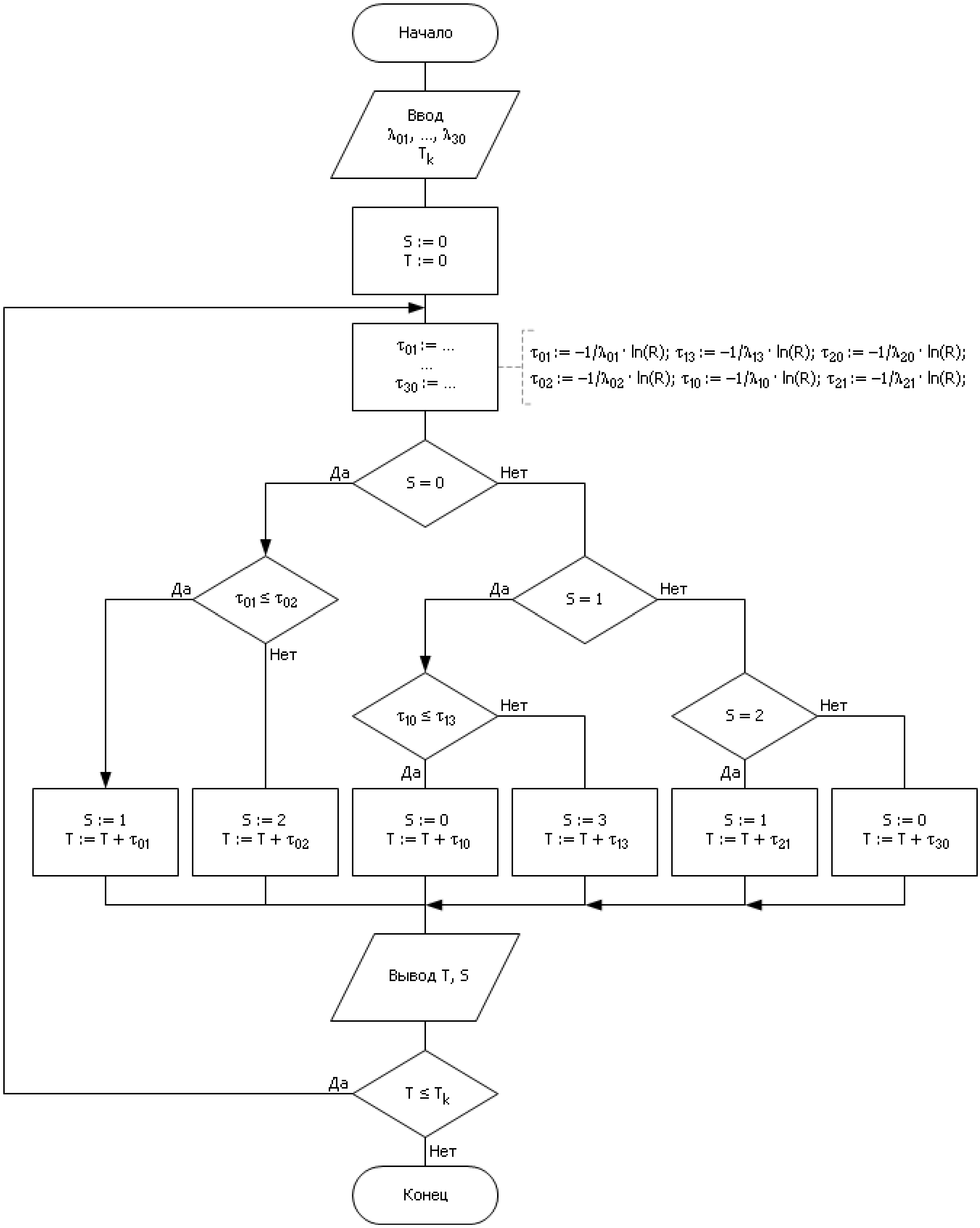


Рис. 3.12. Блок-схема алгоритма моделирования непрерывного марковского процесса на примере имитации работы станка

Очень часто аппарат марковских процессов используется при моделировании компьютерных игр, действий компьютерных героев.

3.3 Работа ремонтной бригады предприятия как СМО

3.3.1 ПОСТАНОВКА ЗАДАЧИ

На предприятии поддерживают процесс производства 2 конвейера, работа которых идет независимо друг о друга, а также имеется ремонтная бригада, осуществляющая по необходимос-ти ремонт оборудования. На рисунке 1 представлен граф, описывающий работу системы. Опре-делить экономически целесообразно ли ускорение ремонта путем использования второй ремонт-ной бригады при следующих данных:
$\lambda_{01} = 2$, $\lambda_{02} = 1$, $\lambda_{10} = 1$, $\lambda_{13} = 3$, $\lambda_{20} = 1$, $\lambda_{23} = 2$, $\lambda_{31} = 2$, $\lambda_{32} = 3$.

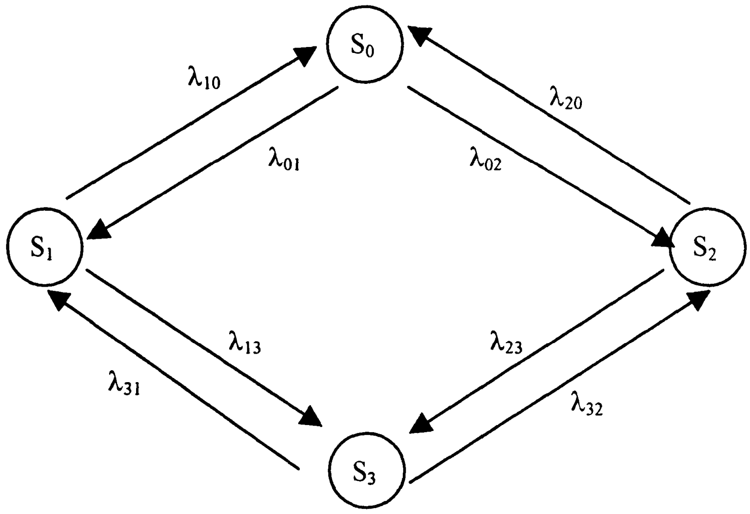


Рис.1. Ориентированный граф, описывающий работу системы.

Перечислим возможные состояния системы:

S_0 — оба конвейера исправны;

S_1— первый конвейер ремонтируется, второй исправен;

S_2— первый конвейер исправен, второй ремонтируется;

S_3 — оба узла ремонтируются.

Дуга, направленная из S_0 в S_1 означает переход в момент отказа первого узла; из S_1 в S_0 переход в момент окончания ремонта этого узла. На графе отсутствуют стрелки из S_0 в S_3 и из S_1 в S_2. Это объясняется тем, что выходы узлов из строя предполагаются независимыми друг от друга и, например, вероятностью одновременного выхода из строя двух узлов (переход от S_0 в S_3) или одновременного окончания ремонтов двух узлов (переход из S_3 в S_0) можно пренебречь.

3.3.2 ОПИСАНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Будем полагать что все переходы системы из состояния S_i в S_j происходят под воздейст-вием простейших потоков событий с интенсивностями состояний λ_{ij}($i, j = 0,1,2,3$).

Так, переход системы из состояния S_0 в состояние S_1 будет происходить под воздействием потока отказов первого узла, а обратный переход из состояния S_1 в S_0 — под воздействием потока и событий, связанных с окончанием ремонтов первого узла и т.п.

Рассматриваемая система S имеет четыре возможные состояния: S_0, S_1, S_2, S_3 . Назовем ве-роятностью i-го состояния вероятность $p_i(t)$ того, что в момент t система будет находиться в состоя-нии S_i. Очевидно, что для любого момента t сумма вероятностей всех состояний равна единице:

$$\sum_{i=0}^{3} p_i(t) = 1. \qquad (*1)$$

Рассмотрим систему в момент t и, задав малый интервал времени Δt, найдем вероятность $p_0(t + \Delta t)$ того, что система в момент $t + \Delta t$ будет находиться в состоянии S_0. Это достигается раз-ными способами.

1. Система в момент t с вероятностью $p_0(t)$, находясь в состоянии S_0 и за время Δt, не выш-ла из него. Вывести систему из этого состояния можно суммарным простейшим потоком с интен-сивностью $(\lambda_{01} + \lambda_{02})$, т.е. с вероятностью приближенно равной $(\lambda_{01} + \lambda_{02})\Delta t$ в соответствии с равенством $p_{\Delta t} = p(T < \Delta t) = 1 - e^{-\lambda\Delta t} \approx \lambda\Delta t.$ (*2) Вероятность же того, что система не выйдет из сос-тояния S_0, равна по теореме умножения вероятностей $p_0(t)\left[1-(\lambda_{01}+\lambda_{02})\Delta t\right]$.

2. Система в момент t с вероятностями $p_1(t)$ (или $p_2(t)$) находилась в состоянии S_1 или S_2 и за время Δt перешла в состояние S_0 . Потоком, интенсивностью λ_{10} (или λ_{20} как следует из рис.1) система перейдет в состояние S_0 с вероятностью приближенно равной $\lambda_{10}\Delta t$ (или $\lambda_{20}\Delta t$). Вероят-ность того, что система будет находиться в состоянии S_0 по этому способу, равна $p_1(t)\lambda_{10}\Delta t$ (или $p_2(t)\lambda_{10}\Delta t$).

Применяя теорему сложения вероятностей, получим

$$p_0(t+\Delta t) = p_1(t)\lambda_{10}\Delta t + p_2(t)\lambda_{20}\Delta t + p_0(t)\left[1-(\lambda_{01}+\lambda_{02})\Delta t\right],$$

откуда

$$\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = p_1(t)\lambda_{10} + p_2(t)\lambda_{20} - p_0(t)(\lambda_{01}+\lambda_{02}).$$

Переходя к пределу в левой части последнего равенства получим:

$$p_0'(t) = \lambda_{10}p_1(t) + \lambda_{20}p_2(t) - (\lambda_{01}+\lambda_{02})p_0(t).$$

Получили дифференциальное уравнение первого порядка, т.е. уравнение, со-держащее как саму неизвестную функцию, так и ее производную первого по-рядка.

Рассуждая аналогично для других состояний системы S, получим систему диф-ференциаль-ных уравнений, которая носит название системы дифференциаль-ных уравнений Колмогорова для вероятностей состояний

$$\begin{cases} p_0'(t) = \lambda_{10}p_1(t) + \lambda_{20}p_2(t) - (\lambda_{01}+\lambda_{02})p_0(t) \\ p_1'(t) = \lambda_{01}p_0(t) + \lambda_{31}p_3(t) - (\lambda_{10}+\lambda_{13})p_1(t) \\ p_2'(t) = \lambda_{02}p_0(t) + \lambda_{32}p_3(t) - (\lambda_{20}+\lambda_{23})p_2(t) \\ p_3'(t) = \lambda_{13}p_1(t) + \lambda_{23}p_2(t) - (\lambda_{31}+\lambda_{32})p_3(t). \end{cases} \qquad (*3)$$

Сформулируем правило составления уравнений Колмогорова. В левой части каждого из них стоит производная вероятностей i-го состояния. В правой части — сумма произведений ве-роятностей всех состояний (из которых идут дуги в данное состояние) на интенсивности соответст-вующих потоков событий минус суммарная интенсивность всех потоков, выводящих систему из дан-ного состояния, умноженная на вероятность данного состояния.

В системе (*3) независимых уравнений на единицу меньше общего числа уравнений, по-этому для решения системы для каждого фиксированного t необходимо добавить еще уравнение (*1). Особенность решения дифференциальных уравнений состоит в том, что необходимо задать начальные условия уравнений, т.е. в данном случае вероятности состояний системы в начальный момент времени $t = 0$. Так, например, систему уравнений (*3) естественно решать при условии, что в начальный момент времени $t = 0$ оба узла исправны и система находится в состоянии S_0, т.е. при начальном условии $p_0(0) = 1, p_1(0) = p_2(0) = p_3(0) = 0.$
Уравнения Колмогорова дают возможность найти все вероятности состояний как функции времени. Особый интерес представляют вероятности системы $p_i(t)$ в предельном стационарном решении, т.е. при $t \to \infty$, которые называются предельными (или финальными) вероятностями состояний.
В теории случайных процессов доказывается, что если число состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое состояние, то пре-дельные вероятности существуют.
Предельная вероятность состояния S_i имеет вполне определенный смысл: она показывает среднее относительное время пребывания системы в этом состоянии. Например, если предельная вероятность состояния S_0, т.е. $p_0 = 0,5$, то это означает, что в среднем половину времени система находится в состоянии S_0.
Так как предельные вероятности постоянны, то, заменяя в уравнениях Колмогорова их производные нулевыми значениями, получим систему линейных алгебраических уравнений, пии-сывающих стационарный режим. Для системы S с графом состояний, изображенном на рис.1, та-кая система уравнений имеет вид

$$\begin{cases}(\lambda_{01}+\lambda_{02})p_0=\lambda_{10}p_1+\lambda_{20}p_2\\(\lambda_{10}+\lambda_{13})p_1=\lambda_{01}p_0+\lambda_{31}p_3\\(\lambda_{20}+\lambda_{23})p_2=\lambda_{02}p_0+\lambda_{32}p_3\\(\lambda_{31}+\lambda_{32})p_3=\lambda_{13}p_1+\lambda_{23}p_2.\end{cases}\qquad(*4)$$

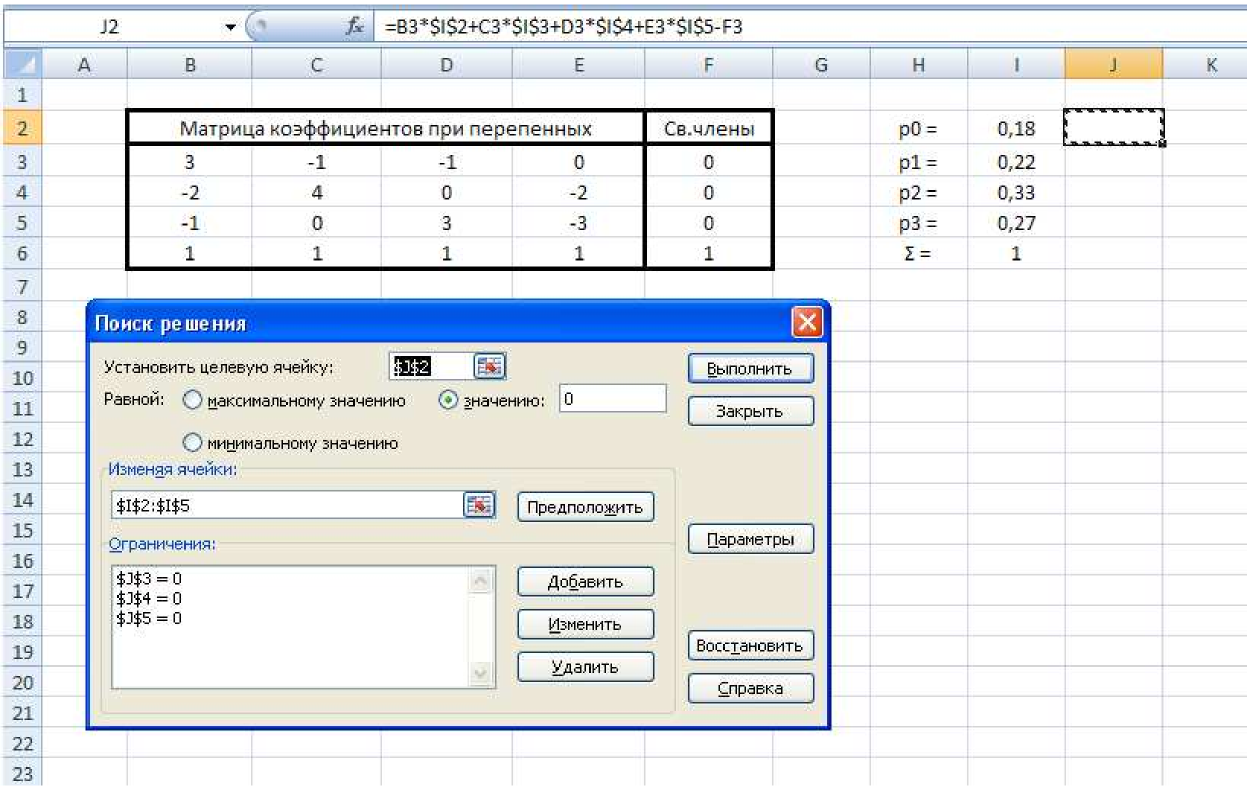
Система (*4) может быть получена непосредственно по размеченному графу состояний, если руководствоваться правилом, согласно которому в левой части уравнений стоит предельная вероятность данного состояния p_i, умноженная на суммарную интенсивность всех потоков, вхо-дящих в i-е состояние, на вероятности тех состояний, из которых эти потоки исходят.

3.3.3 РЕШЕНИЕ ЗАДАЧИ С ПОМОЩЬЮ ПРОГРАММЫ MS EXCEL

Система алгебраических уравнений для нашего случая согласно (*4) имеет вид

$$\begin{cases}3p_0=p_1+p_2\\4p_1=2p_0+2p_3\\3p_2=p_0+3p_3\\p_0+p_1+p_2+p_3=1.\end{cases}\qquad(*5)$$

Решим систему линейных уравнений с помощью надстройки «поиск решения» в MS Excel:

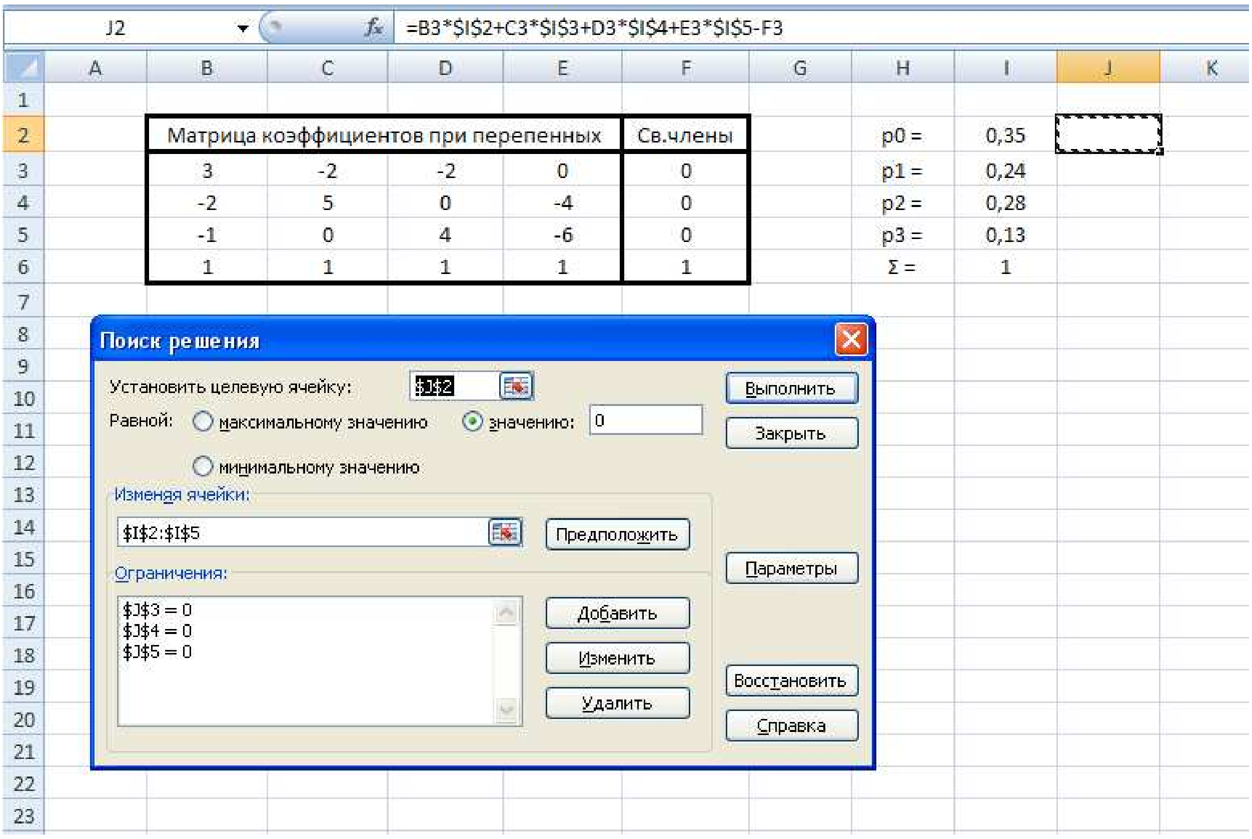


получим p_0 = 0,18; p_1 = 0,22; p_2 = 0,33; p_3 = 27; т.е. в предельном стационарном режиме система S в среднем 18% времени будет находиться в состоянии S_0 (оба конвейера исправны), 22% времени в состоянии S_1 (первый конвейер ремонтируется, второй работает), 33% — в состоянии S_2 (второй конвейер ремонтируется, первый работает) и 27% в состоянии S_3 (оба конвейера ремонтируются). Определим чистый доход от эксплуатации в стационарном режиме рассмотренной сис-темы S в условиях, что в единицу времени исправная работа конвейера один и конвейера два приносит доход соответственно 11 и 7 денежных единиц, а их ремонт требует соответственно затрат 5 и 3 денежных единиц. Оценим экономическую эффективность имеющейся возможности уменьшения вдвое среднего времени ремонта каждого из двух конвейеров, если при этом придется вдвое увеличить затраты на ремонт каждого из них (в единицу времени). Для решения этой задачи с учетом полученных значений p_0, p_1, p_2, p_3 определим долю времени исправной работы первого узла, т.е. $p_0 + p_2$ = 0,18 + 0,33 = 0,51 и долю времени исправной работы второго узла $p_0 + p_1$ = 0,18 + 0,22 = 0,4. В то же время первый узел находится в ремонте в среднем долю времени равную $p_1 + p_3$ = 0,22 + 0,27 = 0,49, а второй узел $p_2 + p_3$ = 0,33 + 0,27 = 0,6. Поэтому средний чистый доход в единицу времени от эксплуатации системы равен Д = 0,51*11 + 0,4*7 – 0,49*5 - 0,6*3 = 4,16 ден. ед. Уменьшение вдвое среднего времени ремонта каждого кон-вейера будет означать увеличение вдвое интенсивностей потока «окончания ремонтов» каждого конвейера, т.е. теперь $\lambda_{10} = 2$, $\lambda_{20} = 2$, $\lambda_{31} = 4$, $\lambda_{32} = 6$ и система уравнений, описывающая стационарный режим системы S, будет иметь вид:

$$\begin{cases} 3p_0 = 2p_1 + 2p_2 \\ 5p_1 = 2p_0 + 4p_3 \\ 4p_2 = p_0 + 6p_3 \\ p_0 + p_1 + p_2 + p_3 = 1. \end{cases}$$

(*6)

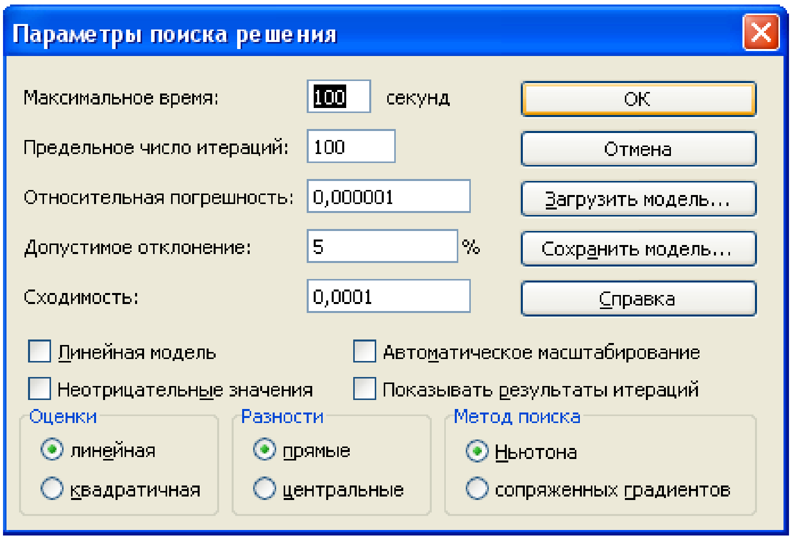
Решим линейную систему уравнений с помощью надстройки «поиск решения» в MS Excel:



получим $p_0 = 0{,}35$; $p_1 = 0{,}24$; $p_2 = 0{,}28$; $p_3 = 0.13$. Учитывая, что $p_0 + p_2 = 0{,}35 + 0{,}28 = 0{,}63$; $p_0 + p_1 = 0{,}35 + 0{,}24 = 0{,}59$; $p_1 + p_3 = 0{,}24 + 0{,}13 = 0{,}37$; $p_2 + p_3 = 0{,}28 + 0{,}13 = 0{,}41$, а затраты на ремонт первого и второго конвейера составляют соответственно 10 и 6 ден. ед., вычислим чистый сред-ний доход в единицу времени: Д1 = 0,63*11 + 0,59*7 - 0,37*10 - 0,41*6 = 4,9 ден. ед.

Так как Д1 (4,9) больше Д (4,16), примерно на 15%, то экономическая целесообразность ускорения ремонта узлов очевидна.

При работе с надстройкой «поиск решения» в MS Excel, для нее использовались сле-дующие параметры:


Параметры поиска решения
Максимальное время:
100
секунд
OK
Предельное число итераций:
100
Отмена
Относительная погрешность:
0,000001
Загрузить модель...
Допустимое отклонение:
5
%
Сохранить модель...
Сходимость:
0,0001
Справка
Линейная модель
Автоматическое масштабирование
Неотрицательные значения
Показывать результаты итераций
Оценки
линейная
квадратичная
Разности
прямые
центральные
Метод поиска
Ньютона
сопряженных градиентов

Тема 4. Системы массового обслуживания

Введение

В реальной жизни часто приходится сталкиваться с ситуациями, в которых необходимо пребывать в состоянии ожидания. Примерами тому может служить очередь покупателей у касс большого магазина, группа пассажирских самолетов, ожидающих разрешения на взлет в аэропорте, ряд вышедших из строя станков и механизмов, поставленных в очередь для починки в ремонтном цехе предприятия и так далее. Во всех перечисленных случаях имеем дело с массовостью и обслуживанием. Изучением таких ситуаций занимается теория систем массового обслуживания (далее СМО).

Избежать ситуации ожидания чаще всего не удается, но можно сократить время ожидания до какого-то терпимого предела.

Встречающийся на каждом шагу феномен ожидания чаще всего является прямым следствием вероятностного характера возникновения потребности в том или ином виде обслуживания и разброса показателей соответствующих обслуживающих систем. Действительно, ни время возникновения потребностей в обслуживании, ни продолжительность обслуживания, как правило, заранее неизвестны.

В теории систем массового обслуживания обслуживаемый объект называется требованием. В общем случае под требованием обычно понимают запрос на удовлетворение некоторой потребности, например, разговор с абонентом, посадка самолета, покупка продуктов, получение материалов на складе.

Средства, обслуживающие требования, называются обслуживающими устройствами или каналами обслуживания. Например, к ним относятся каналы телефонной связи, посадочной полосы, мастера-ремонтники, билетные кассиры, погрузочно-разгрузочные точки на базах и складах.

В теории СМО рассматриваются такие случаи, когда поступление требований происходит через случайные промежутки времени, а продолжительность обслуживания требований не является постоянной, т.е. носит случайный характер.

Цель изучения режима функционирования обслуживающей системы в условиях, когда фактор случайности является существенным, контролировать некоторые количественные показатели функционирования систем массового обслуживания. Такими показателями, в частности, являются среднее время пребывания клиента в очереди или доля времени, в течение которой обслуживающая система простаивает. При этом в первом случае мы оцениваем систему с позиции «клиента», тогда как во втором случае мы оцениваем степень загруженности обслуживающей системы. На интуитивном уровне понятно, что чем больше время ожидания заявки в очереди на обслуживание, тем меньше доля времени простаивания системы обслуживания и наоборот, чем меньше время пребывания требований в очереди на обслуживании, тем больше шансов у обслуживающей системы находиться в состоянии вынужденных простоев. Путем варьирования операционными характеристиками обслуживающей системы может быть достигнут разумный компромисс между требованиями «клиентов» и мощностью обслуживающей системы.

В общем, модели СМО очень распространены и применяются во многих сферах деятельности человека так же и в компьютеризации. Модели СМО удобны для описания отдельных современных вычислительных систем, таких как процессор - винчестер, канал ввода – вывода и т.д. Вычислительная система в целом представляет собой совокупность взаимосвязанных систем. Например: заявка на решение некоторой задачи, проходит несколько этапов обработки, обращения к внешним запоминающим устройствам и устройствам и устройствам ввода – вывода. После выполнения некоторой последовательности таких этапов, заявка считается обслуженной, и она покидает систему.

Теория массового обслуживания опирается на теорию вероятностей и математическую статистику. Первоначальное развитие теории массового обслуживания связано с именем датского ученого А. К. Эрланга (1878—1929), с его трудами в области проектирования и эксплуатации телефонных станций.

Теория массового обслуживания — область прикладной математики, занимающаяся анализом процессов в системах производства, обслуживания, управления, в которых однородные события повторяются многократно.

Большой вклад в развитие этой теории внесли российские математики А. Я. Хинчин, Б. В. Гнеденко, А. Н. Колмогоров Е. С. Вентцель и др.

Предметом теории массового обслуживания является установление зависимостей между характером потока заявок, числом каналов обслуживания, производительностью отдельного канала и эффективным обслуживанием с целью нахождения наилучших путей управления этими процессами. Задачи теории массового обслуживания носят оптимизационный характер и в конечном итоге включают экономический аспект по определению такого варианта системы, при котором будет обеспечен минимум суммарных затрат от ожидания обслуживания, потерь времени и ресурсов на обслуживание и от простоев каналов обслуживания.

4.1 Массовое обслуживание в коммерческой деятельности

Природа массового обслуживания весьма тонка и сложна. Коммерческая деятельность связана с выполнением множества операций на этапах движения, например товарной массы из сферы производства в сферу потребления. Такими операциями являются погрузка товаров, перевозка, разгрузка, хранение, обработка, фасовка, реализация. Кроме таких основных операций процесс движения товаров сопровождается большим количеством предварительных, подготовительных, сопутствующих, параллельных и последующих операций с платежными документами, тарой, деньгами, автомашинами, клиентами и т.п.

Для перечисленных фрагментов коммерческой деятельности характерны массовость поступления товаров, денег, посетителей в случайные моменты времени, затем их последовательное Обслуживание (удовлетворение требований, запросов, заявок) путем выполнения соответствующих операций, время, выполнения которых носит также случайный характер. Все это создает неравномерность в работе, порождает недогрузки, простой и перегрузки в коммерческих операциях. Много неприятностей доставляют очереди. В связи с этим возникают задачи анализа существующих вариантов выполнения всей совокупности операций.

Кроме того, возникают другие задачи, связанные с созданием, организацией и планированием нового экономичного, рационального варианта выполнения множества операций.

Задачи организации массового обслуживания возникают практически во всех сферах человеческой деятельности, например обслуживание продавцами покупателей в магазинах, обслуживание посетителей на предприятиях общественного питания, обслуживание клиентов на предприятиях бытового обслуживания, обеспечение телефонных разговоров на телефонной станции, оказание медицинской помощи больным в поликлинике и т.д. Во всех приведенных примерах возникает необходимость в удовлетворении запросов большого числа потребителей.

Перечисленные задачи можно успешно решать с помощью методов и моделей специально созданной для этих целей теории массового обслуживания (ТМО). Обслуживать необходимо кого-либо или что-либо, что определяется понятием «заявка (требование) на обслуживание», а операции обслуживания выполняются кем-либо или чем-либо, называемыми каналами (узлами) обслуживания.

Роль заявок в коммерческой деятельности выполняют товары, посетители, деньги, ревизоры, документы, а роль каналов обслуживания — продавцы, администраторы, повара, кондитеры, официанты, кассиры, товароведы, грузчики, торговое оборудование и др.

Заявки в силу массовости поступления на обслуживание образуют потоки, которые до выполнения операций обслуживания называются входящими, а после возможного ожидания начала обслуживания, т.е. простоя в очереди, образуют потоки обслуживания в каналах, а затем формируется выходящий поток заявок. В целом совокупность элементов входящего потока заявок, очереди, каналов обслуживания и выходящего потока заявок образует простейшую одноканальную систему массового обслуживания – СМО.

Системы массового обслуживания – это такие системы, в которые в случайные моменты времени поступают заявки на обслуживание, при этом поступившие заявки обслуживаются с помощью имеющихся в распоряжении системы каналов обслуживания.

Примерами систем массового обслуживания могут служить:

- посты технического обслуживания автомобилей;
- посты ремонта автомобилей;
- персональные компьютеры, обслуживающие поступающие заявки или требования на решение тех или иных задач;
- станции технического обслуживания автомобилей;
- аудиторские фирмы;
- отделы налоговых инспекций, занимающиеся приемкой и проверкой текущей отчетности предприятий;
- телефонные станции;
- узлы расчета с покупателями в магазинах;
- рабочие места экономиста, бухгалтера и т.д.

Процедура обслуживания считается завершенной, когда заявка на обслуживание покидает систему. Продолжительность интервала времени, требуемого для реализации процедуры обслуживания, зависит в основном от характера запроса заявки на обслуживание, состояния самой обслуживающей системы и канала обслуживания.

Обслуживание заявок – это процесс удовлетворения потребности. Обслуживание имеет различный характер по своей природе. Однако во всех примерах поступившие заявки нуждаются в обслуживании со стороны какого-либо устройства. В некоторых случаях обслуживание производится одним человеком, в некоторых – группой людей, а в некоторых случаях – техническими устройствами.

Совокупность средств, которые осуществляют обслуживание заявок, называется каналом обслуживания.

Если каналы обслуживания способны удовлетворить одинаковые заявки, то каналы обслуживания называются однородными. Совокупность однородных каналов обслуживания называется обслуживающей системой.

В систему массового обслуживания поступает большое количество заявок в случайные моменты времени, длительность обслуживания которых также является случайной величиной. Последовательное поступление заявок в систему обслуживания называется входящим потоком заявок, а последовательность заявок, покидающих систему обслуживания, — выходящим потоком.

Случайный характер распределения длительности выполнения операций обслуживания наряду со случайным характером поступления требований на обслуживание приводит к тому, что в каналах обслуживания протекает случайный процесс, который может быть назван (по аналогии с входным потоком заявок) потоком обслуживания заявок или просто потоком обслуживания.

Заявки, поступающие в систему обслуживания, могут покинуть ее и будучи не обслуженными. Например, если покупатель не найдет в магазине нужный товар, то он покидает магазин, будучи не обслуженным. Покупатель может покинуть магазин также, если нужный товар имеется, но большая очередь, а покупатель не располагает временем.

Теория массового обслуживания занимается изучением процессов, связанных с массовым обслуживанием, разработкой методов решения типичных задач массового обслуживания.

При исследовании эффективности работы системы обслуживания важную роль играют различные способы расположения в системе каналов обслуживания.

При параллельном расположении каналов обслуживания требование может быть обслужено любым свободным каналом. Примером такой системы обслуживания является расчетный узел в магазинах самообслуживания» где число каналов обслуживания совпадает с числом кассиров-контролеров.

На практике часто обслуживание одной заявки осуществляется последовательно несколькими каналами обслуживания. При этом очередной канал обслуживания начинает работу по обслуживанию заявки после того, как предыдущий канал закончил свою работу. В таких системах процесс обслуживания носит многофазовый характер, обслуживание заявки одним каналом называется фазой обслуживания. Например, если в магазине самообслуживания имеются отделы с продавцами, то покупатели сначала обслуживаются продавцами, а потом уже кассирами-контролерами.

Организация системы обслуживания зависит от воли человека. Под качеством функционирования системы в теории массового обслуживания понимают не то, насколько хорошо выполнено обслуживание, а то, насколько полно загружена система обслуживания, не простаивают ли каналы обслуживания, не образуется ли очередь.

В коммерческой деятельности заявки, поступающие в систему массового обслуживания, выступают с высокими претензиями ещё и на качество обслуживания в целом, которое включает не только перечень характеристик, исторически сложившихся и рассматриваемых непосредственно в теории массового обслуживания, но и дополнительные характерные для специфики коммерческой деятельности, в частности отдельных процедур обслуживания, требования к уровню которых к настоящему времени сильно возросли. В связи с этим необходимо учитывать еще и показатели коммерческой деятельности.

Работу системы обслуживания характеризуют такие показатели, как время ожидания начала обслуживания, длина очереди, возможность получения отказа в обслуживании, возможность простоя каналов обслуживания, стоимость обслуживания и в конечном итоге удовлетворение качеством обслуживания, которое еще включает показатели коммерческой деятельности.

Чтобы улучшить качество функционирования системы обслуживания, необходимо определить, каким образом распределить поступающие заявки между каналами обслуживания, какое количество каналов обслуживания необходимо иметь, как расположить или сгруппировать каналы обслуживания или обслуживающие аппараты для улучшения показателей коммерческой деятельности. Для решения перечисленных задач существует эффективный метод моделирования, включающий и объединяющий достижения разных наук, в том числе математики.

4.2 Моделирование систем массового обслуживания

4.2.1 Потоки событий

Переходы СМО из одного состояния в другое происходят под воздействием вполне определенных событий — поступления заявок и их обслуживания. Последовательность появления событий, следующих одно за другим в случайные моменты времени, формирует так называемый поток событий. Примерами таких потоков в коммерческой деятельности являются потоки различной природы — товаров, денег, документов, транспорта, клиентов, покупателей, телефонных звонков, переговоров. Поведение системы обычно определяется не одним, а сразу несколькими потоками событий. Например, обслуживание покупателей в магазине определяется потоком покупателей и потоком обслуживания; в этих потоках случайными являются моменты появления покупателей, время ожидания в очереди и время, затрачиваемое на обслуживание каждого покупателя. При этом основной характерной чертой потоков является вероятностное распределение времени между соседними событиями. Существуют различные потоки, которые отличаются своими характеристиками.

Поток событий называется *регулярным*, если в нем события следуют одно за другим через заранее заданные и строго определенные промежутки времени. Такой поток является идеальным и очень редко встречается на практике. Чаще встречаются нерегулярные потоки, не обладающие свойством регулярности.

Поток событий называется *стационарным*, если вероятность попадания любого числа событий на промежуток времени зависит только от длины этого промежутка и не зависит от того, как далеко расположен этот промежуток от начала отсчета времени. Стационарность потока означает независимость от времени его вероятностных характеристик, в частности, интенсивность такого потока есть среднее число событий в единицу времени и остается величиной постоянной. На практике обычно потоки могут считаться стационарными только на некотором ограниченном промежутке времени. Обычно поток покупателей, например, в ь. магазине существенно меняется в течение рабочего дня. Однако можно выделить определенные временные интервалы, внутри, которых этот поток допустимо рассматривать как стационарный, имеющий постоянную интенсивность.

Поток событий называется потоком без последствия, если число событий, попадающих на один из произвольно выбранных промежутков времени, не зависит от числа событий, попавших на другой, также произвольно выбранный промежуток, при условии, что эти промежутки не пересекаются между собой. В потоке без последствия события появляются в последовательные моменты времени независимо друг от друга. Например, поток покупателей, входящих в магазин, можно считать потоком без последствия потому, что причины, обусловившие приход каждого из них, не связаны с аналогичными причинами для других покупателей.

Ординарность потока означает практическую невозможность одновременного поступления двух и более требований.

Если поток одновременно обладает свойствами *стационарности, ординарности и отсутствием последствия*, то такой поток называется *простейшим* (или *пуассоновским*) потоком событий. Математическое описание воздействия такого потока на системы оказывается наиболее простым. Поэтому, в частности, простейший поток играет среди других существующих потоков особую роль.

Рассмотрим на оси времени: некоторый промежуток времени τ. Допустим, вероятность попадания случайного события на этот промежуток p, а полное число возможных событий - n. При наличии свойства ординарности потока событий вероятность p должна быть достаточно малой величиной, а n- достаточно большим числом, поскольку рассматриваются массовые явления. В этих условиях для вычисления вероятности попадания на промежуток времени τ некоторого числа событий m можно воспользоваться формулой Пуассона:

$$P_{m,n} = \frac{a^m}{m!}e^{-a}, \qquad (m = \overrightarrow{0,n}),$$

где величина $a = np$ - среднее число событий, попадающих на промежуток времени τ, которое можно определить через интенсивность потока событий λ следующим образом:

$$a = \lambda\tau.$$

Размерность интенсивности потока λ есть среднее число событий в единицу времени. Между n и λ, p и τ имеется следующая связь:

$$n = \lambda t; = \frac{\tau}{t},$$

где t — весь промежуток времени, на котором рассматривается действие потока событий.

Необходимо определить распределение интервала времени T между событиями в таком потоке. Поскольку это случайная величина, найдем ее функцию распределения. Как известно из теории вероятностей, интегральная функция распределения $F(t)$ есть вероятность того, что величина T будет меньше времени t.

$$F(t) = P(T<t).$$

По условию в течение времени T не должно произойти ни одного события, а на интервале времени t должно появиться хотя бы одно событие. Эта вероятность вычисляется с помощью вероятности противоположного события на промежутке времени (0; t), куда не попало ни одного события, т.е. $m= 0$, тогда

$$F(t) = 1 - P_0 = 1 - \frac{a^0}{0!}e^{-a} = 1 - e^{-\lambda t}, t \geq 0.$$

Для малых Δt можно получить приближенную формулу, получаемую заменой функции $e^{-\lambda t}$, только двумя членами разложения в ряд по степеням Δt, тогда вероятность попадания на малый промежуток времени Δt хотя бы одного события составляет

$$P(T < \Delta t) = 1 - e^{-\lambda t} = 1 - \left[1 - \lambda\Delta t + \frac{1}{2}(\lambda\Delta t)^2 - \frac{1}{6}(\lambda\Delta t)^3\right] = \lambda\Delta t.$$

Плотность распределения промежутка времени между двумя последовательными событиями получим, продифференцировав $F(t)$ по времени,

$$f(t) = \lambda e - \lambda^t, \qquad t \geq 0.$$

Пользуясь полученной функцией плотности распределения, можно получить числовые характеристики случайной величины *T:* математическое ожидание *M(T),* дисперсию *D(T)* и среднее квадратическое отклонение σ(*T*).

$$M(T) = \lambda \int_0^{\infty} te^{-\lambda t} * dt = \frac{1}{\lambda}; D(T) = \frac{1}{\lambda^2}; \ \sigma(T) = \frac{1}{\lambda}.$$

Отсюда можно сделать следующий вывод: средний интервал времени *T* между любыми двумя соседними событиями в простейшем потоке в среднем равен 1/λ и его среднее квадратическое отклонение также равно 1/*λ*, где λ — интенсивность потока, т.е. среднее число событий, происходящих в единицу времени. Закон распределения случайной величины, обладающей такими свойствами *M(T)=T*, называется показательным (или экспоненциальным), а величина λ является параметром этого показательного закона. Таким образом, для простейшего потока математическое ожидание интервала времени между соседними событиями равно его среднеквадратическому отклонению.

В этом случае вероятность того, что число заявок, поступающих на обслуживание за промежуток времени *t*, равно *k*, определяется по закону Пуассона:

$$P_k(t) = \frac{(\lambda t)^k}{k!} * e^{-\lambda t},$$

где *λ* — интенсивность поступления потока заявок, среднее число событий в СМО за единицу времени, например

$$\left[\frac{\text{чел.}}{\text{мин}}; \frac{\text{руб.}}{\text{ч}}; \frac{\text{чеков}}{\text{ч}}; \frac{\text{докум.}}{\text{день}}; \frac{\text{кг}}{\text{ч}}; \frac{\text{т}}{\text{год}}\right].$$

Для такого потока заявок время между двумя соседними заявками *T* распределено экспоненциально с плотностью вероятности:

$$f(t) = \lambda e^{-\lambda t}.$$

Случайное время ожидания в очереди начала обслуживания $t_{оч}$ тоже можно считать распределенным экспоненциально:

$$f(t_{оч}) = V * e^{-Vt_{оч}},$$

где v — интенсивность потока прохода очереди, определяемая средним числом заявок, проходящих на обслуживание в единицу времени:

$$v = \frac{1}{T_{оч}},$$

где $T_{оч}$ - среднее время ожидания обслуживания в очереди.

Выходной поток заявок связан с потоком обслуживания в канале, где длительность обслуживания $t_{обс}$ является тоже случайной величиной и подчиняется во многих случаях показательному закону распределения с плотностью вероятности:

$$f(t_{обс}) = \mu * e^{-\mu t_{обс}},$$

где μ — интенсивность потока обслуживания, т.е. среднее число заявок, обслуживаемых в единицу времени:

$$\mu = \frac{1}{t_{обс}}\left[\frac{чел.}{мин};\frac{руб.}{ч};\frac{чеков}{ч};\frac{докум.}{день};\frac{кг}{ч};\frac{т}{год}\right],$$

где $t_{обс}$ - среднее время обслуживания заявок.

Важной характеристикой СМО, объединяющей показатели λ и μ, является интенсивность нагрузки: $\rho = \frac{\lambda}{\mu}$, которая показывает степень согласования входного и выходного потоков заявок канала обслуживания и определяет устойчивость системы массового обслуживания.

Кроме понятия простейшего потока событий часто приходится пользоваться понятиями потоков других типов. Поток событий называется *потоком Пальма,* когда в этом потоке промежутки времени между последовательными событиями T_1, T_2,..., T_k, ..., T_n являются независимыми, одинаково распределенными, случайными величинами, но в отличие от простейшего потока не обязательно распределенными по показательному закону. Простейший поток является частным случаем потока Пальма.

Важным частным случаем потока Пальма является так называемый *поток Эрланга.* Этот поток получается «прореживанием» простейшего потока. Такое «прореживание» производится путем отбора по определенному правилу событий из простейшего потока. Например, условившись учитывать только каждое второе событие из образующих простейший поток, мы получим поток Эрланга второго порядка. Если брать только каждое третье событие, то образуется поток Эрланга третьего порядка и т.д. Можно получить потоки Эрланга любого *k-го* порядка. Простейший поток есть поток Эрланга первого порядка.

Любое исследование системы массового обслуживания начинается с изучения того, что необходимо обслуживать, следовательно, с изучения входящего потока заявок и его характеристик.

Поскольку моменты времени t_i и интервалы времени поступления заявок τ, затем продолжительность операций обслуживания $t_{обс}$ и время ожидания в очереди $t_{оч}$, а также длина очереди $l_{оч}$ — случайные величины, то, следовательно, характеристики состояния СМО носят вероятностный характер, а для их описания следует применять методы и модели теории массового обслуживания.

Перечисленные выше характеристики k, τ, λ, $L_{оч}$,$T_{оч}$,,ν, $t_{обс}$,μ , ρ, P_k являются наиболее общими для СМО, которые являются обычно лишь некоторой частью целевой функции, поскольку необходимо учитывать еще и показатели коммерческой деятельности.

4.2.2 Графы состояний СМО

При анализе случайных процессов с дискретными состояниями и непрерывным временем удобно пользоваться вариантом схематичного изображения возможных состояний СМО (рисунок 1) в виде графа с разметкой его возможных фиксированных состояний. Состояния СМО изображаются обычно либо прямоугольниками, либо кружками, а возможные направления переходов из одного состояния в другое ориентированы стрелками, соединяющими эти состояния.

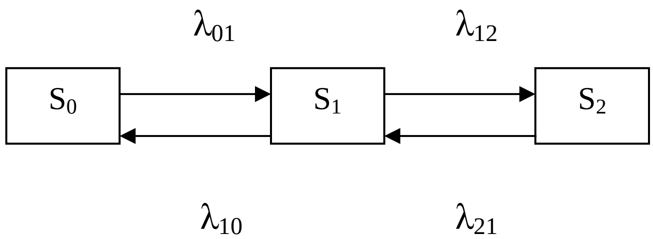


Рисунок 4.1. Размеченный граф состояний СМО

Система может находиться в одном из трех состояний: S_0 — канал свободен, простаивает, S_1 — канал занят обслуживанием, S_2 — канал занят обслуживанием и одна заявка в очереди. Переход системы из состояния S_0 в S_1 происходит под воздействием простейшего потока заявок интенсивностью λ_{01}, а из состояния S_1 в состояние S_0 систему переводит поток обслуживания с интенсивностью λ_{10}. Граф состояний системы обслуживания с проставленными интенсивностями потоков у стрелок называется размеченным. Поскольку пребывание системы в том или ином состоянии носит вероятностный характер, то вероятность $p_i(t)$ того, что система будет находиться в состоянии S_i в момент времени t, называется вероятностью i-го состояния СМО и определяется числом поступивших заявок k на обслуживание.

Случайный процесс, происходящий в системе, заключается в том, что в случайные моменты времени $t_0, t_1, t_2, \ldots, t_k, \ldots, t_n$ система оказывается в том или другом заранее известном дискретном состоянии последовательно. Такая случайная последовательность событий называется *марковской цепью*, если для каждого шага вероятность перехода из одного состояния S_i в любое другое Sj не зависит от того, когда и как система перешла в состояние S_i. Описывается марковская цепь с помощью вероятности состояний, причем они образуют полную группу событий, поэтому их сумма равна единице. Если вероятность перехода не зависит от номера k, то марковская цепь называется однородной. Зная начальное состояние системы обслуживания, можно найти вероятности состояний для любого значения k - числа заявок, поступивших на обслуживание.

4.2.3 Случайные процессы

Переход СМО из одного состояния в другое происходит случайным образом и представляет собой случайный процесс. Работа СМО — случайный процесс с дискретными состояниями, поскольку его возможные состояния во времени можно заранее перечислить. Причем переход из одного состояния в другое происходит скачкообразно, в случайные моменты времени, поэтому он называется процессом с непрерывным временем. Таким образом, работа СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем.

Из множества разновидностей случайных процессов наибольшее распространение в коммерческой деятельности получили такие процессы, для которых в любой момент времени характеристики процесса в будущем зависят только от его состояния в настоящий момент и не зависят от предыстории - от прошлого.

Такие случайные процессы называются процессами без последствия, или марковскими, в которых при фиксированном настоящем будущее состояние СМО не зависит от прошлого. Случайный процесс, протекающий в системе, называется марковским случайным процессом, или «процессом без последствия», если он обладает следующим свойством: для каждого момента времени t_0 вероятность любого состояния $t > t_0$ системы S_i- в будущем ($t > t_0$) зависит только от ее состояния в настоящем (при $t = t_0$) и не зависит от того, когда и каким образом система пришла в это состояние, т.е. от того, как развивался процесс в прошлом.

Марковские случайные процессы делятся на два класса: процессы с дискретными и непрерывными состояниями. Процесс с дискретными состояниями возникает в системах, обладающих только некоторыми фиксированными состояниями, между которыми возможны скачкообразные переходы в некоторые, заранее не известные моменты времени.

Процессы с непрерывными состояниями отличаются непрерывным плавным переходом из одного состояния в другое. Эти процессы более характерны для технических устройств, нежели для экономических объектов, где обычно лишь приближенно можно говорить о непрерывности процесса (например, непрерывном расходовании запаса товара), тогда как фактически всегда процесс имеет дискретный характер.

Марковские случайные процессы с дискретными состояниями в свою очередь подразделяются на процессы с дискретным временем и процессы с непрерывным временем. В первом случае переходы из одного состояния в другое происходят только в определенные, заранее фиксированные моменты времени, тогда как в промежутки между этими моментами система сохраняет свое состояние. Во втором случае переход системы из состояния в состояние может происходить в любой случайный момент времени.

На практике процессы с непрерывным временем встречаются значительно чаще, поскольку переходы системы из одного состояния в другое обычно происходят не в какие-то фиксированные моменты времени, а в любые случайные моменты времени.

Для описания процессов с непрерывным временем используется модель в виде так называемой марковской цепи с дискретными состояниями системы, или непрерывной марковской цепью.

4.2.4 Уравнения Колмогорова

Рассмотрим математическое описание марковского случайного процесса с дискретными состояниями системы S_0, S_1, S_2 (см. рисунок.1) и непрерывным временем. Полагаем, что все переходы системы массового обслуживания из состояния S_i в состояние S_j происходят под воздействием простейших потоков событий с интенсивностями λ_{ij} , а обратный переход — под воздействием другого потока λ_{ij} . Введем обозначение p_i как вероятность того, что в момент времени t система: находится в состоянии S_i . Для любого момента времени t справедливо записать нормировочное условие - сумма вероятностей всех состояний равна 1:

$$\sum_{i=0}^{2} p_i(t) = p_0(t) + p_1(t) + p_2(t) = 1 .$$

Проведем анализ системы в момент времени t, задав малое приращение времени Δt, и найдем вероятность $p_1\ (t + \Delta t)$ того, что система в момент времени $(t + \Delta t)$ будет находиться в состоянии S_1, которое достигается разными вариантами:

а) система в момент t с вероятностью $p_i\,(t)$ находилась в состоянии S_1 и за малое приращение времени Δt так и не перешла в другое соседнее состояние - ни в S_0, ни в S_2. Вывести систему из состояния S_1 можно суммарным простейшим потоком с интенсивностью (λ_{10}+ λ_{12}), поскольку суперпозиция простейших потоков также является простейшим потоком. На этом основании вероятность выхода из состояния S_1 за малый промежуток времени Δt приближенно равна $(\lambda_{10} + \lambda_{12}) * \Delta t$. Тогда вероятность невыхода из этого состояния равна $[1 — (\lambda_{10} + \lambda_{12}) * \Delta t]$. В соответствии с этим вероятность того, что система останется в состоянии S_1 на основании теоремы умножения вероятностей, равна:

$$p_1(t)[1 — (\lambda_{10} + \lambda_{12})*\Delta t];$$

б) система находилась в соседнем состоянии S_0 и за малое время Δt перешла в состояние S_1. Переход системы происходит под воздействием потока λ_{01} с вероятностью, приближенно равной $\lambda_{01}\,\Delta t$.Вероятность того, что система будет находиться в состоянии S_1, в этом варианте равна $p_0(t)\,\lambda_{01}\,\Delta t$;

в) система находилась в состоянии S_2 и за время Δt перешла в состояние S_1 под воздействием потока интенсивностью λ_{21} с вероятностью, приближенно равной $\lambda_{21}\,\Delta t$. Вероятность того, что система будет находиться в состоянии S_1, равна $p_2(t)\,\lambda_{21}\,\Delta t$.

Применяя теорему сложения вероятностей для этих вариантов, получим выражение:

$$p_1(t+\Delta t) = p_1 t[1-(\lambda_{10}+\lambda_{12})\Delta t] + p_0(t)\lambda_{01}\Delta t + p_2(t)\lambda_{21}\Delta t,$$

которое можно записать иначе:

$$\frac{p_1(t+\Delta t)-p_1(t)}{\Delta t} = p_0(t)\lambda_{01} + p_2(t)\lambda_{21} - p_1(t)(\lambda_{10}+\lambda_{12}).$$

Переходя к пределу при $\Delta t \to 0$, приближенные равенства перейдут в точные, и тогда получим производную первого порядка

$$\frac{dp_1}{dt} = p_0\lambda_{01} + p_2\lambda_{21} - p_1(\lambda_{10}+\lambda_{12}),$$

что является дифференциальным уравнением.

Проводя рассуждения аналогичным образом для всех других состояний системы, получим систему дифференциальных уравнений, которые называются уравнениями А.Н. Колмогорова:

$$\begin{cases} \dfrac{dp_0}{dt} = p_1\lambda_{10} - \lambda_{01}p_0 \\ \dfrac{dp_1}{dt} = p_0\lambda_{01} + p_2\lambda_{21} - p_{1(}\lambda_{10} + \lambda_{12}) \\ \dfrac{dp_2}{dt} = p_1\lambda_{12} - \lambda_{21}p_2 \ . \end{cases}$$

Для составления уравнений Колмогорова существуют общие правила.

Уравнения Колмогорова позволяют вычислить все вероятности состояний СМО S_i в функции времени $p_i(t)$. В теории случайных процессов показано, что если число состояний системы конечно, а из каждого из них можно перейти в любое другое состояние, то существуют предельные (финальные) вероятности состояний, которые показывают на среднюю относительную величину времени пребывания системы в этом состоянии. Если предельная вероятность состояния S_0 равна $p_0 = 0{,}2$, то, следовательно, в среднем 20% времени, или 1/5 рабочего времени, система находится в состоянии S_0.

Поскольку предельные вероятности системы постоянны, то, заменив в уравнениях Колмогорова соответствующие производные нулевыми значениями, получим систему линейных алгебраических уравнений, описывающих стационарный режим СМО. Такую систему уравнений составляют по размеченному графу состояний СМО по следующим правилам: слева от знака равенства в уравнении стоит предельная вероятность p_i рассматриваемого состояния S_i умноженная на суммарную интенсивность всех потоков, выводящих из данного состояния S_i, систему, а справа от знака равенства - сумма произведений интенсивности всех потоков, входящих в состояние S_i систему, на вероятность тех состояний, из которых эти потоки исходят. Для решения подобной системы необходимо добавить еще одно уравнение, определяющее нормировочное условие, поскольку сумма вероятностей всех состояний СМО равна 1:

$$\sum_{i=1}^{n} p_i(t) = 1 \, .$$

Для СМО, имеющей размеченный граф из трех состояний S_0, S_1, S_2 рисунок 1, система уравнений Колмогорова, составленная на основе изложенного правила, имеет следующий вид:

Выходящие Входящие

Для состояния $S_0 \rightarrow p_0 * \lambda_{01} \quad = \quad p_1 * \lambda_{10}$

Для состояния $S_1 \rightarrow (\lambda_{10}+\lambda_{12}) \quad = \quad p_0\lambda_{01}+p_{21}\lambda_{21}$

Для состояния $S_2 \rightarrow p_2\lambda_{21} \quad = \quad p_1\lambda_{12}$

$$p_0+p_1+p_2 = 1$$

4.2.5 Процессы «рождения-гибели»

Среди однородных марковских процессов существует класс случайных процессов, имеющих широкое применение при построении математических моделей в областях демографии, биологии, медицины (эпидемиологии), экономики, коммерческой деятельности. Это так называемые процессы «рождения—гибели», марковские процессы со стохастическими графами состояний следующего вида:

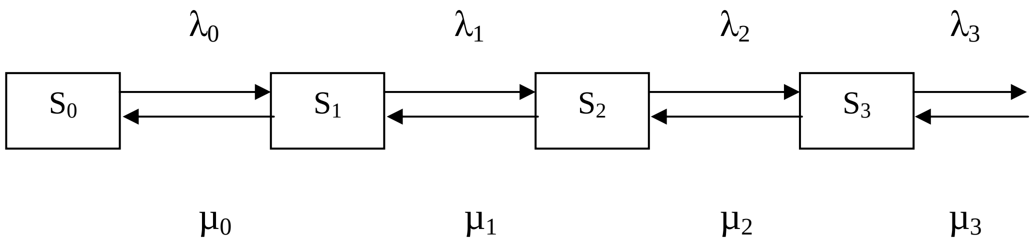


Рисунок 4.2. Размеченный граф процесса «рождения—гибели».

Этот граф воспроизводит известную биологическую интерпретацию: величина $\lambda_к$ отображает интенсивность рождения нового представителя некоторой популяции, причем текущий объем популяции равен *k;* величина μ является интенсивностью гибели (продажи) одного представителя этой популяции, если текущий объем популяции равен *k.* В частности, популяция может быть неограниченной (число *n* состояний марковского процесса является бесконечным, но счетным), интенсивность λ может быть равна нулю (популяция без возможности возрождения).

Для марковского процесса «рождения—гибели», описанного стохастическим графом, приведенным на рисунке 2, найдем финальное распределение. Пользуясь правилами составления уравнений для конечного числа и предельных вероятностей состояния системы S_1, S_2, S_3, ..., S_k, ..., S_n, составим соответствующие уравнения для каждого состояния:

для состояния S_0 - $\lambda_0 p_0 = \mu_0 p_1$;

для состояния S_1 — $(\lambda_1 + \mu_0)p_1 = \lambda_0 p_0 + \mu_1 p_2$, которое с учетом предыдущего уравнения для состояния S_0 можно преобразовать к виду $\lambda_1 p_1 = \mu_1 p_2$.

Аналогично можно составить уравнения для остальных состояний системы S_2, S_3,..., S_k..., S_n. В результате получим следующую систему уравнений:

$$\lambda_0 p_0 = \mu_0 p_2$$

$$\lambda_1 p_1 = \mu_1 p_2$$

$$\dots\dots\dots\dots\dots$$

$$\lambda_{k-1} p_{k-1} = \mu_{k,k} + P_k$$

$$\dots\dots\dots\dots\dots$$

$$\lambda_{n-1} p_{n-1} = \mu_n + P_n$$

$$P_O + P_1 + P_2 + \ldots + P_k + \ldots + P_n = 1.$$

Решая эту систему уравнений, можно получить выражения, определяющие финальные состояния системы массового обслуживания:

$$P_0 = \left(1 + \frac{\lambda 0}{\mu 0} + \frac{\lambda 0 \lambda 1}{\mu 0 \mu 1} + \frac{\lambda 0 \lambda 1 \lambda 2}{\mu 0 \mu 1 \mu 2} + \ldots + \frac{\lambda 0 \lambda 1 \ldots \lambda n}{\mu 0 \mu 1 \ldots \mu n}\right)^{-1};$$

$$P_1 = \frac{\lambda 0}{\mu 0} P_0, \quad P_2 = \frac{\lambda 0 \lambda 1}{\mu 0 \mu 1} P_0, \quad P_3 = \frac{\lambda 0 \lambda 1 \lambda 2}{\mu 0 \mu 1 \mu 2} P_0, \quad P_n = \frac{\lambda 0 \lambda 1 \lambda 2 \ldots \lambda n}{\mu 0 \mu 1 \mu 2 \ldots \mu n}$$

В формулы определения финальных вероятностей состояний p_1, p_2, p_3, ..., p_n входят слагаемые, являющиеся составной частью суммы выражения, определяющей p_0. В числителях этих слагаемых находятся произведения всех интенсивностей, стоящих у стрелок графа состояний, ведущих слева направо до рассматриваемого состояния S_k, а знаменатели представляют собой произведения всех интенсивностей, стоящих у стрелок, ведущих справа налево до рассматриваемого состояния S_k, т.е. μ_0, μ_1, μ_2, μ_3, …, μ_k. В связи с этим запишем эти модели в более компактном виде:

$$P_0 = \left(\sum_{k=1}^{n} \frac{\prod_{j=0}^{k-1} \lambda j}{\prod_{m=1}^{k} \mu m}\right)^{-1}, \quad Pk = \frac{\prod_{j=0}^{k-1} \lambda j}{\prod_{m=1}^{k} \mu m}, \quad k = \overrightarrow{1,n}$$

4.3 Системы массового обслуживания в коммерческой деятельности

Системы массового обслуживания в коммерческой деятельности являются базовой основой самой природы коммерции, поскольку каждый шаг в этой сфере неизменно связан с обслуживанием. Для облегчения процесса моделирования следует пользоваться классификацией СМО по различным признакам, для которых пригодны определенные группы методов и моделей теории массового обслуживания, упрощающие подбор адекватных математических моделей к решению задач обслуживания в коммерческой деятельности.

Существующие варианты заявок, особенности их обслуживания и образования очередей, расположение, количество и организация каналов обслуживания послужили причиной появления большого разнообразия СМО. Наиболее распространенные в коммерческой деятельности виды СМО представлены в виде классификационной структуры на рисунке 3.

В целом структура включает следующие десять основных классификационных признаков: организация потока заявок, количество каналов обслуживания, характер образования очереди, ограничения на очередь, дисциплина очереди, характеристика каналов, расположение каналов, вид ограничений на очередь, правило отбора заявок, наличие и характеристика приоритета. Перечисленные признаки являются ключевыми в проведении исследования и позволяют подобрать перечень необходимых и достаточных характеристик СМО в самом начале решения.

По числу каналов обслуживания СМО разделяются на одноканальные n = 1 и многоканальные, для которых $n \geq 2$. К одноканальным СМО в коммерческой деятельности можно, отнести практически любой вариант локального обслуживания, например выполняемые на рабочем месте операции одним бухгалтером, менеджером, коммерсантом, товароведом, экономистом, торговым аппаратом. В зависимости от взаимного расположения каналов системы подразделяются на СМО с параллельными и с последовательными каналами. В СМО с параллельными каналами входной поток заявок на обслуживание является общим, и поэтому заявки в очереди могут обслуживаться любым свободным каналом. В таких СМО очередь на обслуживание можно рассматривать как общую. В СМО с последовательным расположением каналов каждый канал может рассматриваться как отдельная одноканальная СМО или фаза обслуживания. Выходной поток обслуженных заявок одной СМО является входным потоком для последующей СМО.

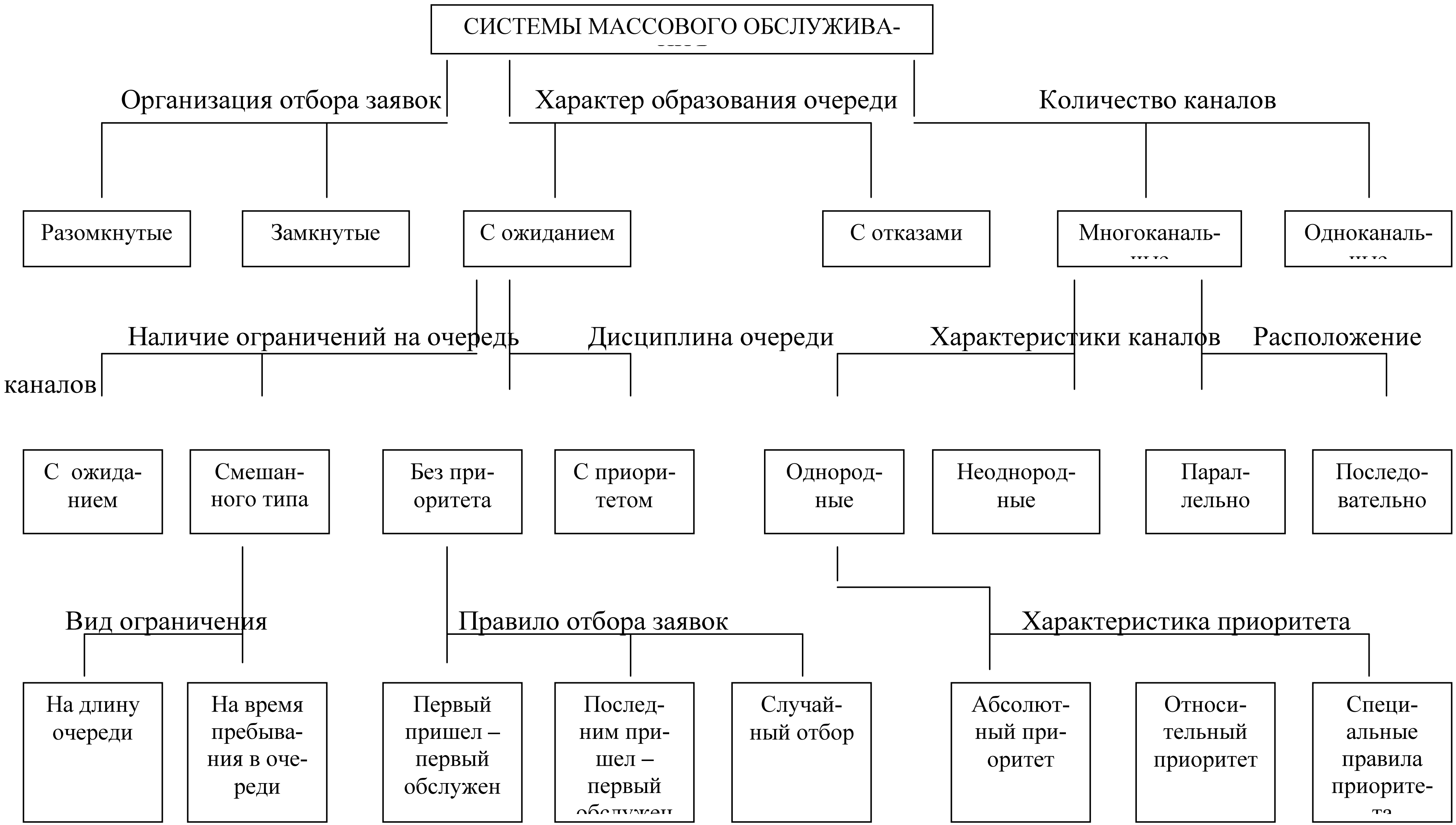


Рисунок 4.3. Классификация систем массового обслуживания

Практически любые заявки в коммерческой деятельности: товары, посетители, торговое оборудование, деньги, документы, проходят множество фаз обслуживания, следовательно, коммерческие системы являются многофазными.

В зависимости от характеристик каналов обслуживания многоканальные СМО подразделяются на СМО с однородными и неоднородными каналами. Отличие состоит в том, что в СМО с однородными каналами заявка потока может обслуживаться любым свободным каналом, а в СМО с неоднородными каналами отдельные заявки обслуживаются только специально для этой цели предназначенными каналами.

К другим сложным СМО коммерческой деятельности относятся посреднические фирмы, мини-маркеты, магазины, торговые павильоны, столовые, кафе, склады, базы, сети коммерческой деятельности районов города.

Модельное описание необходимо для целей анализа, который может учитывать не только характеристики входящих потоков и процессов обслуживания, но и внутреннюю структуру взаимосвязей, взаимопомощи, дисциплину и приоритеты обслуживания. Такой подход позволит более объективно прогнозировать перспективу работоспособности и разрабатывать действенные рекомендации по эффективному перестроению совокупностей существующих СМО, а также разработать рекомендации для вновь проектируемых сложных систем обслуживания.

Каждая простейшая или сложная СМО в зависимости от числа каналов обслуживания и их производительности, от характера потока заявок имеет пропускную способность, позволяющую ей более или менее успешно справляться с потоком заявок. Поэтому предметом исследования СМО в коммерческой деятельности является установление зависимости между характером потока заявок, числом каналов, их расположением, производительностью, правилами и эффективностью обслуживания для успешного достижения поставленной цели при малых затратах на создание и функционирование системы в целом.

СМО в зависимости от возможности образования очереди подразделяются на два основных типа: СМО с отказами обслуживания и СМО с ожиданием (очередью) обслуживания. В СМО с отказами возможен отказ в обслуживании, если все каналы уже заняты обслуживанием, а образовывать очередь и ожидать обслуживание нельзя. В СМО с ожиданиями пришедшая заявка, находящая все каналы обслуживания занятыми, становится в очередь и ожидает, пока не освободится хотя бы один из каналов. Такие системы в коммерческой сфере имеют большое распространение, поскольку все поступающие заявки в коммерческие предприятия вынуждены ожидать начала обслуживания.

В то же время СМО с ожиданием подразделяются на СМО с неограниченным ожиданием, или с неограниченной очередью $L_{оч}$, или временем ожидания $T_{оч}$, и СМО с ограниченными ожиданиями или очередью, в которых накладываются ограничения на максимально возможную длину очереди $max\ L_{оч}= m$ или на максимально возможное время пребывания заявки в очереди $max\ T_{оч} = t_{огр}$, или на время работы системы.

В зависимости от организации потока заявок СМО подразделяются на разомкнутые и замкнутые. В разомкнутых СМО выходной поток обслуженных заявок не связан с входным потоком заявок на обслуживание и имеет неограниченный источник заявок.

В замкнутых СМО (рисунок 4), обслуженные заявки в общем случае после некоторой временной задержки $t_{зад}$ снова поступают на вход СМО и μ_2 заявок ограничен и входит в состав СМО.

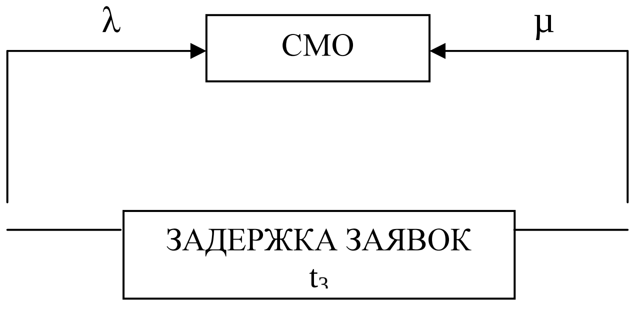


Рисунок 4.4. Модель замкнутой СМО

В замкнутой СМО циркулирует одно и то же конечное число потенциальных заявок. Пока потенциальные заявки циркулируют и не преобразовались на входе СМО в заявку на обслуживание, считается, что они находятся в линии задержки.

Фазы обслуживания разнородных заявок могут осуществляться параллельно или последовательно, одновременно или не одновременно, иметь пересечения в местах схождения потоков заявок.

Типовые варианты СМО определяются также установленной дисциплиной очереди, которая зависит от правил преимущества в обслуживании, т.е. приоритета. Приоритет отбора заявок на обслуживание может быть следующий: первый пришел — первый обслужен; последний пришел — первый обслужен; случайный отбор. Для СМО с ожиданием и обслуживанием по приоритету возможны следующие виды: абсолютный приоритет; относительный приоритет; специальные правила приоритета. Существуют и другие виды СМО: с: поступлением групповых заявок, с каналами разной производительности, с надежными каналами обслуживания, с групповым обслуживанием, со смешанным потоком заявок.

Перечисленные варианты СМО имеют место и распространены в коммерческой деятельности. В своих совокупностях разные типы СМО, последовательно и параллельно объединенные, образуют более сложные многофазные структуры СМО. Моделирование СМО позволяет выявить существенные связи в коммерческой деятельности, а для их описания применить методы и модели теории массового обслуживания и разработать рекомендации по реорганизации, направленные на совершенствование систем массового обслуживания, базируясь на количественном обосновании управленческих решений. Важнейшим началом в этой работе являются замысел и экономико – математическая постановка задач массового обслуживания.

4.4 Экономико-математическая постановка задач массового обслуживания

Правильная или наиболее удачная экономико-математическая постановка задачи в значительной степени определяет полезность рекомендаций по совершенствованию систем массового обслуживания в коммерческой деятельности.

В связи с этим необходимо тщательно проводить наблюдение за процессом обслуживания в системе, поиска и выявления существенных связей, формирования проблемы, выделения цели, определения показателей и выделения экономических критериев оценки работы СМО. В этом случае в качестве наиболее общего, интегрального показателя могут выступать затраты, с одной стороны, СМО коммерческой деятельности как обслуживающей системы, а с другой - затраты заявок, которые могут иметь разную своему физическому содержанию природу.

Повышение эффективности в любой сфере деятельности К. Маркс в конечном счете рассматривал как экономию времени и усматривал в этом один из важнейших экономических законов. Он писал, что экономия времени, равно как и планомерное распределение рабочего времени по различным отраслям производства, остается первым экономическим законом на основе коллективного производства. Этот закон проявляется во всех сферах общественной деятельности.

Для товаров, в том числе и денежных средств, поступающих в коммерческую сферу, критерий эффективности связан со временем и скоростью обращения товаров и определяет интенсивность поступления денежных средств в банк. Время и скорость обращения, являясь экономическими показателями коммерческой деятельности, характеризуют эффективность использования средств, вложенных в товарные запасы. Товарооборачиваемость отражает среднюю скорость реализации среднего товарного запаса. Показатели товарооборачиваемости и уровня запасов тесно связаны известными моделями. Таким образом, можно проследить и установить взаимосвязь этих и других показателей коммерческой деятельности с временными характеристиками.

Следовательно, эффективность работы коммерческого предприятия или организации складывается из совокупности времени выполнения отдельных операций обслуживания, в то же время для населения затраты времени включают время на дорогу, посещение магазина, столовой, кафе, ресторана, ожидание начала обслуживания, ознакомление с меню, выбор продукции, расчет и т.д. Проведенные исследования структуры затрат времени населения свидетельствуют о том, что значительная его часть расходуется нерационально. Коммерческая деятельность в конечном счете направлена на удовлетворение потребности человека. Поэтому усилия моделирования СМО должны включать анализ затрат времени по каждой элементарной операции обслуживания. С помощью соответствующих методов следует создавать модели связи показателей СМО. Это обусловливает необходимость наиболее общие и известные экономические показатели, такие как товарооборот, прибыль, издержки обращения, рентабельность и другие, увязывать в экономико-математических моделях с дополнительно возникающей группой показателей, определяемых спецификой обслуживающих систем и вносимых собственно спецификой теории массового обслуживания.

Например, особенностями показателей СМО с отказами являются: время ожидания заявок в очереди $T_{оч}$=0, поскольку по своей природе в таки системах существование очереди невозможно, то $L_{оч}$=0 и, следовательно, вероятность ее образования $P_{оч}$=0. По числу заявок k определяется режим работы системы, ее состояние: при k=0 – простой каналов, при $1<k< n$ – обслуживание заявок, при $k>n$ – обслуживание и отказ. Показателями таких СМО являются вероятность отказа в обслуживании $P_{отк}$, вероятность обслуживания $P_{обс}$, среднее время простоя канала $t_{пр}$, среднее число занятых $n_з$ и свободных каналов $n_{св}$, среднее время обслуживания $t_{обс}$, абсолютная пропускная способность А.

Для *СМО с неограниченным ожиданием* характерно, что вероятность обслуживания заявки $P_{обс}$=1, поскольку длина очереди и время ожидания начала обслуживания не ограничены, т.е. формально $L_{оч}\to\infty$ и $T_{оч}\to\infty$. В таких системах возможны следующие режимы работы: при k=0 наблюдается простой каналов обслуживания, при $1<k\leq n$ – обслуживание и при $k>n$ – обслуживание и очередь. Показателями эффективности таких СМО являются среднее число заявок в очереди $L_{оч}$, среднее число заявок в системе k, среднее время пребывания заявки в системе $T_{смо}$, абсолютная пропускная способность А.

В *СМО с ожиданием с ограничением на длину очереди*, если число заявок в системе k=0, то наблюдается простой каналов, при $1<k\leq n$ - обслуживание, при $n<k<n+m$ - обслуживание и очередь и при $k>n+m$ - обслуживание, очередь и отказ в ожидании обслуживания. Показателями эффективности таких СМО являются вероятность отказа в обслуживании $P_{\text{отк}}$ - вероятность обслуживания $P_{\text{обс}}$, среднее число заявок в очереди $L_{\text{оч}}$, среднее число заявок в системе $L_{\text{смо}}$ среднее время пребывания заявки в системе $T_{\text{смо}}$, абсолютная пропускная способность А.

Таким образом, перечень характеристик систем массового обслуживания можно представить следующим образом: среднее время обслуживания - $t_{\text{обс}}$, среднее время ожидания в очереди - $T_{\text{оч}}$; среднее время пребывания в СМО - $T_{\text{смо}}$; средняя длина очереди - $L_{\text{оч}}$; среднее число заявок в СМО — $L_{\text{смо}}$; количество каналов обслуживания - n; интенсивность входного потока заявок - λ; интенсивность обслуживания - μ; интенсивность нагрузки - ρ; коэффициент нагрузки - α; относительная пропускная способность - Q; абсолютная пропускная способность - A; доля времени простоя СМО - P_0; доля обслуженных заявок – $P_{\text{обс}}$, доля потерянных заявок – $P_{\text{отк}}$; среднее число занятых каналов - $n_{\text{з}}$; среднее число свободньк каналов — $n_{\text{св}}$; коэффициент загрузки каналов — $К_з$; среднее время простоя каналов – $t_{\text{пр}}$.

Иногда достаточно использовать до десяти основных показателей, чтобы выявить слабые места и разработать рекомендации по совершенствованию СМО.

Это часто связано с решением вопросов согласованной работы цепочки или совокупностей СМО.

Например, в коммерческой деятельности необходимо учитывать еще и экономические показатели СМО: общие затраты – С; издержки обращения – $С_{\text{ио}}$; издержки потребления – $С_{\text{ип}}$; затраты на обслуживание одной заявки – $С_1$; убытки, связанные с уходом заявки, - $С_{\text{у1}}$; затраты на эксплуатацию канала – $С_{\text{к}}$; затраты простоя канала – $С_{\text{пр}}$; капитальные вложения – $С_{\text{кап}}$; приведенные годовые затраты – $С_{\text{пр}}$; текущие затраты – $С_{\text{тек}}$; доход СМО в единицу времени – $Д_1$.

В процессе постановки задач необходимо раскрыть взаимосвязи показателей СМО, которые по своей базовой принадлежности можно разделить на две группы: первая связана с издержками обращения $C_{ио}$, которые определяются числом занятых обслуживанием каналов, затратами на содержание СМО, интенсивностью обслуживания, степенью загрузки каналов, эффективностью их использования, пропускной способностью СМО и др.; вторая группа показателей определяется издержками собственно заявок $C_{ип}$, поступающих на обслуживание, которые образуют входящий поток, ощущают эффективность обслуживания и связаны с такими, показателями, как длина очереди, время ожидания обслуживания, вероятность отказа в обслуживании, время пребывания заявки в СМО и др.

Эти группы показателей противоречивы в том смысле, что улучшение показателей одной группы, например сокращение длины очереди или времени ожидания в очереди путем увеличения числа каналов обслуживания, связано с ухудшением показателей другой группы, поскольку это может привести к увеличению времени простоев каналов обслуживания, затрат на их содержание и т.д. В связи с этим при формализации задач обслуживания вполне естественно стремление построить СМО таким образом, чтобы установить разумный компромисс между показателями собственно заявок и полнотой использования возможностей системы. С этой целью необходимо выбрать обобщенный, интегральный показатель эффективности СМО, включающий одновременно претензии и возможности обеих групп. В качестве такого показателя может быть выбран критерий экономической эффективности, включающий как издержки; обращения $C_{ио}$, так и издержки заявок $C_{ип}$, которые будут иметь оптимальное значение при минимуме общих затрат C. На этом основании целевую функцию задачи можно записать так:

$$C=(C_{ио} + C_{ип})\rightarrow \min$$

Поскольку издержки обращения включают затраты, связанные с эксплуатацией СМО - $C_{экс}$ и простоем каналов обслуживания – $C_{пр}$, а издержки заявок включают потери, связанные с уходом необслуженных заявок – $C_{нз}$ и с пребыванием в очереди - $C_{оч}$, тогда целевую функцию можно переписать с учетом этих показателей таким образом:

$$C = \{(C_{пр}\overrightarrow{n_{св}} + C_{Экз}\overrightarrow{n_З}) + C_{оч}P_{обс}\lambda(T_{0,ч} + \overrightarrow{t_{обс}}) + C_{нзз}P_{отк}\lambda\} \to \min .$$

В зависимости от поставленной задачи в качестве варьируемых, т.е. управляемых, показателей могут быть: количество каналов обслуживания, организация каналов обслуживания (параллельно, последовательно, смешанным образом), дисциплина очереди, приоритетность обслуживания заявок, взаимопомощь между каналами и др. Часть показателей в задаче фигурирует в качестве неуправляемых, которые обычно являются исходными данными. В качестве критерия эффективности в целевой функции могут быть также товарооборот, доход, прибыль или, например, рентабельность, тогда оптимальные значения управляемых , показателей СМО находятся уже при максимизации указанных показателей.

В некоторых случаях следует пользоваться другим вариантом записи целевой функции:

$$C = \{C_{эга} * \overrightarrow{n_З} + C_{пр}(n - \overrightarrow{n_З}) + C_{отк} * P_{отк} * \lambda + C_{сист} * \overrightarrow{n_З}\} \to \min,$$

где $C_{отк}$ и $C_{сист}$ — соответственно издержки заявок, связанные с уходом необслуженных заявок и с пребыванием в системе.

В качестве общего критерия может быть выбран, например, уровень культуры обслуживания покупателей на предприятиях, тогда целевая функция может быть представлена следующей моделью:

$$K_{об} = [(З_{пу} * К_у) + (З_{пв} * К_у) + (З_{пд} * К_д) + (З_{пз} * Кз) + (З_{по} * К_о) ^{+} (З_{кт} * К_{кт})] * К_{мп},$$

где $З_{пу}$ - значимость показателя устойчивости ассортимента товаров;

$К_у$ - коэффициент устойчивости ассортимента товаров;

$З_{пв}$ - значимость показателя внедрения прогрессивных методов продажи товаров;

$К_в$ - коэффициент внедрения прогрессивных методов продажи товаров;

$З_{пд}$ - значимость показателя дополнительного обслуживания;

$К_д$ - коэффициент дополнительного обслуживания;

$З_{пз}$- значимость показателя завершенности покупки;

$К_з$ - коэффициент завершенности покупки;

$З_{по}$ - значимость показателя затрат времени на ожидание в обслуживании;

$К_о$ - показатель затрат времени на ожидание обслуживания;

$З_{кт}$ - значимость показателя качества труда коллектива;

$К_{кт}$ - коэффициент качества труда коллектива;

$К_{мп}$ - показатель культуры обслуживания по мнению покупателей.

Для анализа СМО можно выбирать и другие критерии оценки эффективности работы СМО. Например, в качестве такого критерия для систем с отказами можно выбрать вероятность отказа $P_{отк}$, значение которого не превышало бы заранее заданной величины. Например, требование $Р_{отк} < 0{,}1$ означает, что не менее чем в 90% случаев система должна справляться с обслуживанием потока заявок при заданной интенсивности λ. Можно ограничить среднее время пребывания заявки в очереди или в системе. В качестве показателей, подлежащих определению; могут выступать: либо число каналов *n* при заданной интенсивности обслуживания μ, либо интенсивность μ, при заданном числе каналов.

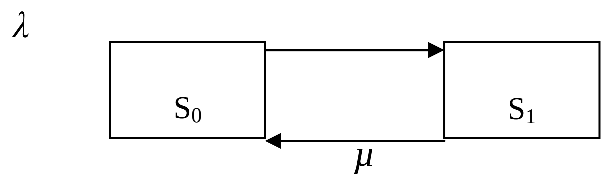
После построения целевой функции необходимо определить условия решения задачи, найти ограничения, установить исходные значения показателей, выделить неуправляемые показатели, построить или подобрать совокупность моделей взаимосвязи всех показателей для анализируемого типа СМО, чтобы в конечном итоге найти оптимальные значения управляемых показателей.

4.5 Модели систем массового обслуживания в коммерческой деятельности

4.5.1 Одноканальная СМО с отказами в обслуживании

Работу одноканальной СМО *n* = 1 можно представить в виде размеченного графа состояний (рисунок 5).

Переходы СМО из одного состояния S_0 в другое S_1 происходят под действием входного потока заявок с интенсивностью λ, а обратный переход- под действием потока обслуживания с интенсивностью μ.



S_0 - канал обслуживания свободен; S_1 - канал занят обслуживанием.

Рисунок 5. Размеченный граф состояний одноканальной СМО

Запишем систему дифференциальных уравнений Колмогорова для вероятностей:

$$\begin{cases} \dfrac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t) \\ \dfrac{dp_1(t)}{dt} = -\mu p_1(t) + \lambda p_0(t) \\ p_0(t) + p_0(t) = 1 \end{cases}$$

Откуда получим дифференциальное уравнение для определения вероятности $p_0(t)$ состояния S_0:

$$\frac{dp_0(t)}{dt} = -(\lambda + \mu)p_0(t) + \mu$$

Это уравнение можно решить при начальных условиях, в предположении, что система в момент $t = 0$ находилась в состоянии S_0, тогда $p_0(0)=1$, $p_1(0) = 0$. В этом случае решение дифференциального уравнения позволяет определить вероятность того; что канал свободен и не занят обслуживанием:

$$p_0(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}$$

Тогда нетрудно получить выражение для определения вероятности занятости канала:

$$p_1(t) = \frac{\lambda}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}$$

Вероятность $p_0(t)$ уменьшается с течением времени и в пределе при $t\to\infty$ стремится к величине

$$p_0 = \frac{\mu}{\lambda + \mu},$$

а вероятность $p_1(t)$ в то же время увеличивается от 0, стремясь в пределе при $t\to\infty$ к величине

$$p_1 = \frac{\lambda}{\lambda + \mu}.$$

Эти пределы вероятностей могут быть получены непосредственно из уравнений Колмогорова при условии

$$\frac{dp_0(t)}{dt} = \frac{dp_1(t)}{dt} = 0 .$$

Функции $p_0(t)$ и $p_1(t)$ определяют переходный процесс в одноканальной СМО и описывают процесс экспоненциального приближения СМО к своему предельному состоянию с постоянной времени $\tau = \frac{1}{\lambda+\mu}$, характерной для рассматриваемой системы.

С достаточной для практики точностью можно считать, что переходный процесс в СМО заканчивается в течение времени, равного 3τ.

Вероятность $p_0(t)$ определяет относительную пропускную способность СМО, которая определяет долю обслуживаемых заявок по отношению к полному числу поступающих заявок, в единицу времени. $P_0(t)$ есть вероятность того, что заявка, пришедшая в момент t, будет принята к обслуживанию. Всего в единицу времени приходит в среднем λ заявок и из них обслуживается λp_0 заявок. Тогда доля обслуживаемых заявок по отношению ко всему потоку заявок определяется величиной

$$Q = \frac{\lambda p_0(t)}{\lambda} = p_0(t)$$

В пределе при $t \to 3\tau$ значение относительной пропускной способности будет равно

$$Q = \frac{\mu}{\lambda+\mu.} .$$

Абсолютная пропускная способность, определяющая число заявок, обслуживаемых в единицу времени в пределе при $t \to \infty$, равна:

$$A = \frac{\lambda * \mu}{\lambda + \mu} = \lambda Q .$$

Соответственно доля заявок, получивших отказ, составляет в этих же предельных условиях

$$P_{\text{отк}} = p_1 = \frac{\lambda}{\lambda + \mu} ,$$

а общее число необслуженных заявок равно

$$\frac{\lambda^2}{\lambda+\mu}.$$

Примерами одноканальных СМО с отказами в обслуживании являются: стол заказов в магазине, диспетчерская автотранспортного предприятия, контора склада, офис управления коммерческой фирмы, с которыми устанавливается связь по телефону.

4.5.2 Многоканальная СМО с отказами в обслуживании

В коммерческой деятельности примерами многоканальных систем массового обслуживания; являются офисы коммерческих предприятий с несколькими телефонными каналами.

Рассмотрим многоканальную СМО с отказами в обслуживании на рисунок 6, на вход которой поступает пуассоновский поток заявок с интенсивностью λ.

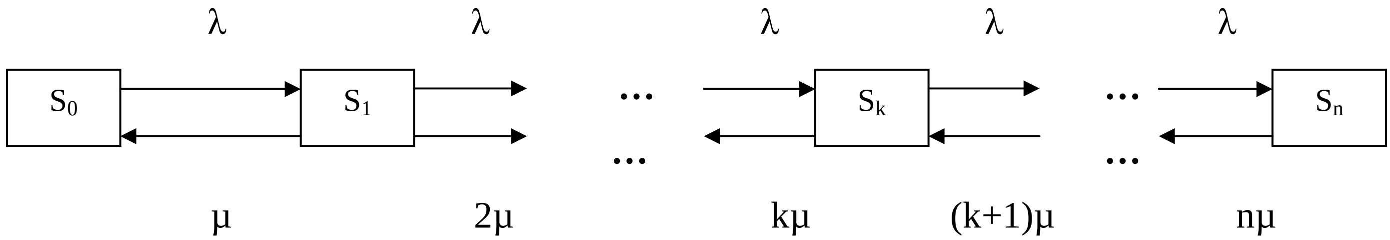


Рисунок 6. Размеченный граф состояний многоканальной СМО с отказами

Поток обслуживания в каждом канале имеет интенсивность μ. По числу заявок СМО определяются ее состояния S_k, представленные в виде размеченного графа:

S_0 — все каналы свободны, k = 0,

S_1 — занят только один канал, k= 1,

S_2 — заняты только два канала, k = 2,

……………………………………………

S_k — заняты k каналов,

……………………………………………

S_n - заняты все n каналов, k = n.

Состояния многоканальной СМО меняются скачкообразно в случайные моменты времени. Переход из одного состояния, например S_0 в S_1, происходит под воздействием входного потока заявок с интенсивностью λ, а обратно — под воздействием потока обслуживания заявок с интенсивностью μ. Для перехода системы из состояния S_k в S_{k-1} безразлично, какой именно из каналов освободится, поэтому поток событий, переводящий СМО, имеет интенсивность kμ, следовательно, поток событий, переводящий систему из S_n в S_{n-1}, имеет интенсивность nμ. Так формулируется классическая задача Эрланга.

Случайный процесс, протекающий в СМО, представляет собой частный случай процесса «рождения—гибели» и описывается системой дифференциальных уравнений Эрланга, которые позволяют получить выражения для предельных вероятностей состояния рассматриваемой системы, называемые формулами Эрланга:

$$P_0 = \left[\sum_{k=0}^{n} \frac{\mathrm{p}^k}{\mathrm{k}!}\right]^{-1};$$

$$P_k = \frac{\rho^k}{k!};$$

$$\rho = \frac{\lambda}{\mu}.$$

Вычислив все вероятности состояний n - канальной СМО с отказами p_0, p_1, p_2, … , p_n, можно найти характеристики системы обслуживания.

Вероятность отказа в обслуживании определяется вероятностью того, что поступившая заявка на обслуживание найдет все n каналов занятыми, система будет находиться в состоянии S_n:

$$p_{\text{отк}} = p_n = p_0 \frac{\rho^n}{n!}; \ \mathrm{k=n}.$$

В системах с отказами события отказа и обслуживания составляют полную группу событий, поэтому

$$P_{\text{отк}} + P_{\text{обс}} = 1.$$

На этом основании относительная пропускная способность определяется по, формуле

$$Q=p_{обс}=1-p_{отк}=1-p_n \,.$$

Абсолютную, пропускную способность СМО можно определить по формуле

$$A=\lambda*p_{обс} \;.$$

Вероятность обслуживания, или доля обслуженных заявок, определяет относительную пропускную способность СМО, которая может быть определена и по другой формуле:

$$Q=p_{обс}=\frac{n_з}{\rho}\,.$$

Из этого выражения можно определить среднее число заявок, находящихся под обслуживанием, или, что то же самое, среднее число занятых обслуживанием каналов

$$\underset{n_з}{\rightarrow}=\rho*p_{обс}=\frac{A}{\mu}\,.$$

Коэффициент занятости каналов обслуживанием определяется отношением среднего числа занятых каналов к их общему числу

$$К_з=\frac{\overrightarrow{n_з}}{n}=\frac{\rho}{n}*p_{обс}$$

Вероятность занятости каналов обслуживанием, которая учитывает среднее время занятости $\underset{t_{зан}}{\rightarrow}$ и простоя $\underset{t_{пр}}{\rightarrow}$ каналов, определяется следующим образом:

$$p_{зан}=\frac{t_з}{\underset{t_з}{\rightarrow}+\underset{t_{пр}}{\rightarrow}}\,.$$

Из этого выражения можно определить среднее время простоя каналов

$$t_{пр}=\underset{t_{обс}}{\rightarrow}\frac{1-p_{зан}}{p_{зан}}\,.$$

Среднее время пребывания заявки в системе в установившемся режиме определяется формулой Литтла

$$T_{смо}=\frac{\overrightarrow{n_з}}{\lambda}\,.$$

4.5.3 Одноканальная СМО с ограниченной длиной очереди

В коммерческой деятельности чаще встречаются СМО с ожиданием (очередью).

Рассмотрим простую одноканальную СМО с ограниченной очередью, в которой число мест в очереди *m* — фиксированная величина. Следовательно, заявка, поступившая в тот момент, когда все места в очереди заняты, не принимается к обслуживанию, не встает в очередь и покидает систему.

Граф простой одноканальной СМО с ограниченной очередью совпадает с графом рисунка 2, описывающим процесс «рождения—гибели», с тем отличием, что при наличии только одного канала обслуживания все интенсивности потоков обслуживания равны.

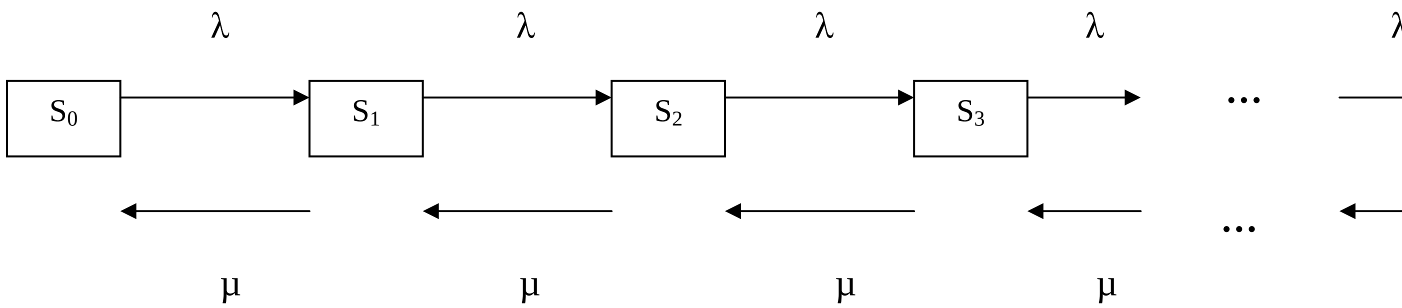


Рисунок 7. Размеченный граф процесса «рождения — гибели»

Состояния СМО можно представить следующим образом:

S_0 - канал обслуживания свободен,

S_1 - канал обслуживания занят, но очереди нет,

S_2 - канал обслуживания занят, в очереди стоит одна заявка,

S_3 - канал обслуживания занят, в очереди стоят две заявки,

……………………………………………….

S_{m+1} - канал обслуживания занят, в очереди все *m* мест заняты, любая следующая заявка получает отказ.

Для описания случайного процесса СМО можно воспользоваться изложенными ранее правилами и формулами. Напишем выражения, определяющие предельные вероятности состояний:

$$\begin{cases} p_1 = \rho * p_0 \\ p_2 = \rho^2 * p_0 \\ p_k = \rho^k * p_0 \\ \dots\dots\dots\dots\dots \\ p_{m+1} = \rho^{m+1} * p_0 \\ p_0 = [1 + \rho + \rho^2 + \rho^3 + \dots + \rho^{m+1}]^{-1} \end{cases}$$

Выражение для p_0 можно в данном случае записать проще, пользуясь тем, что в знаменателе стоит геометрическая прогрессия относительно ρ, тогда после соответствующих преобразований получаем:

$$p_0 = \frac{(1-\rho)}{(1-\rho^{m+2})}$$

Эта формула справедлива для всех ρ, отличных от 1, если же ρ= 1, то p_0 = 1/(m + 2), а все остальные вероятности также равны $1/(m + 2)$. Если предположить $m = 0$, то мы переходим от рассмотрения одноканальной СМО с ожиданием к одноканальной СМО с отказами в обслуживании. Выражение для предельной вероятности p_0 в случае $m = 0$ имеет вид:

$$p_0 = \frac{\mu}{\lambda + \mu}$$

и в случае λ=μ имеет величину p_0=1/2.

Заявка получает отказ, если она поступает в момент времени, когда СМО уже находится в состоянии S_{m+1} и следовательно, все места в очереди m заняты и один канал обслуживает. Поэтому вероятность отказа определяется вероятностью появления состояния S_{m+1}:

$$p_{отк} = p_{m+1} = \rho^{m+1} * p_0 .$$

Относительная пропускная способность, или доля обслуживаемых заявок, поступающих в единицу времени, определяется выражением

$$Q = 1 - p_{отк} = 1 - \rho^{m+1} * p_0 ;$$

абсолютная пропускная способность равна:

$$A = Q * \lambda .$$

Среднее число заявок $L_{оч}$, стоящих в очереди на обслуживание, определяется математическим ожиданием случайной величины k – числа заявок, стоящих в очереди:

$$L_{оч} = M(k) .$$

Случайная величина k принимает следующие только целочисленные значения:

1- в очереди стоит одна заявка,

2- в очереди две заявки,

…………………………………..

m - в очереди все места заняты.

Вероятности этих значений определяются соответствующими вероятностями состояний, начиная с состояния S_2. Закон распределения дискретной случайной величины k изображается следующим образом:

k	1	2	…	m
p_1	p_2	p_3	…	p_{m+1}

Математическое ожидание этой случайной величины равно:

$$L_{оч} = 1 * p_2 + 2 * p_3 + \cdots + m * p_{m+1}.$$

В общем случае при ρ≠1 эту сумму можно преобразовать, пользуясь моделями геометрической прогрессии, к более удобному виду:

$$L_{оч} = \rho^2 * \frac{1 - \rho^m * (m - m * \rho + 1)}{(1-\rho)^2} * p_0.$$

В частном случае при ρ=1, когда все вероятности p_k оказывается равными, можно воспользоваться выражением для суммы членов числового ряда

$$1 + 2 + 3 + \cdots + m = \frac{m(m+1)}{2}.$$

Тогда получим формулу

$$L'_{оч} = \frac{m(m+1)}{2} * p_0 = \frac{m(m+1)}{2}.$$

Применяя аналогичные рассуждения и преобразования, можно показать, что среднее время ожидания обслуживания заявки в очереди определяется формулами Литтла

$$T_{оч} = \frac{L_{оч}}{A} \text{ (при } \rho \neq 1) \text{ и } T^1_{оч} = \frac{L'_{оч}}{A} \text{ (при } \rho = 1).$$

Такой результат, когда оказывается, что $T_{оч} \sim 1/\lambda$, может показаться странным: с увеличением интенсивности потока заявок как будто бы должна возрастать длина очереди и уменьшается среднее время ожидания. Однако следует иметь в виду, что, во - первых, величина $L_{оч}$ является функцией от λ и μ и, во-вторых, рассматриваемая СМО имеет ограниченную длину очереди не более m заявок.

Заявка, поступившая в СМО в момент времени, когда все каналы заняты, получает отказ, и, следовательно, время ее «ожидания» в СМО равно нулю. Это приводит в общем случае (при $\rho \neq 1$) к уменьшению $T_{оч}$ ростом λ, поскольку доля таких заявок с ростом λ увеличивается.

Если отказаться от ограничения на длину очереди, т.е. устремить $m \to \infty$, то случаи $\rho < 1$ и $\rho \geq 1$ начинают существенно различаться. Записанные выше формулы для вероятностей состояний преобразуются в случае $\rho < 1$ к виду

$$\begin{cases} p_0 = 1 - p \\ p_1 = \rho * (1 - \rho) \\ p_2 = \rho^2 * (1 - \rho) \\ \dots\dots\dots\dots\dots\dots \\ p_k = \rho^k * (1 - \rho) \end{cases}.$$

При достаточно большом k вероятность p_k стремится к нулю. Поэтому относительная пропускная способность будет $Q = 1$, а абсолютная пропускная способность станет равной $A = \lambda\, Q = \lambda$, следовательно, обслуживаются все поступившие заявки, причем средняя длина очереди окажется равной:

$$L_{оч} = \frac{\rho^2}{1 - \rho},$$

а среднее время ожидания по формуле Литтла

$$T_{оч} = \frac{L_{оч}}{A}.$$

В пределе $p << 1$ получаем $T_{оч} = \rho/\mu$, т.е. среднее время ожидания быстро уменьшается с увеличением интенсивности потока обслуживания. В противном случае при $\rho \geq 1$ оказывается, что в СМО отсутствует установившийся режим. Обслуживание не успевает за потоком заявок, и очередь неограниченно растет со временем (при $t \to \infty$). Предельные вероятности состояний поэтому не могут быть определены: при $Q = 1$ они равны нулю. Фактически СМО не выполняет своих функций, поскольку она не в состоянии обслужить все поступающие заявки. Нетрудно определить, что доля обслуживаемых заявок и абсолютная пропускная способность соответственно составляют в среднем ρ и μ, однако неограниченное увеличение очереди, а следовательно, и времени , ожидания в ней приводит к тому, что через некоторое время заявки начинают накапливаться в очереди на неограниченно долгое время.

В качестве одной из характеристик СМО используют среднее время $T_{смо}$ пребывания заявки в СМО, включающее среднее время пребывания в очереди и среднее время обслуживания. Эта величина вычисляется по формулам Литтла:

если длина очереди ограничена - среднее число заявок, находящихся в очереди, равно:

$$L_{смо} = \frac{m+1}{2};$$

$$T_{смо} = \frac{L_{смо}}{A} \quad \text{при } \rho \neq 1,$$

тогда среднее время пребывания заявки в системе массового обслуживания (как в очереди, так и под обслуживанием) равно:

$$T_{смо} = \frac{m+1}{2\mu} \quad \text{при } \rho = 1.$$

4.5.4 Одноканальная СМО с неограниченной очередью

В коммерческой деятельности в качестве одноканальной СМО с неограниченным ожиданием является, например, коммерческий директор, поскольку он, как правило, вынужден выполнять обслуживание заявок различной природы: документы, переговоры по телефону, встречи и беседы с подчиненными, представителями налоговой инспекции, милиции, товароведами, маркетологами, поставщиками продукции и решать задачи в товарно-финансовой сфере с высокой степенью финансовой ответственности, что связано с обязательным выполнением запросов, которые ожидают иногда нетерпеливо выполнения своих требований, а ошибки неправильного обслуживания, как правило, экономически весьма ощутимы.

В то же время товары, завезенные для продажи (обслуживания), находясь на складе, образуют очередь на обслуживание (продажу). Длину очереди составляет количество товаров, предназначенных для продажи. В этой ситуации продавцы выступают в роли каналов, обслуживающих товары. Если количество товаров, предназначенных для продажи, велико, то в этом случае мы имеем дело с типичным случаем СМО с ожиданием. Рассмотрим простейшую одноканальную СМО с ожиданием обслуживания, на которую поступает пуассоновский поток заявок с интенсивностью λ и интенсивностью обслуживания μ. Причем заявка, поступившая в момент, когда канал занят обслуживанием, ставится в очередь и ожидает обслуживания. Размеченный граф состояний такой системы приведен на рисунке 8.

Количество возможных состояний ее бесконечно:

S_0 - канал свободен, очереди нет, $k = 0$;

S_1 - канал занят обслуживанием, очереди нет, $k = 1$;

S_2 - канал занят, одна заявка в очереди, $k = 2$;

……………………………………………

S_k - канал занят $(k - 1)$, заявка в очереди.

Модели оценки вероятности состояний СМО с неограниченной очередью можно получить из формул, выведенных для СМО с ограниченной очередью, путем перехода к пределу при $m \to \infty$:

Рисунок 8. Граф состояний одноканальной СМО с неограниченной очередью

$$\begin{cases} p_0 = \dfrac{1-p}{1-p^{m+2}} \\ p_1 = \rho * p_0 \\ p_2 = \rho^2 * p_0 \\ \dots\dots\dots\dots\dots \\ p_k = \rho^k * p_0 \\ \dots\dots\dots\dots\dots \end{cases} \Rightarrow \begin{cases} p_0 = 1-\rho \\ p_1 = \rho(1-\rho) \\ p_2 = \rho^2(1-\rho) \\ \dots\dots\dots\dots\dots \\ p_k = \rho^k(1-\rho) \end{cases}$$

Следует заметить, что для СМО с ограниченной длиной очереди в формуле

$$p_0 = [1 + \rho + \rho^2 + \rho^3 + \dots + \rho^k + \dots + \rho^{m+1}]^{-1}$$

имеет место геометрическая прогрессия с первым членом 1 и знаменателем ρ. Такая последовательность представляет собой сумму бесконечного числа членов при $m \to \infty$. Эта сумма сходится, если прогрессия, бесконечно убывающая при ρ<1, что определяет установившийся режим работы СМО, а при ρ>1 очередь при $t \to \infty$ с течением времени может расти до бесконечности.

Поскольку в рассматриваемой СМО ограничение на длину очереди отсутствует, то любая заявка может быть обслужена, поэтому $p_{обс}=1$, следовательно, относительная пропускная способность $Q=p_{обс}=1$, соответственно $p_{отк}=0$, а абсолютная пропускная способность $A=\lambda * Q=\lambda$, $L_{об} = \rho$.

Вероятность пребывания в очереди k заявок равна:

$$p_k = \rho^k (1-\rho);$$

среднее число заявок в очереди –

$$L_{оч} = \frac{\rho^2}{1-\rho};$$

среднее число заявок в системе –

$$L_{смо} = L_{оч} + L_{об} = \frac{\rho^2}{1-\rho} + \rho = \frac{\rho}{1-\rho};$$

среднее время ожидания обслуживания в очереди –

$$T_{оч} = \frac{\rho}{\mu(1-\rho)} = \frac{\rho^2}{\lambda(1-\rho)};$$

среднее время пребывания заявки в системе –

$$T_{смо} = T_{оч} + t_{обс} = \frac{1}{\mu(1-\rho)} = \frac{\rho}{\lambda(1-\rho)}.$$

Если в одноканальной СМО с ожиданием интенсивность поступления заявок больше интенсивности обслуживания $\lambda > \mu$, то очередь будет постоянно увеличиваться. В связи с этим наибольший интерес представляет анализ устойчивых СМО, работающих в стационарном режиме при $\lambda < \mu, \rho < 1$.

4.5.5 Многоканальная СМО с ограниченной длиной очереди

Рассмотрим многоканальную СМО ($n > 1$), на вход которой поступает пуассоновский поток заявок с интенсивностью λ, а интенсивность обслуживания каждого канала составляет μ, максимально возможное число мест в очереди ограничено величиной m. Дискретные состояния СМО определяются количеством заявок, которые можно записать.

S_0 - все каналы свободны, k=0;

S_1 - занят только один канал (любой), $k=1$;

S_2 - заняты только два канала (любых), k=2;

……………………………………

S_n - заняты все n каналов, $k=n$.

Пока СМО находится в любом из этих состояний, очереди нет. После того как заняты все каналы обслуживания, последующие заявки образуют очередь, тем самым определяя дальнейшее состояние системы:

S_{n+1} - заняты все n каналов и одна заявка стоит в очереди, $k = n+1$;

S_{n+2} - заняты все n каналов и две заявки стоят в очереди, $k = n+2$;

……………………………………..

S_{n+m} - заняты все n каналов и все m мест в очереди, $k = n+m$.

Граф состояний n – канальной СМО с очередью, ограниченной m местами, представлен на рисунке 9.

Рисунок 9. Граф состояний n-канальной СМО с ограничением но длину очереди m

Переход СМО в состояние с большими номерами определяется потоком поступающих заявок с интенсивностью λ, когда как по условию в обслуживании этих заявок принимает участие n одинаковых каналов с интенсивностью потока обслуживания равного μ для каждого канала. При этом полная интенсивность потока обслуживания возрастает с подключением новых каналов вплоть до такого состояния S_n, когда все n каналов окажутся занятыми. С появлением очереди интенсивность обслуживания более не увеличивается, так как она уже достигла максимального значения, равного $n\mu$.

Запишем выражения для предельных вероятностей состояний:

$$\begin{cases} p_1 = \dfrac{\rho}{1!} * p_0 ; \\ p_2 = \dfrac{\rho^2}{2!} * p_0 \dots \dots ; \\ p_k = \dfrac{\rho^k}{k!} * p_0 , 0 \le k \le n ; \\ p_n = \dfrac{\rho^n}{n!} * p_0 ; \end{cases}$$

$$\begin{cases} p_{n+1} = \dfrac{\rho^{n+1}}{n * n!} * p_0 \,; \\ p_{n+2} = \dfrac{\rho^{n+2}}{n^2 * n!} * p_0 \ldots\ldots; \\ p_{n+r} = \dfrac{\rho^{n+r}}{n^r * n!} * p_0 \,, n \le r \le m \,; \\ p_{n+m} = \dfrac{\rho^{n+m}}{n^m * n!} * p_0 = p_{\text{отк}} \,; \end{cases}$$

$$p_0 = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \cdots \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n * n!} + \frac{\rho^{n+2}}{n^2 * n!} + \cdots + \frac{\rho^m}{n^m * n!}\right]^{-1}.$$

Выражение для p_0 можно преобразовать, используя формулу геометрической прогрессии для суммы членов со знаменателем ρ/n:

$$p_0 = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \cdots + \frac{\rho^n}{n!} + \frac{\rho^n * \frac{\rho}{n} - (\frac{\rho}{n})^{m+1}}{n!\,(1 - \frac{\rho}{n})}\right]^{-1} k = 0 \,;$$

$$p_0 \left[\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^n}{n!} * \frac{\frac{\rho}{n} - (\frac{\rho}{n})^{m+1}}{1 - \rho/n}\right]^{-1} = \left[\sum_{k=0}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!\,(n - \rho)} * (1 - (\frac{\rho}{n}))^m\right]^{-1}.$$

Образование очереди возможно, когда вновь поступившая заявка застанет в системе не менее n требований, т.е. когда в системе будет находиться n, $n+1$; $n+2$,..., $(n+m-1)$ требований. Эти события независимы, поэтому вероятность того, что все каналы заняты, равна сумме соответствующих вероятностей p_n, p_{n+1}, p_{n+2},…, p_{n+m-1}. Поэтому вероятность образования очереди равна:

$$p_{\text{оч}} = \sum_{k=n}^{n+m-1} p_k = \frac{\rho^n}{n!} * \frac{1 - (\frac{\rho}{n})^m}{1 - \rho/n} * p_0 \,.$$

Вероятность отказа в обслуживании наступает тогда, когда все n каналов и все m мест в очереди заняты:

$$p_{\text{отк}} = p_{n+m} = \frac{\rho^{n+m}}{n^m * n!} * p_0 \,; k = n + m \,.$$

Относительная пропускная способность будет равна:

$$Q = p_{\text{обс}} = 1 - p_{\text{отк}} \,;$$

абсолютная пропускная способность –

$$A = \lambda * Q \,;$$

среднее число занятых каналов –

$$\overrightarrow{n_з} = \frac{A}{\mu} = \rho * Q \text{ ;}$$

среднее число простаивающих каналов –

$$\overrightarrow{n_{пр}} = n - \overrightarrow{n_з};$$

коэффициент занятости (использования) каналов –

$$K_з = \frac{\overrightarrow{n_з}}{n} \text{ ;}$$

коэффициент простоя каналов –

$$K_{пр} = 1 - K_з \text{ ;}$$

среднее число заявок, находящихся в очередях, -

$$L_{оч} = \frac{\rho^{n+1}}{n * n!} * \frac{1 - (\frac{\rho}{n})^m * (m + 1 - m * \frac{\rho}{n})}{(1 - \frac{\rho}{n})^2} p_0 \text{ ;}$$

в случае если $\rho/n = 1$, эта формула принимает другой вид –

$$L^1_{оч} = \frac{\rho^{n+1}}{n * n!} * \frac{m * (m + 1)}{2} * p_0 \text{ ;}$$

среднее время ожидания в очереди определяется формулами Литтла –

$$T_{оч} = \frac{L_{оч}}{A} \left(\frac{\rho}{n \neq 1}\right);$$

$$T^1_{оч} = L^1_{оч} \left(\frac{\rho}{т = 1}\right).$$

Среднее время пребывания заявки в СМО, как и для одноканальной СМО, больше среднего времени ожидания в очереди на среднее время обслуживания, равное $1/\mu$, поскольку заявка всегда обслуживается только одним каналом:

$$T_{смо} = T_{оч} + \frac{1}{\mu} \left(\frac{\rho}{n \neq 1}\right).$$

4.6 Многоканальная СМО с неограниченной очередью

Рассмотрим многоканальную СМО с ожиданием и неограниченной длиной очереди, на которую поступает поток заявок с интенсивностью λ и которая имеет интенсивность обслуживания каждого канала μ. Размеченный граф состояний представлен на рисунке 10. Он имеет бесконечное число состояний:

S_0 - все каналы свободны, $k = 0$;

S_1 - занят один канал, остальные свободны, $k = 1$;

S_2 - заняты два канала, остальные свободны, $k = 2$;

……………………………………………….

S_n - заняты все n каналов, $k = n$, очереди нет;

S_{n+1} - заняты все n каналов, одна заявка в очереди, $k= n + 1$;

S_{n+r} - заняты все n каналов, r заявок в очереди, $k= n + r$.

……………………………………………….

Вероятности состояний получим из формул для многоканальной СМО с ограниченной очередью при переходе к пределу при $m\to\infty$. Сумма геометрической прогрессии в выражении для p_0 расходится при уровне загрузки $p/n>1$, очередь будет бесконечно возрастать, а при $p/n<1$ ряд сходится, что определяет установившийся стационарный режим работы СМО,

Рисунок 10. Размеченный граф состояний многоканальной СМО с неограниченной очередью

для которого и определим выражения для предельных вероятностей состояний:

$$p_0 = \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \cdots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n!\,(n-\rho)}\right]^{-1};$$

$$p_1 = \frac{\rho}{1!} * p_0\,;\ p_2 = \frac{\rho^2}{2!} * p_0\,;\ \ldots;\ p_n = \frac{\rho^n}{n!} * p_0\,;$$

$$p_{n+1} = \frac{\rho^{n+1}}{n * n!} * p_0\,;\ p_{n+2} = \frac{\rho^{n+2}}{n^2 * n!} * p_0\,;\ \ldots;\ p_{n+r} = \frac{\rho^{n+r}}{n^r * n!} * p_0$$

Поскольку отказа в обслуживании в таких системах не может быть, то характеристики пропускной способности равны:

$$p_{\text{отк}} = 0\,; Q = 1; A = \lambda * Q = \lambda\,;$$

среднее число заявок в очереди –

$$L_{\text{оч}} = \frac{\rho^{n+1}}{n * n! * \left(1 - \frac{\rho}{n}\right)^2} * p_0\,;$$

среднее время ожидания в очереди –

$$T_{\text{оч}} = \frac{\rho^n}{n * \mu * n! * \left(1 - \frac{\rho}{n}\right)^2} * p_0 = \frac{L_{\text{оч}}}{\lambda}\,;$$

среднее число заявок в СМО –

$$L_{\text{смо}} = L_{\text{оч}} + \rho \ ; \ T_{\text{смо}} = \frac{L_{\text{смо}}}{\lambda}.$$

Вероятность того, что СМО находится в состоянии S_0, когда нет заявок и не занято ни одного канала, определяется выражением

$$p_0 = \left[\sum_{k=1}^{n} \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!\,(n-\rho)}\right]^{-1}, k = 0.$$

Эта вероятность определяет среднюю долю времени простоя канала обслуживания.

Вероятность занятости обслуживанием k заявок –

$$p_k = \frac{\rho^k}{k!} * p_0 , 1 \le k \le n.$$

На этом основании можно определить вероятность, или долю времени занятости всех каналов обслуживанием

$$p_n = \frac{\rho^n}{n!} * p_0 , k = n.$$

Если же все каналы уже заняты обслуживанием, то вероятность состояния определяется выражением

$$p_{n+r} = \frac{\rho^{n+r}}{n! * n^r} * p_0 = p_n * \left(\frac{\rho}{n}\right)^r , k > n.$$

Вероятность оказаться в очереди равна вероятности застать все каналы уже занятыми обслуживанием.

$$p_0 = \frac{\rho^{n+1}}{n!\,(n-\rho)} * p_0 , k \ge n.$$

Среднее число заявок, находящихся в очереди и ожидающих обслуживания, равно:

$$L_{\text{оч}} = \frac{n}{n-\rho} * p_{\text{оч}} \ ;$$

среднее время ожидания заявки в очереди по формуле Литтла:

$$T_{\text{оч}} = \frac{L_{\text{оч}}}{A}$$

и в системе

$$T_{\text{смо}} = \frac{L_{\text{смо}}}{A} \ ;$$

среднее число занятых каналов обслуживанием:

$$\overrightarrow{n_з} = \frac{\lambda}{\mu} = \rho \, ;$$

среднее число свободных каналов:

$$\overrightarrow{n_{св}} = n - \rho \, ;$$

коэффициент занятости каналов обслуживанием:

$$K_з = \frac{\overrightarrow{n_з}}{n} = \frac{\rho}{n} \, ;$$

Важно заметить, что параметр ρ характеризует степень согласования входного потока, например покупателей в магазине с интенсивностью потока обслуживания. Процесс обслуживания будет стабилен при $\rho < n$. Если же $\rho \geq n$, в системе будут возрастать средняя длина очереди и среднее время ожидания покупателями начала обслуживания и, следовательно, СМО будет работать неустойчиво.

4.7 Дополнительный материал

ОСНОВНЫЕ ЭЛЕМЕНТЫ СМО

Основными элементами СМО являются источники заявок, их входящий поток, каналы обслуживания и выходящий поток.

Это можно изобразить так:

Входящий по ⟶ *Очер* ⟶ *Каналы* ⟶ *:ивания*

Выходящий поток

В зависимости от формирования СМО, ее различают на:

1. Системы с отказами, в которых при занятости, заявка не встает в очередь и при это она по-кидает систему обслуживания.
2. Системы с неограниченными ожиданиями, в которых заявка встает в очередь, если в мо-мент поступления все каналы были заняты.

Существуют и системы смешанного типа с ожиданием ограниченной длиной очереди, это когда заявка получает отказ, если приходит в момент, когда все

места в очереди заняты. Заявка, попавшая в очередь, обязательно обслуживается.
По числу каналов обслуживания СМО делятся на одноканальные и многоканальные.
В зависимости от расположения источника требований системы могут быть разомкнутыми, (источник заявок находится вне системы) и замкнутыми (источник находится в самой системе).

ХАРАКТЕРИСТИКИ СМО

Перечень характеристик систем массового обслуживания можно представить следующим образом:

1. Среднее время обслуживания;
2. Среднее время ожидания в очереди;
3. Среднее время пребывания в СМО;
4. Средняя длина очереди;
5. Среднее число заявок в СМО;
6. Количество каналов обслуживания;
7. Интенсивность входного потока заявок;
8. Интенсивность обслуживания;
9. Интенсивность нагрузки;
10. Коэффициент нагрузки;
11. Относительная пропускная способность;
12. Абсолютная пропускная способность;
13. Доля времени простоя СМО;
14. Доля обслуженных заявок;
15. Доля потерянных заявок;
16. Среднее число занятых каналов;
17. Среднее число свободных каналов;
18. Коэффициент загрузки каналов;
19. Среднее время простоя каналов;

ПАРАМЕТРЫ СТРУКТУРЫ СМО

Каждая система массового обслуживания обладает определенной структурой, характери-зующейся совокупностью параметров. Основным компонентом структуры СМО являются А-налы обслуживания. В зависимости от числа каналов различают одноканальные и многока-нальные СМО.
Простейшей одноканальной моделью с вероятностным входным потоком и процедурой обслуживания является модель, характеризуемая показательным распределением как длитель-ностей интервалов между поступлениями требо-

ваний, так и длительностей обслуживания. При этом плотность распределения длительностей интервалов между поступлениями требований имеет вид

$f_1(t)=\lambda e^{(-\lambda t)}$, где λ - интенсивность поступления заявок в систему.

Под интенсивностью потока понимают

$\lambda(t)=\lim\limits_{\tau\to 0}\dfrac{m(t,t+\tau)}{\tau}$, где m(t, t+τ) – среднее число событий в интервале (t, t+τ).

Плотность распределения длительностей обслуживания:

$f_2(t)=\mu e^{(-\mu t)}$, где μ - интенсивность обслуживания.

Поток заявок и обслуживания простейшие, т.е. обладающие свойствами стационарности (среднее число событий, воздействующих на систему, в течение единицы времени, остается пос-тоянным), ординарности (вероятность попадания на элементарный участок времени двух и более событий пренебрежимо мала) и отсутствия последействия (для любых непересекающихся участ-ков времени количество событий, попадающих на один из них, не зависит от того, сколько со-бытий попало на другие участки времени).

Для простейшего потока интенсивность λ = const.

Пусть система работает с отказами. Необходимо определить абсолютную и относительную пропускную способности системы. Система имеет два состояния: S_0 – канал свободен и S_1 – канал занят. Обозначим вероятности состояний: $P_0(t)$ – вероятность состояния S_0, $P_1(t)$ – вероятность состояния S_1. Составим систему урав-нений Колмогорова:

$$\begin{cases} \dfrac{\partial}{\partial t}P_0(t)=-\lambda P_0(t)+\mu P_1(t) \\ \dfrac{\partial}{\partial t}P_0(t)=-\lambda P_0(t)+\mu P_1(t) \end{cases},$$

С учетом того, что $P_0(t)+P_1(t)=1$, решение системы такое:

$$\begin{cases} P_0(t)=\dfrac{\lambda e^{(-t(\lambda+\mu))}}{\lambda+\mu}+\dfrac{\mu}{\lambda+\mu}; \\ P_1(t)=1-P_0(t) \end{cases}$$

Для 1-канальной СМО с отказами вероятность $P_0(t)$ есть не что иное, как относительная пропускная способность системы $q: q=P_0(t)$

По истечении большого интервала времени (при $t>\infty$) достигается стационарный ре-жим:

$$P_0=\frac{\mu}{\lambda+\mu}.$$

Абсолютная пропускная способность (А) – среднее число заявок, которое может обслужить СМО в единицу времени:

$$A=\lambda\cdot P_0 \quad \text{или} \quad A=\frac{\lambda\mu}{\lambda+\mu}.$$

Вероятность отказа в обслуживании заявки будет равна вероятности состояния «канал за-нят»:

$$P_{otk} = 1 - P_0.$$

В свою очередь, многоканальные СМО могут содержать одинаковые и различные по производительности каналы обслуживания.

Производительность канала обслуживания обратно длительности обслуживания заявки равной промежутку времени, необходимому каналу обслуживания для обслуживания заявки. В общем случае это случайная величина с функцией распределения $f(t_{обс})$, плотностью распреде-ления $f(t_{обс})$ и математическим ожиданием $\bar{t}_{обс}$. Типы заявок различаются либо законами рас-пределения, либо только математическими ожиданиями при одинаковых законах распределения. При этом принимается допущение о независимости длительностей обслуживания для различных заявок одного типа, вполне корректное для большинства реальных систем. Наряду с математи-ческим ожиданием длительности обслуживания используется понятие интенсивности потока обслуживания $\mu = \frac{1}{\bar{t}_{обс}}$.

Если в момент появления заявки на входе в СМО хотя бы один канал свободен от обслу-живания, ее обслуживание будет начато немедленно. Однако вполне вероятна ситуация, когда все m каналы заняты обслуживанием. В этом случае обслуживание заявки задерживается, и заяв-ка встает в очередь. Таким образом, вторым важным компонентом является очередь, параметром которой является число мест в очереди n. В приоритетных системах очередь может быть разделена на несколько очередей, для каждой из которых должно быть указанно число мест. На число мест в очереди может быть наложено ограничение, это может быть сделано для каждой очереди в отдельности, так и на все очереди в целом. При этом могут быть конфликтные ситуации. При этом на заявку может быть отказ системы принять заявку.

В зависимости от числа мест в очереди различают СМО с отказами, и, соответственно, СМО без отказов. В СМО с отказами число мест в очереди конечно и вследствие вероятностного харак-тера как входящего потока, так и процессов обслуживания, существует ненулевая вероятность то-го, что поступившая на вход СМО заявка застанет все каналы занятыми обслуживанием и все мес-та в очереди занятыми, то есть, она получит отказ. В СМО без отказов заявка либо сразу назна-чается на обслуживание, если в момент ее поступления свободен хотя бы один канал обслужива-ния, либо безусловно принимается в очередь на обслуживание.

ВХОДЯЩИЙ ПОТОК

Входящий поток наиболее распространен в практике это простейший поток заявок, обладающий свойствами стационарности, ординарности и отсутствия последствия.

1. Свойством стационарности, которое выражает неизменность вероятностного режима потока по времени. Это значит, что число требований, посту-

пающих в систему в равные проме-жутки времени, в среднем должно быть постоянным. Например, число вагонов, поступающих под погрузку в среднем в сутки должно быть одинаковым для различных периодов времени.

2. Отсутствие последствия, которое обслуживает взаимную независимость поступления того или иного числа требований на обслуживание в непересекающиеся промежутки времени. Это значит, что число требований, поступающих в данный отрезок времени, не зависит от числа требований, обслуженных в предыдущем промежутке времени.

3. Свойством ординарности, которое выражает практическую невозможность одновре-менного поступления двух или более требований (вероятность такого события неизмеримо у-ла по отношению к рассматриваемому промежутку времени, когда последний устремляют к нулю).

В простейшем потоке заявок позволяет получать характеристики СМО от параметров вхо-дящего потока, что затруднено для других видов потоков заявок. Если СМО обеспечивает желаемую эффективность функционирования системы при прос-тейшем потоке, то обслуживание системой других потоков заявок с одинаковой интенсивностью будет выполнено не хуже.

Входящий поток требований представляет собой совокупность требований, которые посту-пают в систему и нуждаются в обслуживании. Входящий поток требований изучается с целью уста-новления закономерностей этого потока и дальнейшего улучшения качества обслуживания.

В большинстве случаев входящий поток неуправляем и зависит от ряда случайных фак-торов. Число требований, поступающих в единицу времени, случайная величина. Случайной вели-чиной является также интервал времени между соседними поступающими требованиями. Однако среднее количество требований, поступивших в единицу времени, и средний интервал времени между соседними поступающими требованиями предполагаются заданными.

Среднее число требований, поступающих в систему обслуживания за единицу времени, называется интенсивностью поступления требований.

На практике условия простейшего потока не всегда строго выполняются. Часто имеет мес-то нестационарность процесса (в различные часы дня и различные дни месяца поток требований может меняться, он может быть интенсивнее утром или в последние дни месяца). ущеествует так-же наличие последствия, когда количество требований на отпуск товаров в конце месяца зависит от их удовлетворения в начале месяца.

ВЫХОДЯЩИЙ ПОТОК

Совокупность обслуженных и потерянных заявок образует выходящий поток СМО.

В зависимости от структуры выходящего потока различают СМО без потерь и СМО с поте-рями. Для СМО без потерь характерно отсутствие ограничений на

число мест в очереди и на вре-мя пребывания заявки в системе. По этой причине выходящий поток будет состоять лишь из обслуженных заявок.
В свою очередь, поток потерянных заявок может состоять из потока заявок, получивших отказ, и потока не «терпеливых» заявок, покинувших систему, так как их время пребывания пре-высило допустимую величину.
Выходящий поток в общем случае распадается на поток обслуженных и поток потерянных заявок.
Выходящий поток заявок связан с потоком обслуживания в канале, где длительность обслуживания $t_{обс}$ является случайной величиной и подчиняется закону распределения с плотностью

$f\left(t_{обс}\right)=\mu e^{-\mu t}$, где μ – интенсивность потока обслуживания, т.е. среднее число заявок, обслуживаемых в единицу времени:

$\mu = \dfrac{1}{\bar{t}_{обс}}$ (чел./мин, р./ дн., кг/ч, докум./дн.), где $\bar{t}_{обс}$ - среднее время обслуживания.
Важной характеристикой СМО, объединяющей λ и μ, является интенсивность нагрузки

$$\rho = \frac{\lambda}{\mu}.$$

СМО С ОТКАЗАМИ

Заявка, поступившая в систему с отказами, при этом все каналы заняты, получает отказ и покидает систему необслуженной. Показателем качества обслуживания выступает вероятность получения отказа. Предполагается, что все каналы доступны в равной степени всем заявкам, входящий поток является простейшим, длительность (время) обслуживания одной заявки $\left(t_{обс}\right)$ распределена по показательному закону.
По ниже написанным формулам можно рассчитать СМО с отказами:

1. Вероятность простоя каналов обслуживания, когда нет заявок (k = 0):

$$P_0 = \frac{1}{\sum_{k=0}^{n} \frac{\rho^k}{k!}}$$

2. Вероятность отказа в обслуживании, когда поступившая на обслуживание заявка найдет все каналы занятыми (k = n):

$$P_{отк} = P_n = \frac{P_0 \rho^n}{n!};$$

3. Вероятность обслуживания:

$$P_{обс} = 1 - P_{отк};$$

4. Среднее число занятых обслуживанием каналов:

$$\bar{n}_з = \rho P_{обс};$$

5. Доля каналов, занятых обслуживанием:

$$k_з = \frac{\overline{n}_з}{n};$$

6. Абсолютная пропускная способность СМО:

$$A = \lambda P_{обс};$$

СМО С НЕОГРАНИЧЕННЫМ ОЖИДАНИЕМ

Заявка, поступившая в систему с неограниченным ожиданием и нашедшая все каналы занятыми, становится в очередь, ожидая освобождения одного из каналов.

Основной характеристикой качества обслуживания является время ожидания (время пре-бывания заявки в очереди).

Для таких систем характерно отсутствие отказа в обслуживании, т.е. $P_{отк} = 0$ и $P_{обс} = 1$.

Для систем с ожиданием существует дисциплина очереди:

1. Обслуживание в порядке очереди по принципу “первым пришел – первым обслужен”;
2. Случайное неорганизованное обслуживание по принципу “последний пришел – первым обслужен”;
3. Обслуживание с приоритетами по принципу “самые главные вне очереди”.

По ниже написанным формулам можно вычислить СМО с неограниченным ожиданием:

1. Вероятность простоя каналов, когда нет заявок (k = 0):

$$P_0 = \frac{1}{\sum\limits_{k=0}^{n}\frac{\rho^k}{k!}} + \frac{\rho^{n+1}}{n!(n-\rho)};$$ предполагается, что $\frac{\rho}{n} < 1$.

2. Вероятность занятости обслуживанием k заявок:

$$P_k = \frac{\rho^k P_0}{k!},\ 1 \le k \le n;$$

3. Вероятность занятости обслуживанием всех каналов:

$$P_n = \frac{\rho^n P_0}{n!};$$

4. Вероятность того, что заявка окажется в очереди:

$$P_{оч} = \frac{\rho^{n+1}}{n!(n-\rho)} P_0;$$

5. Среднее число заявок в очереди:

$$\overline{L}_{оч} = \frac{\rho^{n+1}}{(n-1)!(n-\rho)^2} P_0;$$

6. Среднее время ожидания заявки в очереди:

$$\bar{t}_{оч} = \frac{\bar{L}_{оч}}{\lambda};$$

7. Среднее время пребывания заявки в СМО:

$$\bar{t}_{смо} = \bar{t}_{оч} + \bar{t}_{обс};$$

8. Среднее число занятых обслуживанием каналов:

$$\bar{n}_з = \rho;$$

9. Среднее число свободных каналов:

$$\bar{n}_{св} = n - \bar{n}_з;$$

10. Коэффициент занятости каналов обслуживания:

$$k_з = \frac{\bar{n}_з}{n};$$

11. Среднее число заявок в СМО:

$$\bar{Z} = \bar{L}_{оч} + \bar{n}_з;$$

СМО С ОЖИДАНИЕМ И ОГРАНИЧЕННОЙ ДЛИНОЙ ОЧЕРЕДИ

Заявка, поступившая в систему с ожиданием с ограниченной длиной очереди и нашедшая все каналы и ограниченную очередь занятыми, покидает систему необслуженной.
Основной характеристикой качества системы является отказ заявке в обслуживании.
Ограничения на длину очереди могут быть из-за:

1. Ограничения сверху времени пребывания заявки в очереди;
2. Ограничения сверху длины очереди;
3. Ограничения общего времени пребывания заявки в системе.

По ниже написанным формулам можно рассчитать СМО с ожиданием и с ограниченной длиной очереди:

1. Вероятность простоя каналов обслуживания, когда нет заявок (k = 0):

$$P_0 = \frac{1}{\left\{\sum_{k=0}^{n}\frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)}\left[1-\left(\frac{\rho}{n}\right)^m\right]\right\}};$$

2. Вероятность отказа в обслуживании:

$$P_{отк} = \frac{\rho^{n+m}}{n!\,n^m}P_0;$$

3. Вероятность обслуживания:

$$P_{обс} = 1 - P_{отк};$$

4. Абсолютная пропускная способность:

$$A = P_{обс} \cdot \lambda;$$

5. Среднее число занятых каналов:

$$\overline{n}_з = \frac{A}{\mu};$$

6. Среднее число заявок в очереди:

$$\overline{L}_{оч} = \frac{\rho^{n+1}}{n \cdot n!} \frac{1 - \left(\frac{\rho}{n}\right)^m \left(m + 1 - \frac{m\rho}{n}\right)}{\left(1 - \frac{\rho}{n}\right)^2} P_0;$$

7. Среднее время ожидания обслуживания:

$$\overline{t}_{оч} = \frac{\overline{L}_{оч}}{\lambda};$$

8. Среднее число заявок в системе:

$$\overline{Z} = \overline{L}_{оч} + \overline{n}_з;$$

9. Среднее время пребывания в системе:

$$\overline{t}_{смо} = \frac{\overline{Z}}{\lambda};$$

ОДНОКАНАЛЬНАЯ СМО С ОЖИДАНИЕМ

Пусть простейший поток заявок на обслуживание – простейший поток с интенсивностью λ.

Интенсивность потока обслуживания равна μ. Заявка, поступившая в момент, когда канал занят, становится в очередь и ожидает обслуживания. Предположим, что СМО не может вместить более N заявок, т.е. заявки, не попавшие в ожидание, покидают СМО. Состояния СМО имеют следующий вид:

S_0 – канал свободен;

S_1 – канал занят, очереди нет;

S_2 – канал занят, одна заявка в очереди;

…………………………………………….

S_n – канал занят, n-1 заявка в очереди;

…………………………………………….

S_N – канал занят, N-1 заявка в очереди.

Процесс в данной системе будет описан системой алгебраических уравнений:

$$\begin{cases} -\rho P_0 + P_1 = 0, n = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 1-\rho P_n + P_{n-1} + \rho P_{n-1} = 0, 0 < n < N \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ -P_n + \rho P_{n-1} = 0, n = N \end{cases}$$

, где $\rho = \frac{\lambda}{\mu}$; n – номер состояния.

Система уравнений имеет следующее решение;

$$\begin{cases} P_0 = \frac{1-\rho}{1-\rho^{(N+1)}} \\ P_n = P_0\rho^n,\ \rho \neq 1,\ n = 1,2,\ldots.,N \\ P_n = \frac{1}{N+1}\ при\ \rho = 1 \end{cases}$$

Выполнение условия $\rho < 1$ необязательно, поскольку число допускаемых в СМО заявок контролируется путем введения ограничения на длину очереди. Определим характеристики одноканальной СМО с ожиданием и ограниченной длиной очереди, равной (N-1):

1. Вероятность отказа обслуживания заявки:

   $P_{otk} = P_N$;

2. Относительная пропускаемая способность СМО:

   $q = 1 - P_{otk}$;

3. Абсолютная пропускная способность СМО:

   $A = q\lambda$;

4. Среднее число находящихся в СМО заявок:

   $L_s = \sum_{n=0}^{N} nP_n$;

5. Среднее время пребывания заявки в СМО:

   $W_s = \frac{L_s}{\lambda(1-P_N)}$;

6. Средняя продолжительность пребывания заявки в очереди:

   $W_q = W_s - \frac{1}{\mu}$;

7. Среднее число заявок в очереди:

   $L_q = \lambda(1-P_N)W_q$;

<u>МНОГОКАНАЛЬНАЯ СМО С ОЖИДАНИЕМ</u>

Пусть входной и выходной поток с интенсивностями λ и μ, соответственно. СМО имеет С каналов обслуживания. Средняя продолжительность обслуживания одного клиента равна $\frac{1}{\mu}$.

Многоканальное СМО с ожиданием и неограниченной очередью может быть описано с помощью системы алгебраических уравнений:

$$\begin{cases} \lambda P_{n-1} - (\lambda + n\mu)P_n + (n+1)\mu P_{n+1} = 0, \text{ при } C > n \ge 1, \\ \lambda P_{n-1} - (\lambda + C\mu)P_n + C\mu P_{n+1} = 0, \text{ при } n \ge C \end{cases}$$

Пусть $\rho = \frac{\lambda}{\mu}$.

При условии $\frac{\rho}{C} < 1$решение системы уравнений имеет вид

$$P_0 = \left(\left(\sum_{n=0}^{C-1} \frac{\rho^n}{n!} \right) + \frac{\rho^C}{C!\left(1 - \frac{\rho}{C}\right)} \right)^{(-1)} .$$

$$P_n = \frac{\rho^n P_0}{n!}; \text{ при } 0 \le n < C,$$

$$P_n = \frac{\rho^n P_0}{C!\,C^{(n-C)}}; \text{ при } C \le n.$$

Здесь P_n есть вероятность того, что в СМО n клиентов находится на обслуживании. Среднее число клиентов в очереди на обслуживание определяется следующей формулой:

$$L_q = \frac{C_\rho P_C}{(C-\rho)^2} .$$

Среднее число находящихся в СМО клиентов (на обслуживании и в очереди):

$$L_s = L_q + \rho .$$

Средняя продолжительность пребывания клиента (заявки) в очереди:

$$W_q = \frac{L_q}{\lambda} .$$

Средняя продолжительность пребывания клиента в СМО:

$$W_s = W_q + \frac{1}{\mu}.$$

ЗАДАЧА ОПТИМАЛЬНОГО ТЕХОБСЛУЖИВАНИЯ АВТОМАШИН (дополнение)

Станция S имеет два вида техобслуживания машин: 1.«VIP» и 2.«Эконом». Каждое из этих видов техобслуживания может в случайный момент времени выйти из строя, после чего мгновенно начинается ремонт, продолжающийся заранее неизвестное случайное время.

Для того чтобы решить эту задачу, перечислим возможные состояния системы:

S_o— оба вида исправны;

S_1— первый вид ремонтируется, второй исправен;

S_2— первый вид исправен, второй ремонтируется;

S_3 — оба вида ремонтируются.

Соответствующий граф приведен на рис. 2.

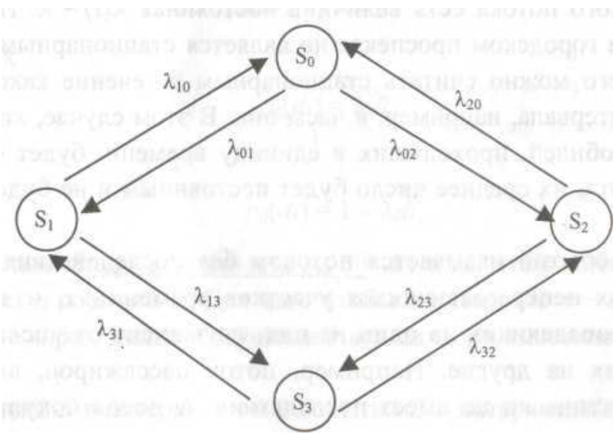


Рис.2 Граф состояний случайного процесса

Дуга, направленная из S_o в S_1 означает переход в момент отказа первого вида; из S_1 в S_o переход в момент окончания ремонта этого вида. На графе отсутствуют стрелки из S_o в S_3 и из S_1 в S_2. Это объясняется тем, что выходы видов обслуживания из строя предполагаются независимыми друг от друга и, например, вероятностью одновременного выхода из строя двух видов (переход от S_o в S_3) или одновременного окончания ремонтов двух видов обслуживания (переход из S_3 в S_o) можно пренебречь.

Предположим, что все переходы системы из состояния S_i в S_j происходят под воздействием простейших потоков событий с интенсивностями состояний λ_{ij} (i, j = 0, 1, 2, 3).

Так, переход системы из состояния S_0 в состояние S_1 будет происходить под воздействием потока отказов первого вида, а обратный переход из состояния S_1 в S_0 – под воздействием потока и событий, связанных с окончанием ремонтов первого вида и т.п.

Граф состояний системы с проставленными у стрелок интенсивностями будем называть размеченным. Рассматриваемая система S имеет четыре возможные состояния : S_0, S_1, S_2, S_3. Назовем вероятностью i-го состояния вероятность $p_i(t)$ того, что в момент t система будет находиться в состоянии S_i. Очевидно, что для любого момента t сумма вероятностей всех состояний равна единице.

Рассмотрим систему в момент t и, задав малый интервал времени Δt, найдем вероятность $p_0(t+\Delta t)$ того, что система в момент $t+\Delta t$ будет находиться в состоянии S_0. Это достигается разными способами.

1. Система в момент t с вероятностью $p_0(t)$, находясь в состоянии S_0 и за время Δt, не вышла из него. Вывести систему из этого состояния можно суммарным простейшим потоком с интенсивностью ($\lambda_{01} + \lambda_{02}$), т. е. в соответствии с вероятностью приближенно равной $(\lambda_{01} + \lambda_{02})\ \Delta t$. Вероятность же того, что система не выйдет из состояния S_0, равна по теореме умножения вероятностей $p_0(t)[1 – (\lambda_{01} + \lambda_{02})\ \Delta t]$.

2. Система в момент t с вероятностями $p_1(t)$ (или $p_2(t)$) находилась в состоянии S_1 и S_2 и за время Δt перешла в состояние S_0. Потоком, интенсивностью λ_{10} (или λ_{20} как следует из рис. 2.) система перейдет в состояние S_0 с вероятностью приближенно равной $\lambda_{10}\Delta t$ (или $\lambda_{20}\Delta t$). Вероятность того, что система будет находиться в состоянии S_0 по этому способу, равна $p_1(t)\ \lambda_{10}\Delta t$ (или $p_2(t)\lambda_{20}\Delta t$). Применяя теорему сложения вероятностей, получим

$$p_0(t + \Delta t) = p_1(t)\ \lambda_{10}\Delta t + p_2(t)\ \lambda_{20}\Delta t + p_0t[1 – (\lambda_{01} + \lambda_{02})\Delta t],$$

откуда

$$(p_0(t + \Delta t) - p_0(t)) / \Delta t = p_1(t)\,\lambda_{10} + p_2(t)\lambda_{20} - p_0(t)(\lambda_{01} + \lambda_{02}).$$

Переходя к пределу в левой части последнего равенства получим:

$$p'_0(t) = \lambda_{10}p_1(t) + \lambda_{20}p_2(t) - (\lambda_{01} + \lambda_{02})p_0(t).$$

Получили дифференциальное уравнение первого порядка, т. е. уравнение, содержащее как саму неизвестную функцию, так и ее производную первого порядка.

Рассуждая аналогично для других состояний системы S, получим систему дифференциальных уравнений, которая носит название системы дифференциальных уравнений Колмогорова для вероятностей состояний

$$\begin{cases} p'_0(t) = \lambda_{10}p_1(t) + \lambda_{20}p_2(t) - (\lambda_{01} + \lambda_{02})p_0(t) \\ p'_1(t) = \lambda_{01}p_0(t) + \lambda_{31}p_3(t) - (\lambda_{10} + \lambda_{13})p_1(t) \\ p'_2(t) = \lambda_{02}p_0(t) + \lambda_{32}p_3(t) - (\lambda_{20} + \lambda_{23})p_2(t) \\ p'_3(t) = \lambda_{13}p_1(t) + \lambda_{23}p_2(t) - (\lambda_{31} + \lambda_{32})p_3(t). \end{cases} \qquad (1.1)$$

Сформулируем правило составления уравнений Колмогорова. В левой части каждого из них стоит производная вероятностей i-го состояния. В правой части – сумма произведений вероятностей всех состояний (из которых идут дуги в данное состояние) на интенсивности соответствующих потоков событий минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность данного состояния.

Особенность решения дифференциальных уравнений состоит в том, что необходимо задать начальные условия уравнений, т.е. в данном случае вероятности состояний системы в начальный момент времени t = 0. Так, например, систему уравнений (1.1) естественно решать при условии, что в начальный момент времени t = 0 оба вида техобслуживания исправны и система находится в состоянии S_0, т.е. при начальном условии $p_0(0) = 1$, $p_1(0) = p_2(0) = p_3(0) = 0$.

Уравнения Колмогорова дают возможность найти все вероятности состояний как функции времени. Особый интерес представляют вероятности системы $p_i(t)$ в предельном стационарном решении, т.е. при $t \to \infty$, которые называются предельными (или финальными) вероятностями состояний.

Предельная вероятность состояния S_i имеет вполне определенный смысл: она показывает среднее относительное время пребывания системы в этом состоянии. Например, если предельная вероятность состояния S_0, т.е. $p_0 = 0,5$, то это означает, что в среднем половину времени система находится в состоянии S_0.

Так как предельные вероятности постоянны, то, заменяя в уравнениях Колмогорова их производные нулевыми значениями, получим систему линейных алгебраических уравнений, описывающих стационарный режим. Для системы S с графом состояний, изображенном на рис. 2, такая система уравнений имеет вид

$$\begin{cases} (\lambda_{01} + \lambda_{02})\, p_0 = \lambda_{10}\, p_1 + \lambda_{20}\, p_2 \\ (\lambda_{10} + \lambda_{13})\, p_1 = \lambda_{01}\, p_0 + \lambda_{31}\, p_3 \\ (\lambda_{20} + \lambda_{23})\, p_2 = \lambda_{02}\, p_0 + \lambda_{32}\, p_3 \\ (\lambda_{31} + \lambda_{32})\, p_3 = \lambda_{13}\, p_1 + \lambda_{23}\, p_2. \end{cases} \qquad (1.2)$$

Система (1.2) может быть получена непосредственно по размеченному графу состояний, если руководствоваться правилом, согласно которому в левой части уравнений стоит предельная вероятность данного состояния p_i, умноженная на суммарную интенсивность всех потоков, входящих в i-е состояние, на вероятности тех состояний, из которых эти потоки исходят.

Найдем предельные вероятности для системы, граф состояний которого изображен на рис.2 при

$$\lambda_{01} = 1,\ \lambda_{02} = 2,\ \lambda_{10} = 2,\ \lambda_{13} = 2,\ \lambda_{20} = 3,\ \lambda_{23} = 1,\ \lambda_{31} = 3,\ \lambda_{32} = 2.$$

Система алгебраических уравнений имеет вид

$$\begin{cases} 3p_0 = 2p_1 + 3p_2 \\ \\ \\ \end{cases}$$

$4p_1 = p_0 + 3p_3$

$4p_2 = 2p_0 + 2p_3$

$p_0 + p_1 + p_2 + p_3 = 1.$

Решаем систему алгебраических уравнений в среде MX Excel.

Для решения оптимизационных задач в табличном процессоре MX Excel используется пункт «Поиск решения» в меню «Данные» на панели инструментов.

Задаем значения уравнений и используем пункт «Поиск решения»:

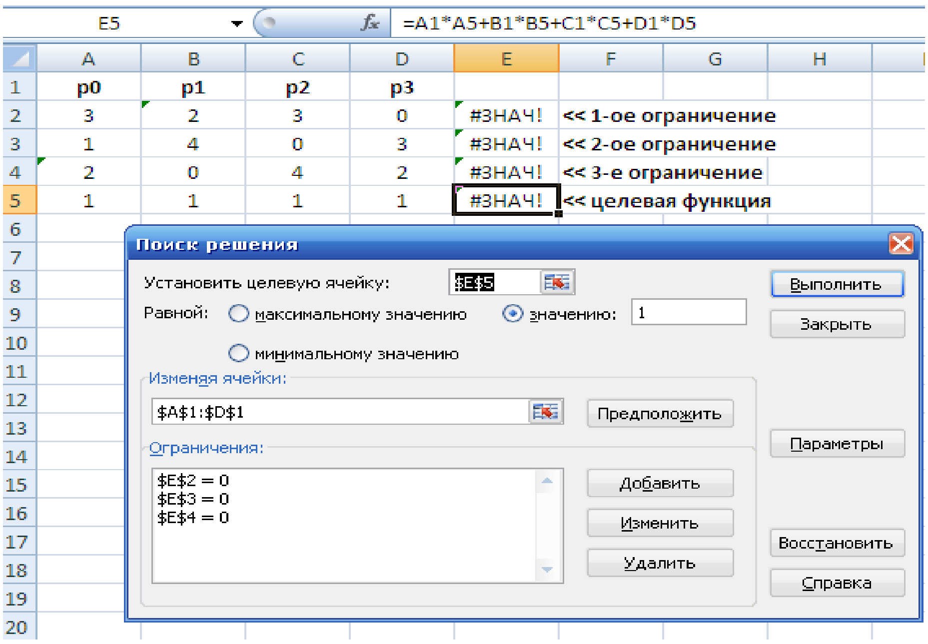


Рис. 3 «Поиск решения»

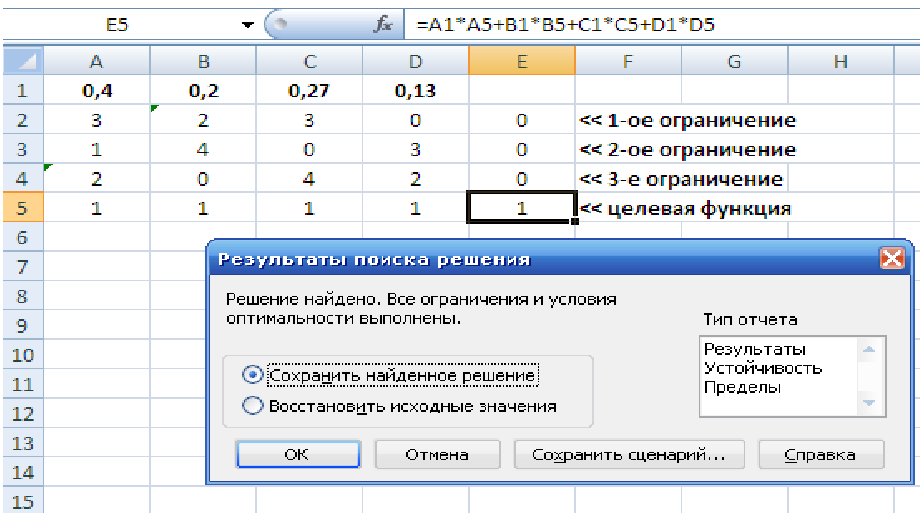
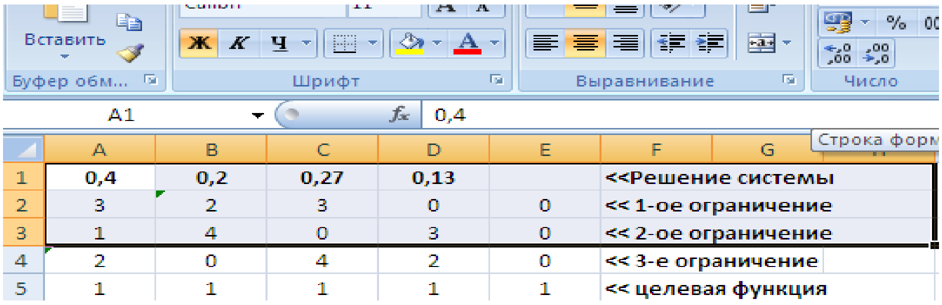


Рис. 3.1. Результаты поиска решения



	A	B	C	D	E	F
1	0,4	0,2	0,27	0,13		<<Решение системы
2	3	2	3	0	0	<< 1-ое ограничение
3	1	4	0	3	0	<< 2-ое ограничение
4	2	0	4	2	0	<< 3-е ограничение
5	1	1	1	1	1	<< целевая функция

Рис. 3.2. Решение системы уравнений

Решив линейную систему уравнений, получаем:

$$p_0 = 0,4;\ p_1 = 0,2;\ p_2 = 0,27;\ p_3 = 0,13;$$

т.е. в предельном стационарном режиме система S в среднем:

40% времени будет находиться в состоянии S_0 (оба узла исправны);

13% времени в состоянии S_1 (первый узел ремонтируется, второй работает);

27% — в состоянии S_2 (второй узел ремонтируется, первый работает);

13% в состоянии S_3 (оба узла ремонтируются).

Определим **чистый доход** от эксплуатации в стационарном режиме рассмотренной системы *S* в условиях, что в единицу времени исправная работа вида «VIP » и вида «Эконом» приносит доход соответственно 10 и 6 денежных единиц, а их ремонт требует соответственно затрат 4 и 2 денежных единиц. Оценим экономическую эффективность имеющейся возможности уменьшения вдвое среднего времени ремонта каждого из видов техобслуживания, если при этом придется вдвое увеличить затраты на ремонт каждого вида (в единицу времени).

Для решения этой задачи, с учетом полученных значений p_0, p_1, p_2, p_3, определим долю времени исправной работы «VIP»:

$$p_0 + p_3 = 0,4 + 0,27 = 0,67$$

и долю времени исправной работы «Эконом»:

$$p_0 + p_1 = 0,4 + 0,2 = 0,6.$$

В то же время первый вид находится в ремонте в среднем долю времени равную:

$$p_1 + p_3 = 0{,}2 + 0{,}13 = 0{,}33,$$

а второй вид:

$$p_2 + p_3 = 0{,}27 + 0{,}13 = 0{,}40.$$

Поэтому средний чистый доход в единицу времени от эксплуатации системы равен

$$Д_1 = 0{,}67 \times 10 + 0{,}6 \times 6 - 0{,}33 \times 4 - 0{,}4 \times 2 = 8{,}18 \text{ ден. ед.}$$

Уменьшение вдвое среднего времени ремонта каждого вида техобслуживания будет означать увеличение вдвое интенсивностей потока «окончания ремонтов» каждого вида, т.е. теперь

$$\lambda_{10} = 4,\ \lambda_{20} = 6,\ \lambda_{31} = 6,\ \lambda_{32} = 4$$

и система уравнений, описывающая стационарный режим системы *S*, будет иметь вид:

$$\begin{cases} 3p_0 = 4p_1 + 6p_2 \\ 4p_1 = p_0 + 6p_3 \\ 7p_2 = 2p_0 + 4p_3 \\ p_0 + p_1 + p_2 + p_3 = 1. \end{cases}$$

Решаем систему алгебраических уравнений в среде MX Excel.

Задаем значения уравнений и используем пункт «Поиск решения»:

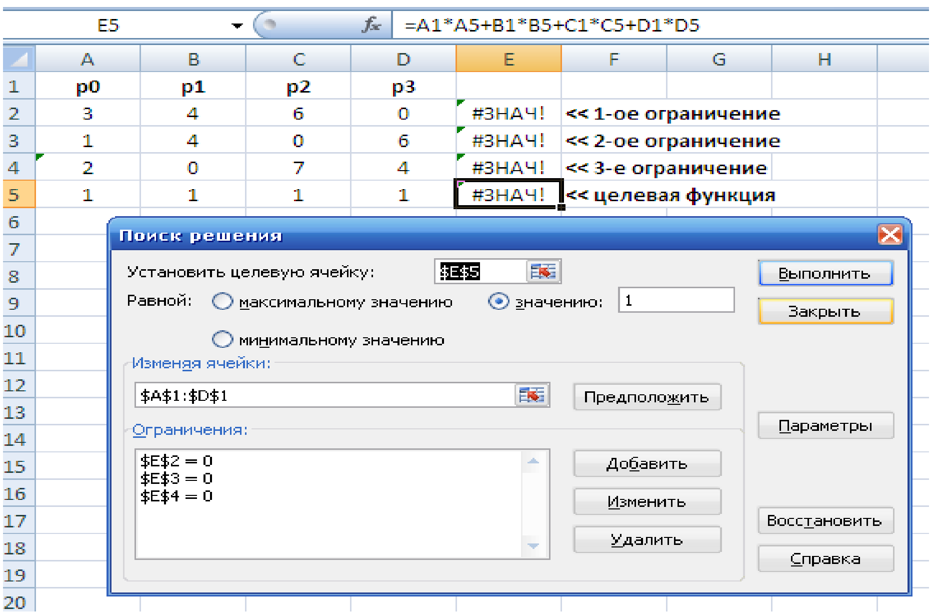
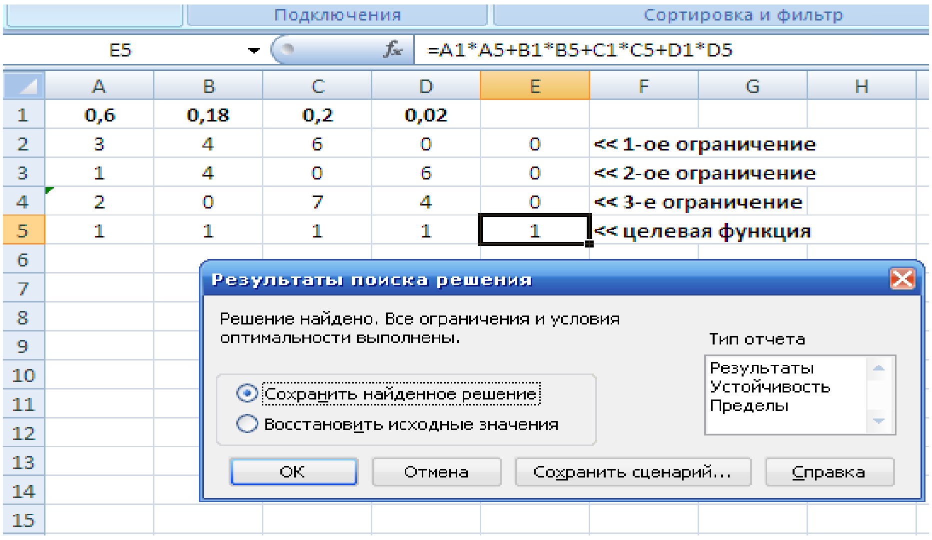
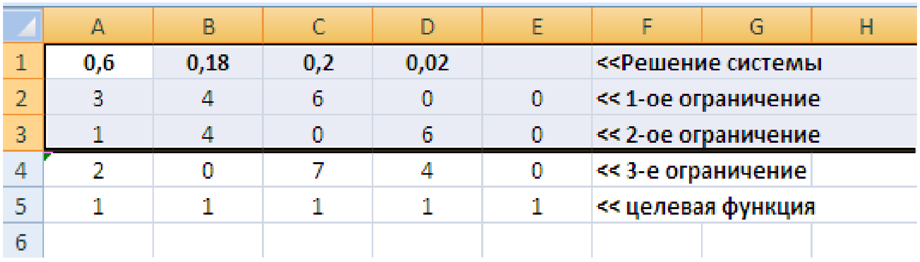


Рис. 4. «Поиск решения»



	A	B	C	D	E	F
1	0,6	0,18	0,2	0,02		
2	3	4	6	0	0	<< 1-ое ограничение
3	1	4	0	6	0	<< 2-ое ограничение
4	2	0	7	4	0	<< 3-е ограничение
5	1	1	1	1	1	<< целевая функция

Рис. 4.1. Результаты поиска решения



	A	B	C	D	E	F
1	0,6	0,18	0,2	0,02		<<Решение системы
2	3	4	6	0	0	<< 1-ое ограничение
3	1	4	0	6	0	<< 2-ое ограничение
4	2	0	7	4	0	<< 3-е ограничение
5	1	1	1	1	1	<< целевая функция

Рис. 4.2. Решение системы уравнений

Решив систему, получаем:

$$p_0 = 0{,}6;\ \ p_1 = 0{,}18;\ p_2 = 0{,}2;\ p_3 = 0{,}02\ .$$

Учитывая, что $p_0 + p_2 = 0{,}6 + 0{,}2 = 0{,}8;$

$$p_0 + p_1 = 0{,}6 + 0{,}18 = 0{,}78;$$

$$p_1 + p_3 = 0{,}18 + 0{,}02 = 0{,}2;$$

$$p_2 + p_3 = 0{,}2 + 0{,}02 = 0{,}22,$$

а затраты на ремонт «VIP» и «Эконом» составляют соответственно 8 и 4 ден. ед., вычислим чистый средний доход в единицу времени:

$$Д_2 = 0{,}8\times10 + 0{,}78\times6 - 0{,}2\times8 - 0{,}22\times4 = 10{,}2 \text{ ден. ед.}$$

Вывод : Так как $Д_2$ больше $Д_1$ (примерно на 20%), то экономическая целесообразность ускорения ремонта видов техобслуживания очевидна.

Тема 5. Теория управления запасами

Основные понятия

Выбирается промежуток времени 1 год. Рассматривается модель одиночного склада. Считается, что на складе хранится запас однотипных изделий (однономенклатурный запас). Спрос на эти изделия может быть постоянным или случайным. Пополняться склад может либо периодически (циклическая модель), либо при снижении запасов до некоторого уровня (уровневая модель).

Объем заказа — это количество заказываемых изделий. Уровень повторного заказа — количество изделий на складе, при котором подается заказ на новые изделия. Время поставки может быть либо мгновенным, либо фиксированным, либо случайным. Штраф за дефицит — это убытки, связанные с отсутствием запаса.

За хранение каждой единицы запаса берется определенная плата C_h. D — годовой спрос на изделия. Стоимость подачи заказа C_o — это накладные расходы, связанные с реализацией заказа (затраты на подготовительно-заготовочные операции, не зависят от объема заказа). Вся теория будет строиться с целью минимизации суммарных издержек.

5.1 Основная модель управления запасами

Предпосылки основной модели: 1) спрос равномерный и постоянный; 2) время поставки постоянно; 3) отсутствие запасов недопустимо; 4) каждый раз заказывается постоянное количество — оптимальный размер заказа.

Издержки *TC* = подача заказов + хранение min, (1)

где q — оптимальный размер заказа;

$q/2$ — средний объем хранимого запаса.

Решением этой оптимизационной задачи будет значение:

$$q = \sqrt{\frac{2C_0D}{C_h}}\ .$$

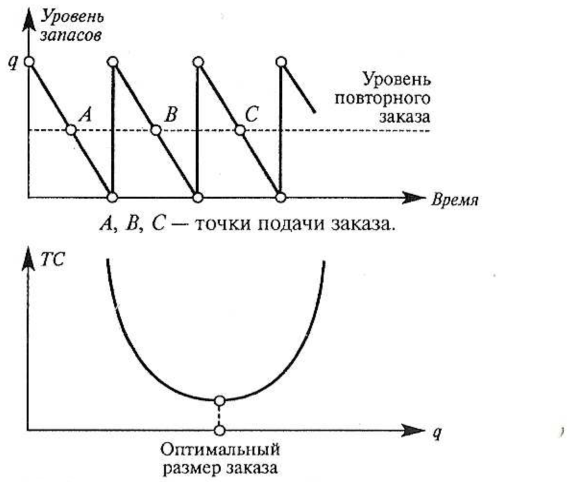


Рисунок 5.1- Решение оптимизационной задачи

Пример. Годовой спрос D = 1500 единиц, стоимость подачи заказа C_o = 150 рублей/заказ, издержки хранения одной единицы C_h = 45 рублей/год, время доставки 6 дней, 1 год = 300 рабочих дней. Найдем оптимальный размер заказа, издержки, уровень повторного заказа.

Оптимальный размер заказа:

$$q = \sqrt{\frac{2C_0D}{}} = \sqrt{\frac{2\times150\times1500}{}} =$$ 100 единиц.

Издержки: TC(q) =

За 300 рабочих дней реализуется 1500 единиц, за 6 дней доставки — x единиц. 300/6 = 1500/x. Отсюда x= 1500×6/300 = 30 единиц. Каждый раз, когда на складе остается 30 единиц, подается заказ на 100 единиц.

Годовой спрос D = 1500 единиц, каждый раз заказывается q= 100 единиц. Поэтому всего за год будет подано D/q = 1500/100 = 15 заказов. Говорят, что за год пройдет 15 циклон. Расстояние между циклами 1/(D/q) = q/D = 100/1500= 1/15 лет = 300×(1/15) = 20 рабочих дней.

5.2 Модель экономического размера партии

Технологический процесс может быть организован на основе производства партии продукции: чередование процессов производства и реализации произведенного.

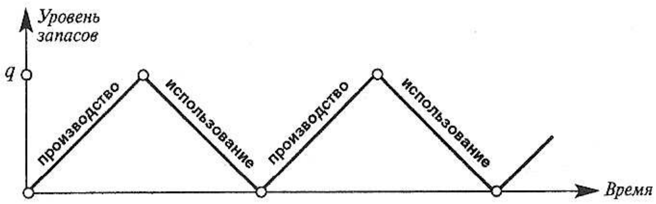


Рисунок 5.2 – Модель экономического размера заказа

Каким должен быть размер q партии продукции? Обозначим через C_s, стоимость организации производственного цикла (фиксированные издержки производства).

Издержки ТС = стоимость организации технологического процесса + хранение min,

где q — экономичный размер партии.

Решение этой задачи:

$$q = \sqrt{\frac{2C_sD}{C_h}}$$

Пример. Годовой спрос D = 14800 единиц, стоимость организации производственного цикла C_s= 100 рублей, издержки хранения одной единицы C_h = 8 рублей/год.

Экономичный размер партии равен:

$$q = \sqrt{\frac{2C_sD}{C_h}} = \sqrt{\frac{2 \times 100 \times 14800}{45}} \approx 608 \text{ единиц.}$$

То есть надо произвести 608 единиц, остановить производство, реализовать всю произведенную продукцию и вновь запустить производство. И так далее.

Издержки ТС равны:

Число циклов за год D/q = 14800/608 ≈ 24,3. Расстояние между циклами q/D≈ ≈ 0,04лет≈ 15 дней.

5.3. Скидка на количество

Очень часто, если заказываемое количество товара больше определенного числа, предоставляется скидка. В этом случае снижаются расходы на закупку, но увеличиваются затраты на хранение.

Общие издержки = закупка + издержки $TC(q)$

где C — закупочная цена.

Необходимо выяснить, стоит ли воспользоваться скидкой.

Пример. Годовой спрос D = 1000 единиц, стоимость подачи заказа C_o = 40 рублей/заказ, закупочная цена С = 50 рублей/единицу, годовая стоимость хранения одной единицы составляет 25% ее цены, Можно получить скидку 3% у поставщиков, если размер заказа будет не меньше 200 единиц (уровень, нарушающий цену). Стоит ли воспользоваться скидкой?

Так как годовая стоимость хранения одной единицы составляет 25% ее цены, то C_h = 0,25×С = 0,25×50 = 12,5 руб./единицу.

Найдем общие издержки в случае основной модели.

$$q = \sqrt{\frac{2C_0D}{C_h}} = \sqrt{\frac{2 \times 40 \times 1000}{12{,}5}} = 80 \text{ единиц.}$$

Общие издержки равны: $TC = CD +$

$$\frac{C_0 D}{q} + \frac{C_h q}{2} = 50 \times 1000 + \frac{40 \times 1000}{80} + \frac{12{,}5 \times 80}{2} = 51000 \text{ руб.} \frac{}{\text{год}}.$$

Если воспользоваться скидкой, то новая закупочная цена равна:

C = 0,97×50 = 48,5 рублей/единицу.

Поэтому C_h = 0,25×C = 0,25×48,5 = 12,125 рублей/единицу.

В этом случае оптимальный размер заказа равен:

$$q = \sqrt{\frac{2C_0 D}{C_h}} = \sqrt{\frac{2 \times 40 \times 1000}{12{,}125}} \approx 81 \text{ единица}.$$

Но скидка предоставляется, если объем заказа $q \geq 200$. Поэтому положим q=200.

Тогда общие издержки равны:

$$TC = 49906{,}5 \text{ руб.} \frac{}{\text{год}}.$$

Мы видим, что общие издержки уменьшились. Поэтому следует воспользоваться скидкой, заказывая каждый раз 200 единиц.

Число циклов за год равно D/q=1000/200=5, а интервал между циклами q/D=200/1000=1/5 лет=73дня.

5.4 Модель производства партии продукции

Ранее была рассмотрена модель экономичного размера партии (сначала товар производят, потом используют и т. д.). Разрешим теперь использование товара по мере его производства.

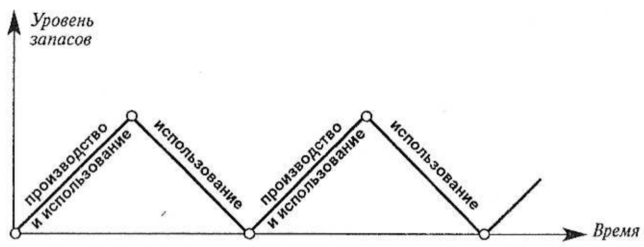


Рисунок 5.3 – Модель производства партии продукции

Пусть P — темп производства, D — темп использования. Произведя q единиц продукции, производство прекращаем. Так как мы начинаем использовать произведенную продукцию сразу же, не дожидаясь остановки производства, то в момент этой остановки на складе будет не q единиц (как в модели экономичного размера партии), а меньше.

Издержки TC = стоимость организации технологического процесса + +хранение = $\frac{C_S D}{q} + \frac{C_h (P-D) q}{2P} \to$ min,

где q – экономичный предел партии.

$$\frac{C_h P_q}{2P} - \frac{C_h D_q}{2P} = \frac{C_h q}{2} - \frac{C_h q}{2} \times \frac{D}{P} = \frac{C_h q}{2\left(1 - \frac{D}{P}\right)}$$

Решение этой задачи:

$$q = \sqrt{\frac{2C_S D}{C_h}} \times \sqrt{\frac{P}{P-D}}.$$

Пример. Компания выпускает электрические ножи. Она в среднем может производить 150 ножей/день. Спрос — 40 ножей/день. Годовые издержки хранения C_h = 8 руб./нож. Стоимость организации производственного цикла C_s = 100 рублей. Найдем экономичный размер партии, издержки, число циклов за год, расстояние между циклами.

P= 150 ножей/день = 54750 ножей/год, D = 40 ножей/день = 14600 ножей/год (напомним, что вся теория строится для временного интервала 1 год).

Экономичный размер партии равен:

$$q=\sqrt{\frac{2C_sD}{C_h}}\times\sqrt{\frac{P}{P-D}}=\sqrt{\frac{2\times 100\times 14600}{8}}\times\sqrt{\frac{54750}{54750-14600}}\approx 705$$ ед

иниц.

Издержки равны: TC =

$$\frac{C_sD}{q}+\frac{C_h(P-D)q}{2P}=\frac{100\times 14600}{705}+\frac{8\times\left(54750-14600\right)\times 705}{2\times 54750}\approx 4138{,}92$$

руб./год.

Таким образом, производим 705 ножей, останавливаем производство. Ножи реализуются сразу не дожидаясь остановки производства. Как только ножи закончатся, тут же запускаем производственный процесс. Число циклов за год равно D/q = 14600/705 ≈ 20,7, а интервал между циклами

q/D = 705/14600 ≈ 0,048 лет ≈ 18 дней.

5.5 Модель планирования дефицита

В некоторых случаях издержки хранения являются очень высокими. Поэтому имеет смысл допустить регулярные интервалы времени, когда товар на складе отсутствует.

Издержки TC = подача заказов + хранение + штраф за дефицит. Возможны два подхода:

1) полученная новая продукция не идет на выполнение заявок на товар во время его отсутствия;

2) часть полученной новой продукции идет на погашение всех заявок, оставленных во время отсутствия запасов.

Случай невыполнения заявок

Рассмотрим эти случаи подробнее.

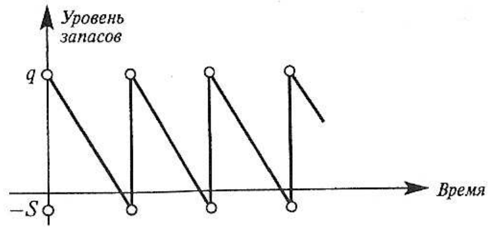


Рисунок 5.4 – Случай невыполнения заявок

S — максимальный размер дефицита (максимально возможное число единиц товара, которое могло бы быть реализовано за время его отсутствия в каждом цикле). На графике периоды дефицита условно изображаются ниже оси времени. C_b — годовая стоимость отсутствия единицы продукции в запасе (потеря доверия клиентов, непроданная продукция и т. д.). При использовании моделей управления запасами расходы из-за дефицита вычислить очень трудно.

Издержки *TC* = подача заказов + хранение + штраф за дефицит $=\frac{C_0 D}{q+S}+\frac{C_h q^2}{2(q+S)}++\frac{C_b S^2}{2(q+S)}\rightarrow$ min,

где q – оптимальный размер заказа;

S — максимальный размер дефицита.

Решениями этой задачи будут величины:

$$q=\sqrt{\frac{2C_0 D}{C_h}}\times\sqrt{\frac{C_b}{C_h+C_b}}.\ S=\sqrt{\frac{2C_0 D}{C_b}}\times\sqrt{\frac{C_h}{C_h+C_b}}.$$

Пример. Годовой спрос *D* = 500 единиц, стоимость подачи заказов C_o = 40 рублей/заказ, издержки хранения одной единицы C_h = 5 рублей/год, годовая стоимость отсутствия запасов C_b = 100 рублей/единицу.

Сравним 2 модели: основную и с дефицитом (заявки не выполняются).

Основная модель:

$$q=\sqrt{\frac{2C_0 D}{C_h}}=\sqrt{\frac{2\times 40\times 500}{5}}\approx 89 \text{ единиц.}$$

$$TC = \frac{C_0 D}{q} + \frac{C_h q}{2} = \frac{40 \times 500}{89} + \frac{5 \times 89}{2} \approx 447$$ руб./год.

Модель с дефицитом:

$$q = \sqrt{\frac{2C_0 D}{C_h}} \times \sqrt{\frac{C_b}{C_h + C_b}} = 89\sqrt{\frac{100}{5 + 100}} \approx 87 \text{ единиц.}$$

$$S = \sqrt{\frac{2C_0 D}{C_b}} \times \sqrt{\frac{C_h}{C_h + C_b}} = \sqrt{\frac{2 \times 40 \times 500}{100}} \times \sqrt{\frac{5}{5 + 100}} \approx 4 \text{ единицы.}$$

$$TC = \frac{C_0 D}{q + S} + \frac{C_h q^2}{2(q + S)} + \frac{C_b S^2}{2(q + S)} = \frac{40 \times 500}{87 + 4} + \frac{5 \times 87^2}{2 \times (87 + 4)} + \frac{100 \times 4^2}{2 \times (87 + 4)} \approx\approx 437 \text{ р}$$

Таким образом, в модели с дефицитом годовые издержки меньше.

Случай выполнения заявок

В случае выполнения заявок максимальный уровень запасов будет равен не *q*, а *(g — S)*.

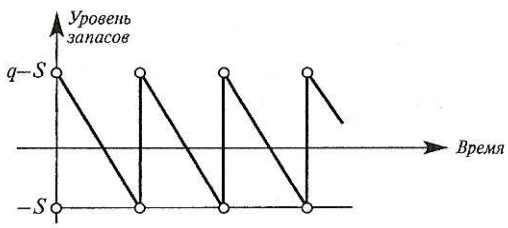


Рисунок 5.5 – Случай выполнения заявок

Издержки *TC* = подача заказов + хранение + штраф за дефицит

$$= \frac{C_0 D}{q} + + \frac{C_h (q - S)^2}{2q} + \frac{C_b S^2}{2q} \rightarrow \min,$$

где q – оптимальный размер заказа;

S – максимальный размер дефицита. Решение задачи:

$$q = \sqrt{\frac{2C_0 D}{C_h}} \times \sqrt{\frac{C_h + C_b}{C_b}}, S = \sqrt{\frac{2C_0 D}{C_b}} \times \sqrt{\frac{C_h}{C_h + C_b}}.$$

Пример. Годовой спрос D = 3000 единиц, стоимость подачи заказов C_o = 25 рублей/заказ, издержки хранения одной единицы C_h = 120 рублей/год, годовая стоимость отсутствия запасов C_b = 225 рублей/единицу. Модель с дефицитом (заявки выполняются).

Найдем издержки.

$$q = \sqrt{\frac{2C_oD}{C_h}} \times \sqrt{\frac{C_h + C_b}{C_b}} = \sqrt{\frac{2 \times 25 \times 3000}{120}} \times \sqrt{\frac{120 + 225}{225}} \approx 44 \text{ единицы.}$$

$$S = \sqrt{\frac{2C_oD}{C_b}} \times \sqrt{\frac{C_h}{C_h + C_b}} = \sqrt{\frac{2 \times 25 \times 3000}{225}} \times \sqrt{\frac{120}{120 + 225}} \approx 15 \text{ единиц.}$$

$$TC = \frac{C_oD}{q} + \frac{C_h(q - S)^2}{2q} + \frac{C_bS^2}{2q} = \frac{25 \times 3000}{44} + \frac{120 \times (44 - 15)^2}{2 \times 44} + \frac{120 \times 15^2}{2 \times 44} \approx\approx 3427$$

5.6 Неопределенность и основная модель управления запасами

Грубо говоря, основная модель — это заказ постоянного количества единиц в заранее определенные моменты времени, то есть фиксированный заказ в фиксированное время. На практике спрос часто не является постоянным, поэтому основная модель мало приспособлена для практических нужд. Будем ее видоизменять, чтобы учесть непостоянство спроса. Самое простое, что можно сделать, — отказаться от одного из двух заявленных условий.

Случай 1. Фиксированный заказ в случайное время. Как до некоторого заданного только на складе запасы понизятся заранее уровня, подается заказ на фиксированное количество единиц. Это - уровневая система повторного заказа.

Уровневая система повторного заказа позволяет реагировать на колебания спроса и подходит для самых разных категорий запас

Случай 2. Случайный заказ в фиксированное время. Заранее определяем, в какие моменты времени будут сделаны заказы. Обычно они выбираются с определенной периодичностью. При наступлении этих моментов подаются заказы, объем которых равен разности между заранее выбранным числом и количеством единиц на складе в тот момент времени. Это — циклическая система повторного заказа.

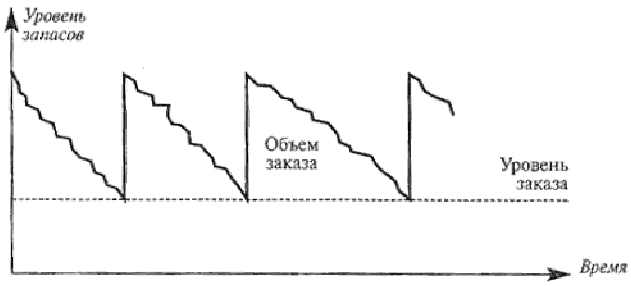


Рисунок 5.6 – Циклическая система повторного заказа

Циклическая система повторного заказа позволяет добиваться скидок за оптовые закупки, способствует ритмичной работе отдела закупок, но не способна реагировать на колебания спроса. Средний размер запаса при использовании циклической системы повторного заказа больше, чем при использовании уровневой системы повторного заказа.

Рассмотрим эти модели подробнее.

5.7 Уровневая система повторного заказа

Достижение минимальной стоимости

Чтобы учесть непостоянство спроса, вводят резервный запас.

Издержки TC = подача заказов + хранение основного запаса + хранение резервного запаса + штраф за дефицит.

Сначала считаем, что спрос постоянный. При помощи основной модели находим оптимальный размер заказа q. Именно такое количество мы и будем заказывать каждый раз. Когда заказывать? Оптимальный размер заказа q позволяет вычислить первые два слагаемых в выражении для издержек. Как выбрать резервный запас? Чем больше (меньше) резервный запас, тем меньше (больше) штраф за дефицит и тем больше (меньше) стоимость хранения резервного запаса. Методом проб и ошибок мы должны подобрать резервный запас, минимизирующий два последних слагаемых в выражении для издержек.

Пример. Средний годовой спрос D = 150 единиц за 300 рабочих дней, стоимость подачи заказов C_o = 50 рублей/заказ, издержки хранения одной единицы C_h = 12 рублей/год, годовая стоимость отсутствия запасов C_b = 20 рублей/единицу. Время поставки 4 дня.

Спрос на товар в течение поставки, шт.	0	1	2	3	4	5	6	Сумма
Частота	2	8	13	10	7	5	5	50

За время поставки спрос 6 единиц наблюдался 5 раз, спрос 5 единиц наблюдался 5 раз и т. д. Всего было 50 наблюдений. Минимизируем общую стоимость запасов.

Из основной модели оптимальный размер заказа равен:

$$q = \sqrt{\frac{2C_oD}{C_h}} = \sqrt{\frac{2 \times 50 \times 150}{12}} \approx 35 \text{ единиц.}$$

Таков объем заказа. Когда заказывать?

Издержки TC = подача заказов + хранение основного запаса + хранение резервного запаса + штраф за дефицит = 12×(резервный запас) + 20×(математическое ожидание) ≈ 424,29 + 12×(резервный запас) + 20×(математическое ожидание).

Надо подобрать резервный запас, минимизирующий два последних слагаемых.

Число циклов за год D/q — 150/35 ≈ 4,3.

Средний спрос за день 150/300 = 0,5, время поставки 4 дня. Поэтому средний спрос в течение поставки 4×0,5 = 2 (если бы получилось дробное число, то его надо округлить до ближайшего меньшего целого числа). Найдем вероятность (относительную частоту) для каждого значения спроса за время поставки. Для этого частоту каждого значения спроса разделим на 50 (общее число наблюдений).

Спрос на товар в течение поставки, шт.	0	1	2	3	4	5	6	Сумма
Частота	2	8	13	10	7	5	5	50
Вероятность	0,04	0,16	0,26	0,20	0,14	0,10	0,10	1

С помощью основной модели мы учитываем спрос 0, 1, 2 изделия за время поставки, так как средний спрос в течение поставки равен 2. Чтобы учесть спрос 3, 4, 5, 6 (а свыше 6 спрос за время поставки не наблюдался), необходим резервный запас (соответственно 1, 2, 3, 4). Мы начнем с наибольшего значения резервного запаса 4. Вычислим сумму двух последних слагаемых в выражении для издержек. После этого каждый раз мы будем понижать резервный запас на 1 и пересчитывать сумму двух последних слагаемых в выражении для издержек. Сначала сумма будет понижаться, а затем возрастать. Смена убывания на возрастание говорит о том, что резервный запас найден. Составим таблицу.

Таблица 5.1

Резервный запас	Покрытый спрос	Математическое ожидание числа нехваток запасов в течение		Стоимость, рублей/год		
		цикла	года	резервного запаса 12×(резервный запас)	нехватки запасов 20×(матожидание)	общая
4	6	0	0	12×4=48	0	48+0=48
3	5	1×0,1=0,1	4,3×0,1= =0,43	12×3=36	20×0,43= =8,6	36+8,6= =44,6
2	4	2×0,1+ 1×0,1=0,3	4,3×0,3= 1,29	12×2=24	20×1,29= =25,8	25,8+24= =49,8

Поясним, как заполняется таблица 1.3.1.

Второй столбец. Покрытый спрос = резервный запас + 2 (средний спрос за время поставки).

Третий столбец. Если покрытый спрос равен 6, то нехватки запасов не возникает. Если покрытый спрос равен 5, то возникает нехватка в 1 единицу при спросе 6. Вероятность спроса 6 равна 0,1 (см. предыдущую таблицу). Поэтому математическое ожидание нехватки 1×0,1 = 0,1. Если покрытый спрос равен 4, то возникает нехватка 2 при спросе 6 и 1 при спросе .5. Поэтому математическое ожидание нехватки 1×0,1 + 2×0,1 = 0,3. Это числа для одного цикла.

Число циклов за год — 4,3. Поэтому числа третьего столбца умножим на 4,3 и результаты запишем в четвертом столбце. Числа четвертого столбца умножим на 20 и результаты запишем в шестом столбце.

Числа первого столбца умножаем на 12 и результаты пишем в пятом столбце. Седьмой столбец равен сумме пятого и шестого столбцов.

Итоговая сумма в седьмом столбце сначала понизилась с 48 до 44,6, а затем начала повышаться. Поэтому целесообразно иметь резервный запас равный 3 (покрытый спрос 5) и нет необходимости исследовать резервный запас 1.

Издержки *TC*= 424,29 + 12×(резервный запас) + 20×(математическое ожидание) = 424,29 + 44,6 = 468,89 руб./год.

Таким образом, каждый раз, когда на складе остаются 5 единиц, надо заказывать 35 единиц.

Достижение минимального уровня обслуживания

Задается вероятность нехватки запасов в течение цикла. Тогда минимальный уровень обслуживания = 1 — вероятность нехватки запасов. По уровню обслуживания находим необходимый резервный запас.

Более высокий уровень обслуживания означает более высокий резервный запас. Но издержки на поддержание большого резервного запаса могут быть очень высокими.

Издержки *TC* = подача заказов + хранение основного запаса +хранение резервного запаса.

Пример. Вернемся к предыдущему примеру.

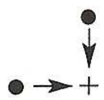
Разрешается 1 нехватка запасов в 5 циклов. Тогда вероятность нехватки запасов в течение цикла равна 1/5 = 0,2.

Минимальный уровень обслуживания равен: 1 — вероятность нехватки запасов = 1 — 0,2 = 0,8.

q = 35 единиц, средний спрос в течение поставки = 2 (см. пример 48). Заполним таблицу.

Спрос	Вероятность	Кумулятивная вероятность
0	0,04	0,04
1	0,16	0,20
2	0,26	0,46
3	0,20	0,66
4	0,14	0,80
5	0,10	0,90
6	0,10	1,00

Порядок заполнения последнего столбца: двигаемся сверху вниз и вычисляем значения по правилу:



Для получения числа данной строки 3-го столбца к числу предыдущей строки 3-го столбца прибавляем число данной строки 2-го столбца: 0,04; 0,04 + 0,16 = 0,20; 0,20 + 0,26 = 0,46 и т. д. Это — кумулятивная (накопленная) вероятность. Для проверки: последнее число всегда равно 1, Смотрим, куда в последнем столбце попадает наш уровень обслуживания 0,8. Он соответствует спросу 4, то есть резервный запас = 4 — 2 = 2. Каждый раз, когда на складе остаются 4 единицы, надо заказывать 35 единиц. Издержки ТС = 424,29 + 12×(резервный запас) = 424,29 + 12×2 = 448,29 рублей/год.

Замечание. Методы, рассмотренные в предыдущих двух примерах, можно применять и в случае, когда спрос подчиняется какому-либо распределению (например, нормальному или распределению Пуассона).

5.8 Циклическая система повторного заказа

Пусть T— интервал повторного заказа.

Издержки $TC = \frac{C_0}{T} + \frac{C_h DT}{2} \to \min . T = \sqrt{\frac{2C_0}{C_h D}}$.

После этого надо задать уровень запасов, который определяет размер подаваемого заказа. Например, если взять за уровень 120 единиц, а на момент подачи заказа на складе 45 единиц, то надо заказывать 120 - 45 = 75 единиц.

Пример. Для данных примера 48 найдем интервал повторного заказа.

$T = \sqrt{\frac{2C_0}{C_h D}} = \sqrt{\frac{2 \times 50}{12 \times 150}} \approx$ 0,24 года = 0,24×300 = 72 дня, то есть заказы надо подавать через 72 дня.

В циклической системе можно также использовать один из двух критериев: достижение минимального уровня обслуживания и достижение минимальной стоимости.

5.9 Другие вопросы управления запасами

Целью построенной нами теории была минимизация издержек. Можно было строить теорию с целью максимизации прибыли. Мы считали, что склад был безграничным. Но, скорее всего, надо вводить ограничение на площадь склада.

Запас у нас был однономенклатурным. В реальной жизни запас всегда многономенклатурный. Для упрощения ситуации здесь можно воспользоваться эффектом Парето: 20% товаров контролируют 80% стоимости запасов.

Сокращение номенклатуры продукции может обеспечить существенную экономию средств. Активное сокращение номенклатуры достигается за счет использования стандартных компонентов. Реактивное сокращение номенклатуры осуществляется периодически. Все это позволяет уменьшить расходы на содержание запасов и сократить число поставщиков.

Бывают ситуации (например, плохой урожай зерновых), когда предприятию требуется максимально увеличить свои запасы. Иногда расходы на поддержание запасов сырья могут оказаться гораздо ниже затрат на закупку из-за роста цен на сырье. В таких случаях не следует слепо полагаться на формулы, а необходимо воспользоваться опытом специалистов, отвечающих за снабжение предприятия.

Построенные модели — очень упрощенные. Если мы хотим рассмотреть более сложные ситуации, то следует воспользоваться имитационным моделированием.

Тема 6. Построение выборочной функции спроса

Функция спроса часто встречается в экономических учебниках, но при этом обычно не рассказывается, как она получена. Между тем оценить ее по эмпирическим данным не так уж трудно. Мы часто выясняем ожидаемый спрос с помощью простого приема — спрашиваем потенциальных потребителей: «Какую максимальную цену Вы заплатили бы за такой-то товар?» Пусть для определенности речь идет о конкретном учебном пособии по менеджменту. В одном из экспериментов выборка состояла из 20 опрошенных. Они назвали следующие максимально допустимые для них цены: **40, 25, 30, 50, 35, 20, 50, 32, 15, 40, 20, 40, 45, 30, 50, 25, 35, 20, 35, 40.**

Первым делом названные величины надо упорядочить в порядке возрастания. Результаты представлены в табл. 6.1. В первом столбце — номера различных численных значений (в порядке возрастания), названных потребителями. Во втором столбце приведены сами значения цены, названные ими. В третьем столбце указано, сколько раз названо то или иное значение.

Таким образом, 20 потребителей назвали 9 конкретных значений цены (максимально допустимых, или приемлемых для них значений), каждое из значений, как видно из третьего столбца, названо от 1 до 4 раз. Теперь легко построить выборочную функцию спроса в зависимости от цены. Она будет представлена в четвертом столбце, который заполним снизу вверх. Если мы будем предлагать товар по цене свыше 50 руб., то его не купит никто из опрошенных. При цене 50 руб. появляются 3 покупателя. Записываем 3 в четвертый столбец в девятую строку. А если цену понизить до 45? Тогда товар купят четверо - тот единственный, для кого максимально возможная цена — 45, и те трое, кто был согласен на большую цену - 50 руб. Таким образом, легко заполнить столбец 4, действуя по правилу: значение в клетке четвертого столбца равно сумме значений в нахо дящейся слева клетке третьего столбца и в лежащей снизу клетк четвертого столбца. Например, за 30 руб. купят товар 14 человек, за 20 руб. —19.

***Таблица 6.1.* Эмпирическая оценка функции спроса и ее использование**

№ п/п	Цена	Ni	Спрос	При-быль	При-быль (р-	При-быль
1	15	1	20	100	0	-
2	20	3	19	190	95	
3	25	2	16	240	160	0
4	30	2	14	280	210	70
5	32	1	12	264	204	84
6	35	3	11	275	220	ПО
7	40	4	8	240	200	120
8	45	1	4	140	120	80
9	50	3	3	120	105	75

Зависимость спроса от цены — это зависимость четвертого столбца от второго. Табл. 6.1 дает нам девять точек такой зависимости. Зависимость можно представить на рисунке, в координатах «спрос - цена». Если абсцисса — это спрос, а ордината — цена, тог девять точек на кривой спроса, перечисленные в порядке возрастания абсциссы, имеют вид: (3; 50), (4; 45), (8; 40), (11; 35), (12; 32), (14; 30), (16; 25), (19; 20), (20; 15).

Эти девять точек можно использовать для построения кривой спроса каким-либо графическим или расчетным способом, например, методом наименьших квадратов (см. главу 5). Кривая спроса убывает, имея направления от левого верхнего угла к правому. Однако заметны отклонения от гладкого вида функции, связанные, в частности с естественным пристрастием потребителей к круглым числам. Заметьте, все опрошенные, кроме одного, назвали числа, кратные 5 руб.

Данные табл. 6.1 могут быть использованы для выбора цены продавцом-монополистом (или действующем на рынке монополи стической конкуренции). Пусть расходы на изготовление единицы товара равны 10 руб. (например, оптовая цена книги — 10 руб.). По какой цене ее продавать на том рынке, функцию спроса для которого мы только что нашли? Для ответа на этот вопрос вычислим суммарную прибыль, т.е. произведение прибыли на одном экземпляре $(p$ - 10) на число проданных (точнее, запрошенных) экземпляров $Dip)$. Результаты приведены в пятом столбце. Максимальная прибыль, равная 280 руб., достигается при цене 30 руб. за экземпляр. При этом из 20 потенциальных покупателей окажутся в состоянии заплатить за книгу 14, т.е. 70%.

Если же удельные издержки производства, приходящиеся на одну книгу (или оптовая цена), повысятся до 15 руб., то данные столбца 6 показывают, что максимальная прибыль, равная 220 руб. (она, разумеется, меньше, чем в предыдущем случае), достигается при более высокой цене — 35 руб. Эта цена доступна 11 потенциальным покупателям, т.е. 55% от всех возможных покупателей. При дальнейшем повышении издержек, скажем, до 25 руб., как вытекает из данных столбца 7, максимальная прибыль, равная 120 руб., достигается при цене 40 руб. за единицу товара, что доступно 8 лицам, т.е. покупателей. Отметьте, что при повышении оптовой цены на руб. оказалось выгодным увеличить розничную лишь на 5, поскольку более резкое повышение привело бы к такому сокращению Р а, которое перекрыло бы эффект от повышения удельной прибыли, приходящейся на одну проданную книгу). Представляет интерес анализ оптимального объема выпуска различных значениях удельных издержек (табл. 6.2).

Таблица 6.2 **Прибыль при различных значениях издержек**

№ п/	Цена	Спрос	Прибыль	Прибыль	Прибыль	Прибыль	Прибыль
1	15	20	200				
2	20	19	285	0			
3	25	16	320	80			
4	30	14	350*	140	0		
5	32	12	324	144	24		
6	35	11	330	165*	55	0	
7	40	8	280	160	80*	40	0
8	45	4	160	100	60	40	20
9	50	3	135	90	60	45*	30*

В табл. 6.2 звездочками указаны максимальные значения прибыли при том или ином значении издержек, не включенном в табл. 6.1. Для легкости обозрения результаты об оптимальных объемах выпуска и соответствующих ценах из табл. 6.1 и 6.2 приведены в табл. 6.3.

Таблица 6.3 **Зависимость оптимального выпуска и цены от издержек**

Издержки	5	10	15	20	25	30	35	40
Оптимальный	14	14	11	11	8	8	3	3
Цена	30	30	35	35	40	40	50	50

Как видно из табл. 6.3, с ростом издержек оптимальный выпуск падает, а цена растет. При этом изменение издержек на 5 единиц может вызывать, а может и не вызывать повышения цены. В этом проявляется микроструктура функции спроса — небольшое повышение цены может привести к тому, что значительные группы покупателей откажутся от покупок, и прибыль упадет.

Этот эффект напоминает известное в экономической теории разделение налогового бремени между производителем и потребителем. Неверно говорить, что производитель перекладывает издержки или, конкретно, налоги, на потребителя, повышая цену на их величину, поскольку при этом сокращается спрос (и выпуск), а потому и прибыль производителя.

Дальнейшее ясно — если оптовая цена будет повышаться, то и дающая максимальную прибыль розничная цена также будет повышаться, и все меньшая доля покупателей сможет приобрести товар. Крайняя точка — оптовая цена, равная 45 руб. Тогда только трое (15%) купят товар за 50 руб., а прибыль продавца составит только 15 руб. Наглядно видно, что повышение издержек производства приводит к ориентации производителя на наиболее богатые слои населения, но и повышение цен (до оптимального для монополиста-производителя уровня) не приводит к повышению прибыли, напротив, она снижается, и при этом большинство потенциальных потребителей не в состоянии купить товар. Таково влияние инфляции издержек на экономическую жизнь.

Отметим, что рыночные структуры не в состоянии обеспечить всех желающих - это просто не выгодно. Так, из 20 опрошенных лишь 14, т.е. 70%, могут рассчитывать на покупку даже при минимальных издержках и ценах.

7. Решение типовых задач междисциплинарного экзамена

7.1 Функция спроса часто встречается в экономических учебниках, но при этом обычно не рассказывается, как она получена. Между тем оценить ее по эмпирическим данным не так уж трудно. Мы часто выясняем ожидаемый спрос с помощью простого приема — спрашиваем потенциальных потребителей: «Какую максимальную цену Вы заплатили бы за такой-то товар?» Пусть для определенности речь идет о конкретном учебном пособии по менеджменту. В одном из экспериментов выборка состояла из 20 опрошенных. Они назвали следующие максимально допустимые для них цены: 40, 25, 30, 50, 35, 20, 50, 32, 15, 40, 20, 40, 45, 30, 50, 25, 35, 20, 35, 40.

Первым делом названные величины надо упорядочить в порядке возрастания. Результаты представлены в табл. 5.1. В первом столбце — номера различных численных значений (в порядке возрастания), названных потребителями. Во втором столбце приведены сами значения цены, названные ими. В третьем столбце указано, сколько раз названо то или иное значение.

Таким образом, 20 потребителей назвали 9 конкретных значений цены (максимально допустимых, или приемлемых для них значений), каждое из значений, как видно из третьего столбца, названо от 1 до 4 раз. Теперь легко построить выборочную функцию спроса в зависимости от цены. Она будет представлена в четвертом столбце, который заполним снизу вверх. Если мы будем предлагать товар поцене свыше 50 руб., то его не купит никто из опрошенных. При цене 50 руб. появляются 3 покупателя. Записываем 3 в четвертый столбец в девятую строку. А если цену понизить до 45? Тогда товар купят четверо - тот единственный, для кого максимально возможная цена — 45, и те трое, кто был согласен на большую цену - 50 руб. Таким образом, легко заполнить столбец 4, действуя по правилу: значение в клетке четвертого столбца равно сумме значений в нахо дящейся слева клетке третьего столбца и в лежащей снизу клетк четвертого столбца. Например, за 30 руб. купят товар 14 человек, за 20 руб. —19.

***Таблица 7.1.* Эмпирическая оценка функции спроса и ее использование**

№ п/п	Цена *pi*	*Ni*	Спрос	Прибыль (p-10)D(pi)	Прибыль (p-15)D(pi)	Прибыль *(p-25)D(pi)*
1	15	1	20	100	0	-
2	20	3	19	190	95	-
3	25	2	16	240	160	0
4	30	2	14	280	210	70
5	32	1	12	264	204	84
6	35	3	11	275	220	110
7	40	4	8	240	200	120

8	45	1	4	140	120	80
9	50	3	3	120	105	75

Зависимость спроса от цены — это зависимость четвертого столбца от второго. Табл. 5.2 дает нам девять точек такой зависимости. Зависимость можно представить на рисунке, в координатах «спрос - цена». Если абсцисса — это спрос, а ордината — цена, то девять точек на кривой спроса, перечисленные в порядке возрастания абсциссы, имеют вид: (3; 50), (4; 45), (8; 40), (11; 35), (12; 32), (14; 30), (16; 25), (19; 20), (20; 15).

Эти девять точек можно использовать для построения кривой спроса каким-либо графическим или расчетным способом, например, методом наименьших квадратов (см. курс «Эконометрика»). Кривая спроса убывает, имея направления от левого верхнего угла к правому. Однако заметны отклонения от гладкого вида функции, связанные, в частности с естественным пристрастием потребителей к круглым числам. Заметьте, все опрошенные, кроме одного, назвали числа, кратные 5 руб.

Данные табл. 5.1 могут быть использованы для выбора цены продавцом-монополистом. Пусть расходы на изготовление единицы товара равны 10 руб. (например, оптовая цена книги — 10 руб.). По какой цене ее продавать на том рынке, функцию спроса для которого мы только что нашли? Для ответа на этот вопрос вычислим суммарную прибыль, т.е. произведение прибыли на одном экземпляре (p - 10) на число проданных (точнее, запрошенных) экземпляров *Dpi).* Результаты приведены в пятом столбце. Максимальная прибыль, равная 280 руб., достигается при цене 30 руб. за экземпляр. При этом из 20 потенциальных покупателей окажутся в состоянии заплатить за книгу 14, т.е. 70%. Если же удельные издержки производства, приходящиеся на одну книгу (или оптовая цена), повысятся до 15 руб., то данные столбца 6 показывают, что максимальная прибыль, равная 220 руб. (она, разумеется, меньше, чем в предыдущем случае), достигается при более высокой цене — 35 руб. Эта цена доступна 11 потенциальным покупателям, т.е. 55% от всех возможных покупателей. При дальнейшем повышении издержек, скажем, до 25 руб., как вытекает из данных столбца

7, максимальная прибыль, равная 120 руб., достигается при цене 40 руб. за единицу товара, что доступно 8 лицам, т.е. покупателей. Отметьте, что при повышении оптовой цены на 10 руб. оказалось выгодным увеличить розничную лишь на 5, поскольку более резкое повышение привело бы к такому сокращению спроса, которое перекрыло бы эффект от повышения удельной прибыли, приходящейся на одну проданную книгу).

Таблица 7.2 **Прибыль при различных значениях издержек**

№ п/п	Цена *Pi*	Спрос D(pi)	Прибыль (р-5) D(pi)	Прибыль (р -20) D(pi)	Прибыль (Р-30) D(pi)	Прибыль (Р-35) D(pi)	Прибыль (Р-40) D(pi)
1	15	20	200	-	-	-	
2	20	19	285	0	-	-	
3	25	16	320	80	-	_	_
4	30	14	350*	140	0	-	_
S	32	12	324	144	24	-	_
6	35	11	330	165*	55	0	_
7	40	8	280	160	80*	40	0
8	45	4	160	100	60	40	20
9	50	3	135	90	60	45*	30*

В табл. 7.2 звездочками указаны максимальные значения прибыли при том или ином значении издержек, не включенном в табл. 5.1.

7.2 Условие задачи. Компания, специализирующаяся на программных разработках, планирует принять участие в тендере на получение госзаказа на разработку информационной системы. Тендер проводится закрытым способом, т.е. предложения подаются участниками в запечатанных конвертах и неизвестны другим участникам. По оценкам компании, участие в тендере обойдется в 5 000 руб., а выполнение заказа в 95 000 руб. Из опыта предыдущих тендеров известно, что с вероятностью 30% конкуренции вообще не будет. Кроме того, известно, что цена подобного тендера имеет условные вероятности представленные в таблице.

Цена тендера	Вероятность
Менее 115 000	0,2

От 115 000 до 120 000	0,4
От 110 000 до 125 000	0,3
Более 125 000	0,1

Необходимо принять решение, участвовать ли в аукционе и, если да, то с какой ценой. Необходимо выбрать решение, которое максимизирует ожидаемую прибыль. Как изменится решение, если перед подачей заявки станет известно, что несколько фирм-конкурентов участвует в тендере.

Решение. Рассмотрим на примере данной задачи три составляющие для ситуации принятия решений в условиях вероятностно заданной информации.

1. Компания должна принять решение - участвовать в аукционе или нет. Если она все же принимает решение участвовать, то возникает вопрос, с какой ценой. Ясно, что минимальные затраты в этом случае составят 5 000 + 95 000 = 100 000. Отсюда следует, что цена менее 100 000 лишена смысла (отсутствие прибыли). К сожалению, отсутствие полной вероятностной информации о возможных ценах аукциона приводит к тому, что мы имеем возможность рассмотреть лишь стратегии участия с ценами 115 000, 120 000 и 125 000.
2. Далее, нам необходимо описать исходы и их вероятности. Поскольку компания точно знает расходы на участие в тендере (5 000) и стоимость выполнения работы (95 000), то вся неопределенность заключается в стратегиях, которые изберут конкуренты. Мы предполагаем, что поведение других игроков подчиняется тем же закономерностям, что и раньше, поэтому мы используем вероятностные данные по ценам предыдущих тендеров.
3. Теперь необходимо количественно оценить (вычислить стоимостную оценку) ситуации для компании при различных выборах стратегии и различных вариантах разрешения неопределенности в поведении других участников. Если компания принимает решение не участвовать в тендере, то нет ни затрат, ни прибыли, т.е. оценка ситуации равна 0. Если компания участвует в тендере и не проходит по цене, то она теряет 5 000. Если компания участвует в тендере с ценой B (115, 120, 125 тыс.) и проходит по цене, то ее ожидаемая прибыль составит B-100 000, где, как отмечалось ранее, 100 000 = 95 000 + 5 000.

Оценки ситуации иногда удобно записать в виде платежной матрицы, где строки соответствуют стратегиям, которые выбирает сторона, принимающая решение, а столбцы соответствуют стратегиям конкурентов. Элементы данной матрицы A = \atj отражают стоимостную оценку ситуаций при различных сочетаниях стратегий участников тендера (см. табл.).

	Нет конкуренции	Не более 115	От 115 до 120	От 120 до 125	Более 125
Неучастие	0	0	0	0	0
115 000	15 000	-5 000	15 000	15 000	15 000
120 000	20 000	-5 000	-5 000	20 000	20 000
125 000	25 000	-5 000	-5 000	-5 000	25 000
Вероятности	0,3	0,14 = 0,7 * 0,2	0,28 =0,7 * 0,4	0,21 = 0,7 * 0,3	0,07 =0,7*0,1

Наиболее универсальным подходом для выбора оптимальной стратегии является подход, при котором выбирается стратегия, обеспечивающая максимальную ожидаемую оценку.

Смысл данной оценки состоит в вычислении математического ожидания выигрыша первого игрока при применении им стратегии i. Опираясь на данные платежной матрицы, вычислим оценки для всех стратегий первого игрока, результаты оформим в виде табл.

Стратегии игрока	Расчет ожидаемой прибыли	Значение прибыли
Не участвовать	0(1)	0
115 000	15 000(0,3 + 0,28 + 0,21 + 0,07) - 5	12 200
120 000	20 000(0,3 + 0,21 + 0,07) - 5 000(0,14 +	9 500
125 000	25 000(0,3 + 0,07)- 5 000(0,14 + 0,28 +	6 100

Необходимо отметить, что значение ожидаемой прибыли не есть значение реальных выигрышей, а есть среднеожидаемое значение выигрышей при различных стратегиях других участников. Таким образом, наиболее эффективной представляется стратегия участия в аукционе с ценой 115 000, так как именно этой стратегии соответствует максимальное значение ожидаемой прибыли.

Дерево решений. Дерево решений представляет из себя некий графический инструмент, который помогает производить действия, описанные в изложенном примере, а именно: описание возможных стратегий игрока, принимающего решение, описание неопределенных исходов(неизвестные стратегии второй стороны) и их вероятностей, вычисление выигрышей по стратегиям первого игрока, выбор стратегии с максимальным значением выигрыша. Как правило, применение этого графического средства предполагает использование следующих соглашений:

1. деревья решений состоят из вершин (круги, квадраты и треугольники) и ветвей (линии);

2. вершины соответствуют определенным моментам времени. Вершины-решения (квадраты) соответствуют моментам времени, когда ЛПР (лицо принимающее решение) принимает решение. Вершины-вероятности (круги) соответствуют моментам времени, когда разрешается одна из неопределенностей. Оконечные вершины (треугольники) соответствуют окончанию задачи, когда

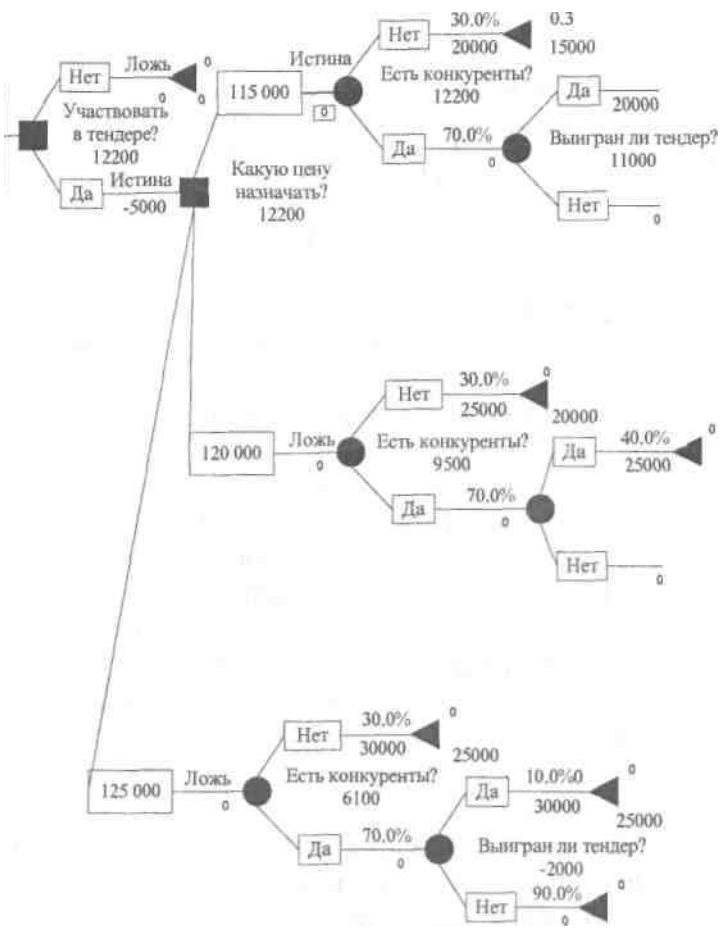
все решения приняты, все неопределенности разрешились и все платежи произошли;

3. развитие ситуации во времени происходит согласно графической схеме слева направо;

4. ветви, идущие из вершин-решений, соответствуют различным возможным решениям. Ветви, идущие из вершин-вероятностей, соответствуют различным возможным вариантам разрешения неопределенности и не являются объектом чьего-либо управления;

5. вероятности соответствуют ветвям, исходящим из вершин-вероятностей. Эти вероятности являются условными вероятностями при условии свершения события, соответствующего вершине, из которой они исходят. Поэтому сумма вероятностей по всем ветвям, исходящим из одной вершины, равна 1; 6) каждой оконечной вершине соответствуют два числовых значения. Первое — это вероятность прихода в данную оконечную вершину и второе— значение платежа, соответствующее данному сценарию развития событий.

На рисунке изображено дерево решений, соответствующее рассмотренной задаче о проведении тендера **(см. Приложение 1).**



Процедура принятия оптимального решения после построения дерева решений производится методом обратного хода и заключается в следующем.

1. Для каждой вероятностной вершины (круги) вычислим среднеожидаемое значение выигрыша по всем альтернативам, исходящим изданной вершины. Например,

0.8(15 000) + (0.2(-5 000) = 11 000.

Далее для другой вершины

0.3(15 000) + 0.7(11 000) = 12 200.

2. Для каждой вершины-решения мы приписываем максимальное значение из ожидаемых значений выигрыша, соответствующих различным вариантам решения, исходящим из данной вершины. То решение, на котором достигается максимум, помечается пометкой «истина», иначе помечаем решение как «ложь». После данной процедуры расстановки пометок на вершинах-решениях оптимальная стратегия определяется путем следования слева направо по вершинам, помеченным как «истина».

Рассмотренный пример относится к так называемым одношаговым играм, в которых на первом шаге принимаются все решения, а далее разрешаются все неопределенности. В более сложных случаях принятие решений чередуется с разрешением некоторых неопределенностей, причем решения, принятые на очередном шаге, порождают свое множество неопределенностей (неопределенных факторов), которые далее разрешаются. Такие ситуации принятия решений называются многошаговыми (позиционными) играми. Рассмотрим пример подобной ситуации.

7.3. Условие задачи. Компания решает вопрос о представлении нового продукта на общенациональный рынок. Неопределенность заключается в том, как отреагирует рынок на новый продукт. Рассматривается вопрос об апробации нового продукта первоначально на некотором региональном рынке. Таким образом, первоначальное решение, которое необходимо принять компании — это проводить ли первоначальный маркетинг продукта на региональном уровне. Компания предполагает, что выход на региональный уровень потребует затрат на 3 млн. руб., а выход на общенациональный рынок потребует вложения 90 млн. руб. Если не проводить первоначальных пробных продаж на региональном уровне, то решение о выходе на общенациональный рынок можно принять незамедлительно.

Компания рассматривает результаты продаж как успешные, средние или отрицательные в зависимости от объемов продаж. Для регионального уровня этим градациям соответствуют объемы в 200, 100 и 30 тыс. экземпляров, а для общенационального 6 000, 3 000 и 900 тыс. экземпляров соответственно. Исходя из данных по результатам региональных тестирований аналогичных видов продукции компания оценивает вероятности указанных трех исходов как 0,3, 0,6 и 0,1. Кроме того, исследуя данные о соотношении результатов региональных продаж с последующими продажами на общенациональном рынке, компания сумела оценить следующие условные вероятности (см. табл.).

			Условные вероятности продаж на общенациональном рынке		
			успешные	средние	отрицательные
Вероятности продаж на	0,3	Успешные	0,8	0,15	0,05
	0,6	Средние	0,3	0,5	0,2
	0,1	Отрицатель-	0,05	0,25	0,7

Кроме этого известно, что каждая продажа приносит прибыль в 18 руб. как при продаже на региональном рынке, так и на общенациональном.

Задача состоит в принятии обоснованной стратегии выхода (или невыхода) на рынок с новой товарной позицией.

Решение. (**см. Приложение 2).**

Как и в предыдущей задаче, рассмотрим три основные составляющие для ситуаций подобного типа. Далее построим дерево решений для данной задачи (см. рис.).

Приведем расчет характеристик дерева решений:

$$(0{,}8)(0{,}3) + (0{,}3)(0{,}6) + (0{,}05)(0{,}1) = 0{,}425$$

$$(0{,}15)(0{,}3) + (0{,}5)(0{,}6) + (0{,}25)(0{,}1) = 0{,}37$$

$$(0{,}05)(0{,}3) + (0{,}2)(0{,}6) + (0{,}7)(0{,}1) = 0{,}205$$

$$6\ 000 \cdot 18 = 108\ 000$$

$$3\ 000 \cdot 18 = 54\ 000$$

$$900 \cdot 18 = 16\ 200$$

$$108\ 000 - 90\ 000 = 18\ 000$$

$$54\ 000 - 90\ 000 = -72\ 000$$

$$16\ 200 - 90\ 000 = -73\ 800 \quad 0{,}425 \cdot 108\ 000 + 0{,}37 \cdot 54\ 000 + 0{,}205 \cdot 16\ 200 - 90\ 000 = -20\ 799$$

$$200 \cdot 18 = 3\ 600$$

$$100 \cdot 18 = 1\ 800$$

$$30 \cdot 18 = 540$$

$$0{,}8 \cdot 108\ 000 + 0{,}15 \cdot 54\ 000 + 0{,}05 \cdot 16\ 200 - 90\ 000 + 3\ 600 - 3\ 000 = 5\ 910$$

$$3\ 600 - 3\ 000 = 600$$

$$108\ 000 - 90\ 000 + 3\ 600 - 3\ 000 = 18\ 600$$

$$54\ 000 - 90\ 000 + 3\ 600 - 3\ 000 = -35\ 400$$

$$16\ 200 - 90\ 000 + 3\ 600 - 3\ 000 = -73\ 200$$

$$0{,}8 \cdot 0.3 = 0{,}24$$

$$0{,}15 \cdot 0{,}3 = 0{,}045$$

$$0{,}05 \cdot 0{,}3 = 0{,}015$$

$$0{,}3 \cdot 108\ 000 + 0{,}5 \cdot 54\ 000 + 0{,}2 \cdot 16\ 200 - 90\ 000 + 1\ 800 - 3\ 000 = -28\ 560$$

$$1\ 800 - 3\ 000 = -1\ 200$$

$$108\ 000 - 90\ 000 + 1\ 800 - 3\ 000 = 16\ 800$$

$$54\ 000 - 90\ 000 + 1\ 800 - 3\ 000 = -37\ 200$$

$$16\ 200 - 90\ 000 + 1\ 800 - 3\ 000 = -75\ 000$$

$$0{,}05 \cdot 108\ 000 + 0{,}25 \cdot 54\ 000 + 0{,}7 \cdot 16\ 200 - 90\ 000 + 540 - 3\ 000 = -62\ 220$$

$$540 - 3\ 000 = -2\ 460$$

$$108\ 000 - 90\ 000 + 540 - 3\ 000 = 15\ 540$$

$$54\ 000 - 90\ 000 + 540 \cdot 3\ 000 = -38\ 460$$

$$16\ 200 - 90\ 000 + 540 \cdot 3\ 000 = -76\ 260.$$

Таким образом, анализируя дерево решений, можно сформулировать оптимальную стратегию компании, а именно: следует произвести предварительную продажу на региональном рынке и развернуть продажи на общенациональном уровне, только если результаты на региональном уровне оказались успешными.

8.Типовые варианты задачи №3 междисциплинарного экзамена

1. В интересах продвижения товара на рынок продовцом-монополистом были проведены маркетинговые исследования. На вопрос «Какую максимальную цену Вы заплатили бы за товар?» потенциальные потребители назвали следующие максимально допустимые для них цены 40, 45, 70, 50, 35, 20, 50, 72, 15, 40, 80, 40, 45, 90, 50, 25, 35, 50, 35 и 40 руб. Выборку считать репрезативной. Необходимо предложить решение по цене товара, которое максимизирует ожидаемую прибыль. Расчеты провести для себестоимости производства 20 руб. и 30 руб.

2. Инвестор принимает решение о вложении средств в 3 инвестиционных фонда открытого типа: простой, специальный и глобальный. Прибыль зависит от ситуации на рынке. Аналитики считают, что имеется 20%-ая вероятность ухудшения ситуации на рынке ценных бумаг, 40%-ая - что рынок останется умеренным и 40%-ая - рынок будет возрастать. Значения процентов прибыли от суммы инвестиции при трех возможностях развития рынка представлены в таблице. Обосновать решения исходя из критериев Гурвица, Сэвиджа и Лапласа. Для критерия Гурвица принять значение коэффициента риска 0,3.

Альтернатива (фонды)	Процент прибыли от инвестиций (%)		
	Ухудшающийся рынок	Умеренный рынок	Растущий рынок
Простой	3	5	20
Специальный	-10	4	25
Глобальный	1	4	30

3. У компании А существует 4 варианта инвестирования средств в различные отрасли. На рынке возможна одна из 4 ситуаций (В1, В2, В3 и В4) которая может привести к получению прибыли или убытков (в млн. руб.). Для известной платежной матрицы обосновать решения исходя из критериев Гурвица, Сэвиджа и Лапласа. Для критерия Гурвица принять значение коэффициента риска 0,3. Как изменится решение, если появится информация о возможных вероятностях появления ситуаций В1, В2, В3 и В4 - 0,2, 0,2, 0,1 и 0,5 соответственно.

	В1	В2	В3	В4
А1	-3	5	2	-3
А2	-6	-7	3	0
А3	3	4	-3	-14
А4	-2	0	-4	8

4. Компании А и В занимают монопольное положение на некотором сегменте фондового рынка. У компании А существует две стратегии поведения на данном рынке, у компании В - четыре. Платежная матрица при различных сочетаниях поведения компаний А и В представлена в таблице (в млн. руб.). Считать возможную игру как игру двух лиц с нулевой суммой. Может ли привести данная платежная матрица к равновесной ситуации (имеется ли в игре седловая точка)? Решить игру в смешанных стратегиях за компанию А.

	B1	B2	B3	B4
A1	0	6	-3	4
A2	4	-2	3	-4

5. Компания решает вопрос о представлении нового продукта на общенациональный рынок. Неопределенность заключается в том, как отреагирует рынок на новый продукт. Рассматривается вопрос об апробации нового продукта первоначально на некотором региональном рынке. Выход на региональный уровень потребует затрат 1 млн. руб., а на общенациональный рынок потребует вложения 40 млн. руб. Если не проводить первоначальных пробных продаж на региональном уровне, то решение о выходе на общенациональный рынок можно принять незамедлительно. Компания рассматривает результаты продаж как успешные, средние или отрицательные в зависимости от объемов продаж. Для регионального уровня этим градациям соответствуют объемы в 100, 50 и 10 тыс. экземпляров, а для общенационального 1 000, 500 и 70 тыс. экземпляре соответственно. Исходя из данных по результатам региональных тестирований аналогичных видов продукции, компания оценивает вероятности указанных трех исходов как 0,2, 0,6 и 0,2. Кроме того, исследуя соотношения результатов региональных продаж с последующими продажами на общенациональном рынке, компания сумела оценить соответствующие условные вероятности, приведенные в таблице.

		Условные вероятности продаж на общенациональном рынке			
			успешные	средние	отрицательные
Вероятности продаж на региональном рынке	0,2	Успешные	0,8	0,15	0,05
	0,6	Средние	0,3	0,5	0,2
	0,2	Отрицательные	0,05	0,25	0,7

Каждая продажа приносит прибыль в 10 руб. как при продаже на региональном рынке, так и на общенациональном. Обосновать стратегию выхода (или невыхода) на рынок с новой товарной позицией.

6. Компания, специализирующаяся на программных разработках, планирует принять участие в тендере на получение госзаказа на разработку информационной системы. Тендер проводится закрытым способом, т.е. предложения подаются участниками в запечатанных конвертах и неизвестны другим участникам. По оценкам компании, участие в тендере обойдется в 7 000 руб., а выполнение заказа в 80 000 руб. Из опыта предыдущих тендеров известно, что с вероятностью 40% конкуренции вообще не будет. Кроме того, известно, что цена подобного тендера имеет условные вероятности представленные в таблице.

Цена тендера	Вероятность
Менее 100 000	0,2
От 100 000 до 110 000	0,4
От 110 000 до 120 000	0,3
Более 120 000	0,1

Необходимо принять решение, участвовать ли в аукционе и, если да, то с какой ценой. Необходимо выбрать решение, которое максимизирует ожидаемую прибыль. Как изменится решение, если перед подачей заявки станет известно, что несколько фирм-конкурентов участвует в тендере.

Основная литература

1. Математические методы и модели исследования операций. Учебник. Под ред. д.э.н. прф. В.А.Колемаева. – М.: Издательство «ЮНИТИ-ДАНА», 2008. – 592 с.
2. О.А. Косоруков, А.В. Мищенко. Исследование операций. Учебник для ВУЗов. Под общ. ред. д.э.н. прф. Н.П.Тихомирова. – М.: Издательство «Экзамен», 2003. – 448 с.
3. Е.В. Бережная, В.И. Бережной. Математические методы моделирование экономических систем. Учебное пособие. – М.:Финансы и статистика. 2001. – 368с.
4. Г.И.Просветов. Математические методы и модели в экономике: задачи и решения. Учебно-практическое пособие. – М.: Альфа-пресс, 2008. – 344с.

Дополнительная литература

1. Введение в математическое моделирование. Учебное пособие. Под ред. П.В. Трусова. – М.: Издательство «Логос», 2005. – 440 с.
2. Е.С. Вентцель. Исследование операций. – М.: «Советское радио», 1972. – 552 с.