

РАЗДЕЛ 1. АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Лекция 2. Принципы аналитического моделирования

Допустим, что нам нужно решить «Задачу о разрезаниях», то есть надо предсказать, сколько потребуется разрезов в виде прямых линий, чтобы разделить фигуру (**рис. 2.1**) на заданное число кусков (для примера достаточно, чтобы фигура была выпуклой).

Попробуем решить эту задачу вручную.

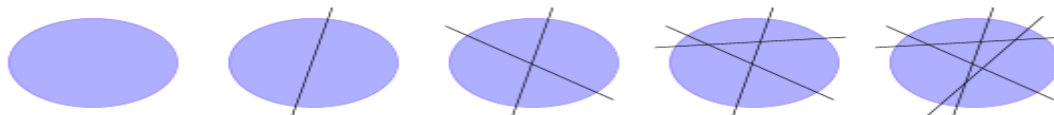


Рис. 2.1. Задача о разрезании фигуры на заданное число кусков

Из **рис. 2.1** видно, что при 0 разрезах образуется 1 кусок, при 1 разрезе образуется 2 куска, при двух – 4, при трех – 7, при четырех – 11. Можете ли вы сейчас сказать наперед, сколько потребуется разрезов для образования, например, 821 куска? Наверное, нет! Почему вы затрудняетесь? – Вам неизвестна закономерность $K = f(P)$, где K – количество кусков, P – количество разрезов. Как обнаружить закономерность? Составим таблицу, связывающую известные нам числа кусков и разрезов.

Таблица 2.1

Таблица соответствия разрезов и получающихся фрагментов фигуры

Разрезы	0	1	2	3	4	...
Куски	1	2	4	7	11	...

Пока закономерность не ясна. Поэтому рассмотрим разности между отдельными экспериментами, посмотрим, чем отличается результат одного эксперимента от другого. Поняв разницу, мы найдем способ перехода от одного результата к другому, то есть закон, связывающий K и P .

Таблица 2.2

Таблица соответствия разрезов и получающихся фрагментов фигуры

Разрезы	0	1	2	3	4	...
Куски	1	2	4	7	11	...
Первые разности	–	$1 = 2 - 1$	$2 = 4 - 2$	$3 = 7 - 4$	$4 = 11 - 7$...

Уже кое-какая закономерность проявилась, не правда ли? Вычислим вторые разности.

Таблица 2.3

Таблица соответствия разрезов и получающихся фрагментов фигуры

Разрезы	0	1	2	3	4	...
Куски	1	2	4	7	11	...

Первые разности	–	1 = 2 – 1	2 = 4 – 2	3 = 7 – 4	4 = 11 – 7
------------------------	---	-----------	-----------	-----------	------------	-----	-----

Вторые разности	–	1 = 2 – 1	1 = 3 – 2	1 = 4 – 3
------------------------	---	-----------	-----------	-----------	-----	-----

Очевидно, что далее продолжать процедуру вычисления разностей смысла нет.

Теперь все просто. Функция f называется **производящей функцией**. Если она линейна, то первые разности равны между собой. Если она квадратичная, то вторые разности равны между собой. И так далее.

Функция f есть частный случай формулы Ньютона:

$$a + \frac{b \cdot n}{1} + \frac{c \cdot n(n-1)}{1 \cdot 2} + \frac{d \cdot n(n-1)(n-2)}{1 \cdot 2 \cdot 3} + \frac{e \cdot n(n-1)(n-2)(n-3)}{1 \cdot 2 \cdot 3 \cdot 4} + \dots$$

Коэффициенты a, b, c, d, e для нашей *квадратичной* функции f находятся в первых ячейках строк экспериментальной таблицы 2.4.

Таблица 2.4

Таблица соответствия разрезов и получающихся фрагментов фигуры

Разрезы	0	1	2	3	4	...	
Куски	$a = 1$	2	4	7	11	...	
Первые разности	–	$b = 1$	2 = 4 – 2	3 = 7 – 4	4 = 11 – 7
Вторые разности	–	$c = 1$	1 = 3 – 2	1 = 4 – 3	

Итак, закономерность есть, и она такова:

$$K = a + b \cdot p + c \cdot p \cdot (p-1)/2 = 1 + p + p \cdot (p-1)/2 = 0.5 \cdot p^2 + 0.5 \cdot p + 1.$$

Теперь, когда закономерность определена, можно решить обратную задачу и ответить на поставленный вопрос: сколько надо выполнить разрезов, чтобы получить 821 кусок? $K = 821, K = 0.5 \cdot p^2 + 0.5 \cdot p + 1, p = ?$

Решаем квадратное уравнение $821 = 0.5 \cdot p^2 + 0.5 \cdot p + 1$, находим корни: $p = 40$.

Подведем итоги. Сразу угадать решение мы не смогли. Поставить эксперимент оказалось затруднительно. Пришлось построить модель, то есть найти закономерность между переменными. Модель получилась в виде уравнения. Добавив к уравнению вопрос и уравнение, отражающее известное условие, образовали задачу. Поскольку задача оказалась типового вида (канонического), то ее удалось решить одним из известных методов. Поэтому задача оказалась решена. И еще очень важно отметить, что модель отражает причинно-следственные связи. Между переменными построенной модели действительно есть крепкая связь. Изменение одной переменной влечет за собой изменение другой. Мы ранее сказали, что «модель играет системообразующую и смыслообразующую роль в научном познании, позволяет понять

явление, структуру изучаемого объекта, установить связь причины и следствия между собой». Это означает, что модель позволяет определить причины явлений, характер взаимодействия ее составляющих. Модель связывает причины и следствия через законы, то есть переменные связываются между собой через уравнения или выражения.

Но!!! Сама математика не дает возможности выводить из результатов экспериментов какие-либо законы или модели, как это может показаться после рассмотренного только что примера. Математика это только способ изучения объекта, явления, и, причем, один из нескольких возможных способов мышления. Есть еще, например, религиозный способ или способ, которым пользуются художники, эмоционально-интуитивный, с помощью этих способов тоже познают мир, природу, людей, себя.

Итак, гипотезу о связи переменных A и B надо вносить самому исследователю, извне, сверх того. А как это делает человек? Посоветовать внести гипотезу легко, но как научить этому, объяснить это действие, а значит, опять-таки, как его формализовать? Подробно мы покажем это в будущем курсе «Моделирование систем искусственного интеллекта».

А вот почему это надо делать извне, отдельно, дополнительно и сверх того, поясним сейчас. Носит это рассуждение имя Геделя, который доказал теорему о неполноте – нельзя доказать правильность некоторой теории (модели) в рамках этой же теории (модели). Модель более высокого уровня преобразует эквивалентно модель более низкого уровня из одного вида в другой. Или генерирует модель более низкого уровня по эквивалентному опять же ее описанию. А вот саму себя она преобразовать не может. Модель строит модель. И эта пирамида моделей (теорий) бесконечна.

Движемся дальше. Сложность задачи часто диктует тот способ представления модели, который будет использоваться при ее описании. Покажем это на примере простейшей задачи.

Задача 1. Пусть два объекта (например, пешеход и велосипедист) движутся друг другу навстречу (рис. 2.2) со скоростями V_1 и V_2 соответственно. Необходимо узнать: когда и где встретятся эти объекты?

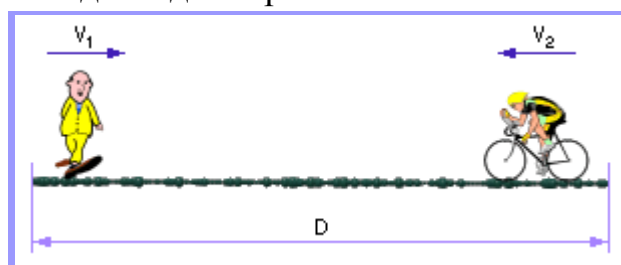


Рис. 2.2. Задача о встрече

Аналитический способ представления задачи 1

Аналитический явный способ

Эта модель весьма далека от реальности. Что-либо изучить на ней представляется проблематичным, так как из нее можно найти только время T и место встречи S . Идеализация заключается в том, что дорога считается идеально прямой, без уклонов и подъемов, скорости объектов считаются посто-

янными, желания объектов не меняются, силы безграничны, отсутствуют помехи для движения, модель не зависит от величин D , V_1 , V_2 (они могут быть сколь угодно большими или малыми).

$$\begin{aligned} T_1 &:= D / (V_1 + V_2) \\ S_1 &:= V_1 \cdot T_1 \end{aligned}$$

Реальность обычно не имеет ничего общего с такой постановкой задачи. Но за счет большой идеализации (идеализации большого порядка) получается очень простая модель, которая может быть разрешена в общем виде (аналитически) математическими способами. Так формулируются чаще всего алгоритмические модели, где протянута цепочка вычислений от исходных данных к выходу. Поэтому мы применили в записи знак присваивания ($:=$). После вычисления правой части выражения ее значение присваивается переменной, стоящей в левой части. Далее значение этой переменной применено в правой части следующего выражения. Это показано на [рис. 2.3](#).

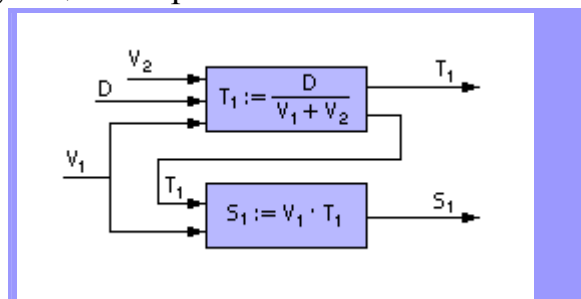


Рис. 2.3. Схема решения задачи о встрече (аналитический явный способ)

Аналитический неявный способ

В данной формулировке за счет использования знака уравнивания получена связь переменных $f(T, V_1, V_2, D, S) = 0$ в виде системы уравнений. Установив знак «?» на различные переменные, можно формулировать при необходимости целый ряд произвольных задач, например так:

$$\begin{aligned} T_1 \cdot (V_1 + V_2) &= D \\ S_1 &= V_1 \cdot T_1 \\ T_1 &= ? \end{aligned}$$

При этом задачи формулируются пользователем и не предусматриваются специально моделировщиком. То есть модель *имеет вид объекта*. Мы получили более качественную модель. Идеализация ее велика, но за счет неявной формы записи появилась возможность изменения задачи, изучения на ней целого ряда проблем.

Линейные регрессионные модели

В целях исследований часто бывает удобно представить исследуемый объект в виде ящика, имеющего входы и выходы, не рассматривая детально его внутренней структуры. Конечно, преобразования в ящике (на объекте) происходят (сигналы проходят по связям и элементам, меняют свою форму

и т. п.), но при таком представлении они происходят скрыто от наблюдателя.

По степени информированности исследователя об объекте существует деление объектов на три типа «ящиков»:

- «**белый ящик**»: об объекте известно все;
- «**серый ящик**»: известна структура объекта, неизвестны количественные значения параметров;
- «**черный ящик**»: об объекте неизвестно ничего.

Черный ящик условно изображают как на **рис. 2.4**.

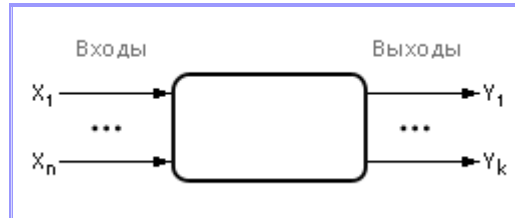


Рис. 2.4. Обозначение черного ящика на схемах

Значения на входах и выходах черного ящика можно наблюдать и измерять. Содержимое ящика неизвестно.

Задача состоит в том, чтобы, зная множество значений на входах и выходах, построить модель, то есть определить функцию ящика, по которой вход преобразуется в выход. Такая задача называется **задачей регрессионного анализа**.

В зависимости от того, доступны входы исследователю для управления или только для наблюдения, можно говорить про активный или пассивный эксперимент с ящиком.

Пусть, например, перед нами стоит задача определить, как зависит выпуск продукции от количества потребляемой электроэнергии. Результаты наблюдений отобразим на графике (см. **рис. 2.5**). Всего на графике n экспериментальных точек, которые соответствуют n наблюдениям.

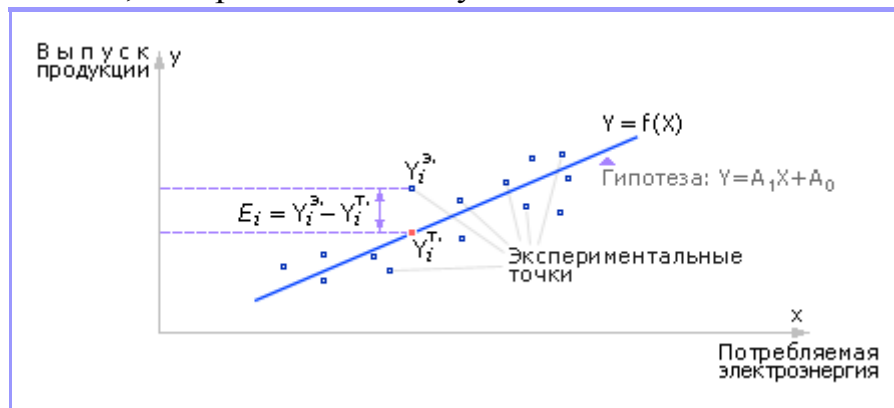


Рис. 2.5. Графический вид представления результатов наблюдения над черным ящиком

Для начала предположим, что мы имеем дело с черным ящиком, имеющим один вход и один выход. Допустим для простоты, что зависимость между входом и выходом линейная или почти линейная. Тогда данная модель будет называться **линейной одномерной регрессионной моделью**.

1) Исследователь вносит гипотезу о структуре ящика

Рассматривая экспериментально полученные данные, предположим, что они подчиняются линейной гипотезе, то есть выход Y зависит от входа X линейно, то есть гипотеза имеет вид: $Y = A_1X + A_0$ (рис. 2.5).

2) Определение неизвестных коэффициентов A_0 и A_1 модели
 Линейная одномерная модель (рис. 2.6).

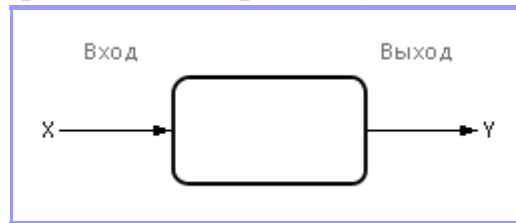


Рис. 2.6. Одномерная модель черного ящика

Для каждой из n снятых экспериментально точек вычислим ошибку (E_i) между экспериментальным значением ($Y_i^{\text{Эксп.}}$) и теоретическим значением ($Y_i^{\text{Теор.}}$), лежащим на гипотетической прямой $A_1X + A_0$ (см. рис. 2.5):

$$E_i = (Y_i^{\text{Эксп.}} - Y_i^{\text{Теор.}}), i = 1, \dots, n; \quad E_i = Y_i - A_0 - A_1 \cdot X_i, i = 1, \dots, n.$$

Ошибки E_i для всех n точек следует сложить. Чтобы положительные ошибки не компенсировали в сумме отрицательные, каждую из ошибок возводят в квадрат и складывают их значения в суммарную ошибку F уже одного знака:

$$E_i^2 = (Y_i - A_0 - A_1 \cdot X_i)^2, i = 1, \dots, n. \quad F(A_0, A_1) = \sum_{i=1}^n E_i^2$$

Цель метода – минимизация суммарной ошибки F за счет подбора коэффициентов A_0, A_1 . Другими словами, это означает, что необходимо найти такие коэффициенты A_0, A_1 линейной функции $Y = A_1X + A_0$, чтобы ее график проходил как можно ближе одновременно ко всем экспериментальным точкам. Поэтому данный метод называется **методом наименьших квадратов**.

$$F(A_0, A_1) = \sum_{i=1}^n E_i^2 = \sum_{i=1}^n (Y_i - A_0 - A_1 X_i)^2 \Rightarrow \min_{A_0, A_1}$$

Суммарная ошибка F является функцией двух переменных A_0 и A_1 , т. е. $F(A_0, A_1)$, меняя которые, можно влиять на величину ошибки (см. рис. 2.7).

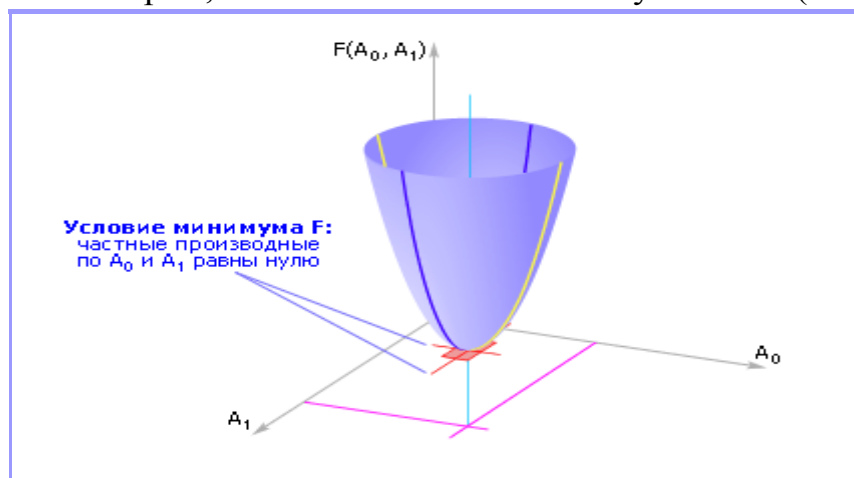


Рис. 2.7. Примерный вид функции ошибки

Чтобы суммарную ошибку минимизировать, найдем частные производные от функции F по каждой переменной и приравняем их к нулю (усло-

вие экстремума):

$$\frac{\partial F}{\partial A_0} = -2 \sum_{i=1}^n (Y_i - A_0 - A_1 X_i) = 0$$

$$\frac{\partial F}{\partial A_1} = -2 \sum_{i=1}^n (Y_i - A_0 - A_1 X_i) X_i = 0$$

Раскроем скобки и получим систему из двух линейных уравнений:

$$\sum_{i=1}^n A_0 + A_1 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i,$$

$$A_0 \sum_{i=1}^n X_i + A_1 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$

Для нахождения коэффициентов A_0 и A_1 методом Крамера представим систему в матричной форме:

$$\begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ A_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix}$$

Решение имеет вид:

$$A_0 = \frac{\sum_{i=1}^n Y_i \sum_{i=1}^n X_i^2 - \sum_{i=1}^n X_i Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

$$A_1 = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

Вычисляем значения A_0 и A_1 .

3) Проверка

Чтобы определить, принимается гипотеза или нет, нужно, во-первых, рассчитать ошибку между точками заданной экспериментальной и полученной теоретической зависимости и суммарную ошибку:

$$E_i = (Y_i^{\text{Эксп.}} - Y_i^{\text{Теор.}}), i = 1, \dots, n,$$

$$F(A_0, A_1) = \sum_{i=1}^n E_i^2$$

Далее, необходимо найти значение σ по формуле $\sigma = \sqrt{\frac{F}{n}}$, где F – сум-

марная ошибка, n – общее число экспериментальных точек.

Если в полосу, ограниченную линиями $Y^{\text{Теор.}} - S$ и $Y^{\text{Теор.}} + S$ (**рис. 2.8**), попадает 68.26% и более экспериментальных точек $Y_i^{\text{Эксп.}}$, то выдвинутая нами гипотеза принимается. В противном случае выбирают более сложную гипотезу или проверяют исходные данные. Если требуется большая уверенность в результате, то используют дополнительное условие: в полосу, ограниченную линиями $Y^{\text{Теор.}} - 2S$ и $Y^{\text{Теор.}} + 2S$, должны попасть 95.44% и более экспериментальных точек $Y_i^{\text{Эксп.}}$.

Расстояние S связано с σ следующим соотношением:

$S = \sigma / \sin(\beta) = \sigma / \sin(90^\circ - \arctg(A_1)) = \sigma / \cos(\arctg(A_1))$, что проиллюстрировано на **рис. 2.9**.

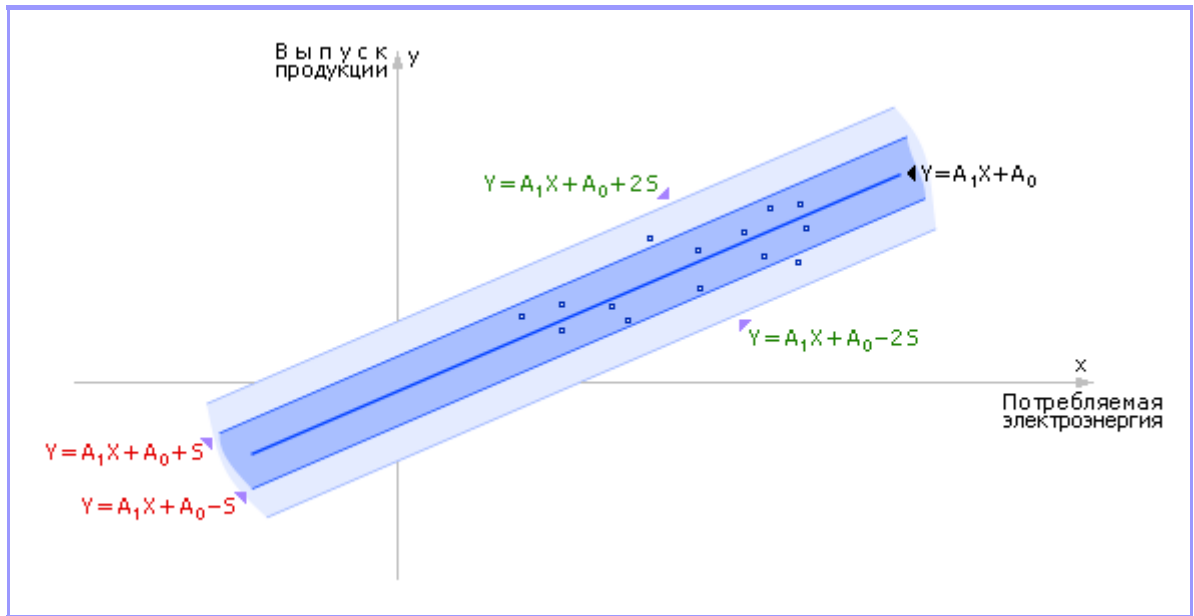


Рис. 2.8. Исследование допустимости принятия гипотезы

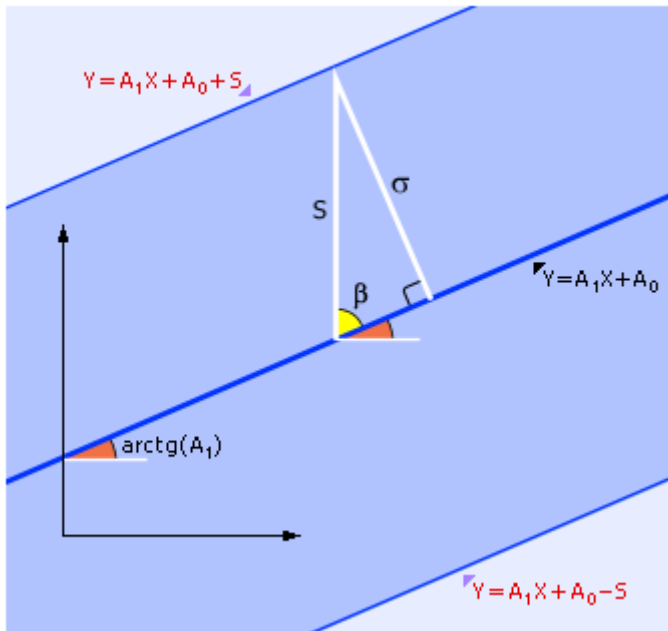


Рис. 2.9. Связь значений σ и S

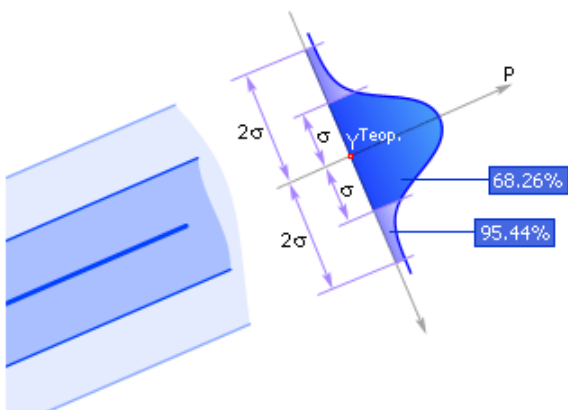


Рис. 2.10. Иллюстрация нормального закона распределения ошибок

Условие принятия гипотезы выведено из нормального закона распределения случайных ошибок (см. рис. 2.10). P – вероятность распределения нормальной ошибки. Наконец, приведем на рис. 2.11 графическую схему реализации одномерной линейной регрессионной модели.

Линейная множественная модель

Предположим, что функциональная структура ящика снова имеет линейную зависимость, но количество входных сигналов, действующих одновременно на объект, равно m (см. рис. 2.11):

$$Y = A_0 + A_1 \cdot X_1 + \dots + A_m \cdot X_m.$$

Так как подразумевается, что мы имеем экспериментальные данные о всех входах и выходах черного ящика, то можно вычислить ошибку между экспериментальным ($Y_i^{\text{Эксп.}}$) и теоретическим ($Y_i^{\text{Теор.}}$) значением Y для каждой i -ой точки (пусть, как и прежде,

число экспериментальных точек равно n):

$$E_i = (Y_i^{\text{Эксп.}} - Y_i^{\text{Теор.}}), i = 1, \dots, n;$$

$$E_i = Y_i - A_0 - A_1 \cdot X_{1i} - \dots - A_m \cdot X_{mi}, i = 1, \dots, n.$$

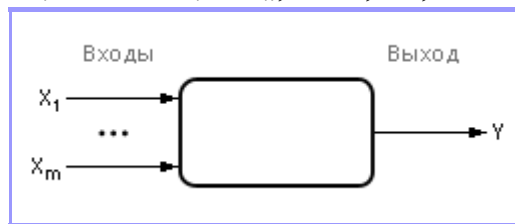


Рис. 2.11. Обозначение многомерного черного ящика на схемах

Минимизируем суммарную ошибку F :

$$F(A_0, A_1, \dots, A_m) = \sum_{i=1}^n E_i^2 \Rightarrow \min_{A_0, A_1, \dots, A_m}$$

Она зависит от параметров A_0, A_1, \dots, A_m . Для нахождения экстремума приравняем частные производные F по неизвестным A_0, A_1, \dots, A_m к нулю:

$$\frac{\partial F}{\partial A_j} = 0, j = \overline{0, m}$$

Получим систему из $m + 1$ уравнения с $m + 1$ неизвестными, которую следует решить, чтобы определить коэффициенты линейной множественной модели A_0, A_1, \dots, A_m . Для нахождения коэффициентов методом Крамера представим систему в матричном виде:

$$\begin{pmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} & \dots & \sum_{i=1}^n X_{mi} \\ \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{1i}X_{1i} & \sum_{i=1}^n X_{2i}X_{1i} & \dots & \sum_{i=1}^n X_{mi}X_{1i} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{1i}X_{2i} & \sum_{i=1}^n X_{2i}X_{2i} & \dots & \sum_{i=1}^n X_{mi}X_{2i} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n X_{mi} & \sum_{i=1}^n X_{1i}X_{mi} & \sum_{i=1}^n X_{2i}X_{mi} & \dots & \sum_{i=1}^n X_{mi}X_{mi} \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ \dots \\ A_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_i X_{1i} \\ \sum_{i=1}^n Y_i X_{2i} \\ \dots \\ \sum_{i=1}^n Y_i X_{mi} \end{pmatrix}$$

Вычисляем коэффициенты A_0, A_1, \dots, A_m .

Далее, по аналогии с одномерной моделью (см. «Проверка»), для каждой точки вычисляется ошибка E_i ; затем находится суммарная ошибка F и значения σ и S с целью определить, принимается ли выдвинутая гипотеза о линейности многомерного черного ящика или нет.

Нелинейные регрессионные модели

Полиномиальная множественная регрессионная модель

Если черный ящик (рис. 2.12) имеет, например, два входа, а зависимость выхода от входов напоминает квадратичную, то целесообразно вы-

брать такую гипотезу:

$$Y = A_0 + A_1 \cdot X_1 + A_2 \cdot X_2 + A_3 \cdot X_1 \cdot X_2 + A_4 \cdot X_1 \cdot X_1 + A_5 \cdot X_2 \cdot X_2.$$

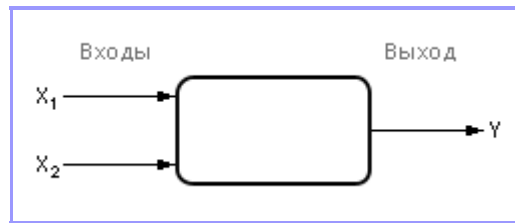


Рис. 2.12. Обозначение двумерной модели черного ящика на схемах

Обозначим: $Z_1 = X_1 \cdot X_2$; $Z_2 = X_1 \cdot X_1$; $Z_3 = X_2 \cdot X_2$ и подставим эти выражения в предыдущую формулу:

$$Y = A_0 + A_1 \cdot X_1 + A_2 \cdot X_2 + A_3 \cdot Z_1 + A_4 \cdot Z_2 + A_5 \cdot Z_3.$$

Таким образом, данная задача сведена к линейной множественной модели. А модель черного ящика выглядит теперь, как показано на **рис. 2.13**.

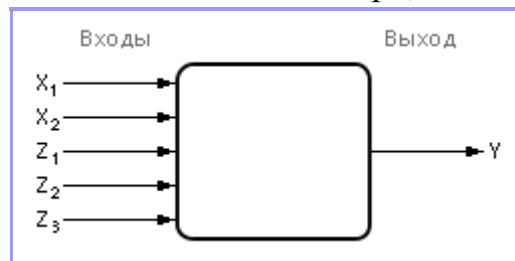


Рис. 2.13. Преобразованная модель черного ящика

Мультипликативная регрессионная модель (рис. 2.14).

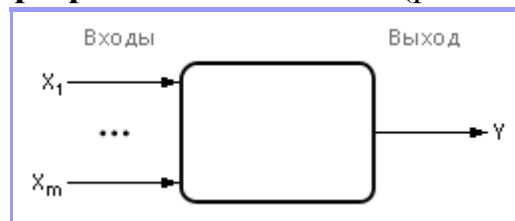


Рис. 2.14. Обозначение модели многомерного черного ящика на схемах

$$Y = A_0 \cdot X_1^{A_1} \cdot X_2^{A_2} \cdot \dots \cdot X_m^{A_m}.$$

Прологарифмируем левую и правую части данного уравнения:

$$\ln(Y) = \ln(A_0) + A_1 \cdot \ln(X_1) + A_2 \cdot \ln(X_2) + \dots + A_m \cdot \ln(X_m).$$

Обозначим: $W = \ln(Y)$, $B_0 = \ln(A_0)$, $Z_1 = \ln(X_1)$, $Z_2 = \ln(X_2)$, ..., $Z_m = \ln(X_m)$.

Получим: $W = B_0 + A_1 \cdot Z_1 + A_2 \cdot Z_2 + \dots + A_m \cdot Z_m$.

Т. е. вновь осуществлен переход к линейной множественной модели.

Обратная регрессионная модель (рис. 2.15).

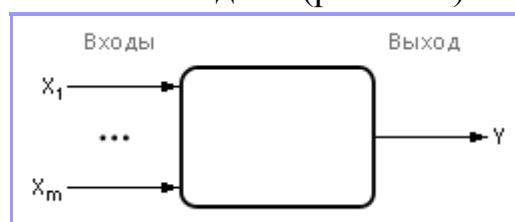


Рис. 2.15. Обозначение модели многомерного черного ящика на схемах

$$Y = k / (A_0 + A_1 X_1 + \dots + A_m X_m).$$

Заменим: $W = 1/Y$, $a_i = A_i/k$, получим линейную множественную модель:

$$W = a_0 + a_1 \cdot X_1 + \dots + a_m \cdot X_m.$$

Экспоненциальная модель

$$Y = e^{B_0 + B_1 X_1 + B_2 X_2 + \dots + B_m X_m}.$$

Прологарифмируем левую и правую части уравнения:

$$\ln(Y) = B_0 + B_1 \cdot X_1 + B_2 \cdot X_2 + \dots + B_m \cdot X_m.$$

Выполним замену $W = \ln(Y)$ и получим:

$$W = B_0 + B_1 \cdot X_1 + B_2 \cdot X_2 + \dots + B_m \cdot X_m.$$

Далее пользуемся выражением для линейной множественной модели.

Моделирование систем массового обслуживания

Большой класс систем, которые сложно изучить аналитическими способами, но которые хорошо изучаются методами статистического моделирования, сводится к системам массового обслуживания (СМО).

В СМО подразумевается, что есть *типовые пути* (каналы обслуживания), через которые в процессе обработки проходят *заявки*. Принято говорить, что заявки *обслуживаются* каналами. Каналы могут быть разными по назначению, характеристикам, они могут сочетаться в разных комбинациях; заявки могут находиться в очередях и ожидать обслуживания. Часть заявок может быть обслужена каналами, а части могут отказать в этом. Важно, что заявки, с точки зрения системы, абстрактны: это то, что желает обслуживаться, то есть пройти определенный путь в системе. Каналы являются также абстракцией: это то, что обслуживает заявки.

Заявки могут приходиться неравномерно, каналы могут обслуживать разные заявки за разное время и так далее, количество заявок всегда весьма велико. Все это делает такие системы сложными для изучения и управления, и проследить все причинно-следственные связи в них не представляется возможным. Поэтому принято представление о том, что обслуживание в сложных системах носит случайный характер. Примеры СМО сведены в табл. 2.5.

Таблица 2.5

Примеры систем массового обслуживания

СМО	Заявки	Каналы
Автобусный маршрут и перевозка пассажиров	Пассажиры	Автобусы
Производственный конвейер	Детали	Станки, склады
Эскадрилья самолетов, которая «обслуживается» зенитками ПВО	Самолеты	Зенитные орудия
Ствол и рожок автомата, которые «обслуживают» патроны	Патроны	Ствол, рожок
Электрические заряды, перемещающиеся в некотором устройстве	Заряды	Каскады устройства

Но все эти системы объединены в один класс СМО, поскольку **подход** к их изучению един. Он состоит в том, что, **во-первых**, с помощью генератора случайных чисел разыгрываются случайные числа, которые имитируют СЛУЧАЙНЫЕ моменты появления заявок и время их обслуживания в каналах. Но в совокупности эти случайные числа, конечно, подчинены *статистическим* закономерностям.

К примеру, пусть сказано: «заявки в среднем приходят в количестве 5 штук в час». Это означает, что времена между приходом двух соседних заявок случайны, например: 0.1; 0.3; 0.1; 0.4; 0.2, как это показано на **рис. 2.16**, но в сумме они дают в среднем 1 (обратите внимание, что в примере это не точно 1, а 1.1 – но зато в другой час эта сумма, например, может быть равной 0.9); и только *за достаточно большое время* среднее этих чисел станет близким к одному часу.

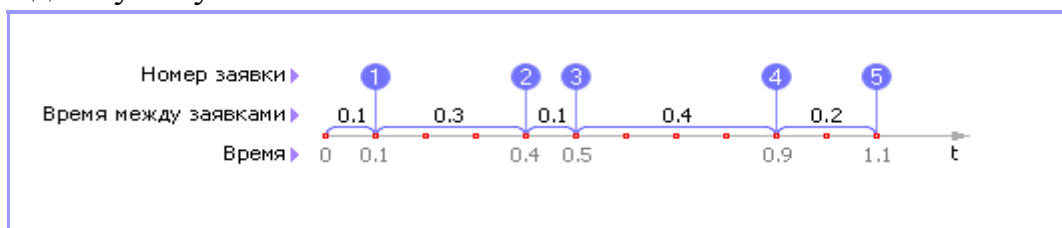


Рис. 2.16. Случайный процесс прихода заявок в СМО

Результат (например, пропускная способность системы), конечно, тоже будет случайной величиной на отдельных промежутках времени. Но измеренная на большом промежутке времени, эта величина будет уже, в среднем, соответствовать точному решению. То есть для характеристики СМО интересуются ответами в статистическом смысле.

Итак, систему испытывают случайными входными сигналами, подчиненными заданному статистическому закону, а в качестве результата принимают статистические показатели, усредненные по времени рассмотрения или по количеству опытов (см. **рис. 2.17**).

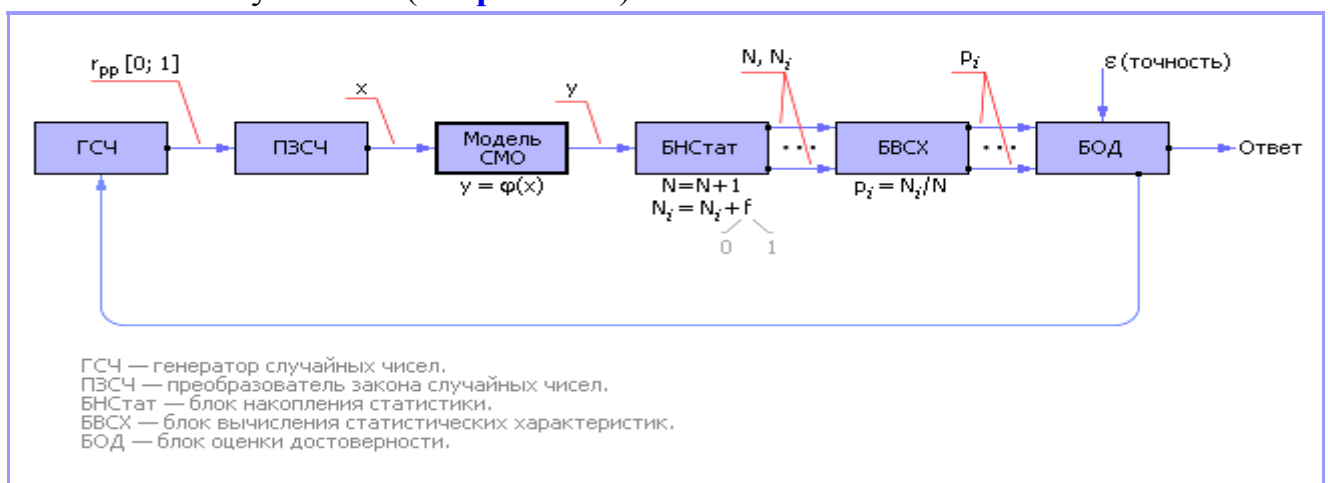


Рис. 2.17. Схема статистического эксперимента для изучения систем массового обслуживания

Во-вторых, все модели СМО собираются типовым образом из небольшого набора элементов (канал, источник заявок, очередь, заявка, дисциплина

обслуживания, стек, кольцо и так далее), что позволяет имитировать эти задачи *типовым* образом. Для этого модель системы собирают из конструктора таких элементов. Неважно, какая конкретно система изучается, важно, что схема системы собирается из одних и тех же элементов. Разумеется, структура схемы будет всегда различной.

Перечислим некоторые основные понятия СМО.

Каналы – то, что обслуживает; бывают **горячие** (начинают обслуживать заявку в момент ее поступления в канал) и **холодные** (каналу для начала обслуживания требуется время на подготовку). **Источники заявок** – порождают заявки в случайные моменты времени, согласно заданному пользователем статистическому закону. **Заявки**, они же **клиенты**, входят в систему (порождаются источниками заявок), проходят через ее элементы (обслуживаются), покидают ее обслуженными или неудовлетворенными. Бывают **нетерпеливые заявки** – такие, которым надоело ожидать или находиться в системе и которые покидают по собственной воле СМО. Заявки образуют **потоки** – поток заявок **на входе системы**, поток **обслуженных** заявок, поток **отказанных** заявок. Поток характеризуется количеством заявок определенного сорта, наблюдаемым в некотором месте СМО за единицу времени (час, сутки, месяц), то есть поток есть величина статистическая.

Очереди характеризуются правилами стояния в очереди (дисциплиной обслуживания), количеством мест в очереди (сколько клиентов максимум может находиться в очереди), структурой очереди (связь между местами в очереди). Бывают **ограниченные** и **неограниченные** очереди. Перечислим важнейшие дисциплины обслуживания. **FIFO** (First In, First Out – первым пришел, первым ушел): если заявка первой пришла в очередь, то она первой уйдет на обслуживание. **LIFO** (Last In, First Out – последним пришел, первым ушел): если заявка последней пришла в очередь, то она первой уйдет на обслуживание (пример – патроны в рожке автомата). **SF** (Short Forward – короткие вперед): в первую очередь обслуживаются те заявки из очереди, которые имеют меньшее время обслуживания.

Дадим яркий пример, показывающий, как правильный выбор той или иной дисциплины обслуживания позволяет получить ощутимую экономию по времени. Пусть имеется два магазина. В магазине № 1 обслуживание осуществляется в порядке очереди, то есть здесь реализована дисциплина обслуживания FIFO (см. **рис. 2.18**).

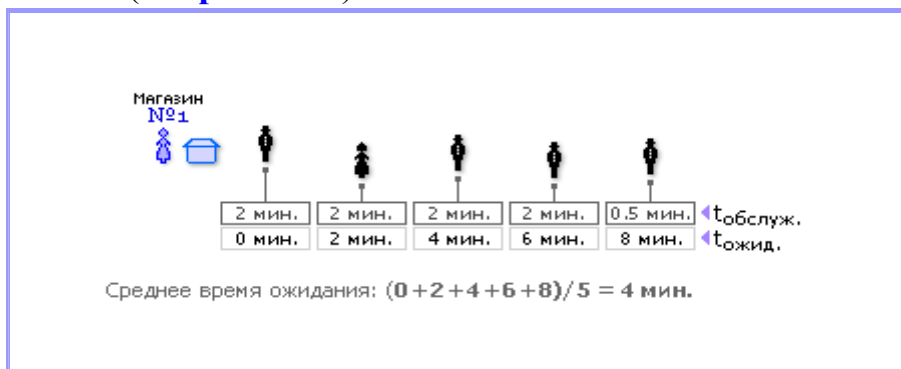


Рис. 2.18. Организация очереди по дисциплине FIFO

Время обслуживания $t_{\text{обслуж.}}$ на рис. 2.18 показывает, сколько продавец затратит на одного покупателя. Понятно, что при покупке штучного товара продавец затратит меньше времени на обслуживание, чем при покупке, скажем, сыпучих продуктов, требующих дополнительных манипуляций (набрать, взвесить, высчитать цену и т. п.). Время ожидания $t_{\text{ожид.}}$ показывает, через какое время очередной покупатель будет обслужен продавцом.

В магазине № 2 реализована дисциплина SF (см. рис. 2.19), означающая, что штучный товар можно купить вне очереди, так как время обслуживания $t_{\text{обслуж.}}$ такой покупки невелико.

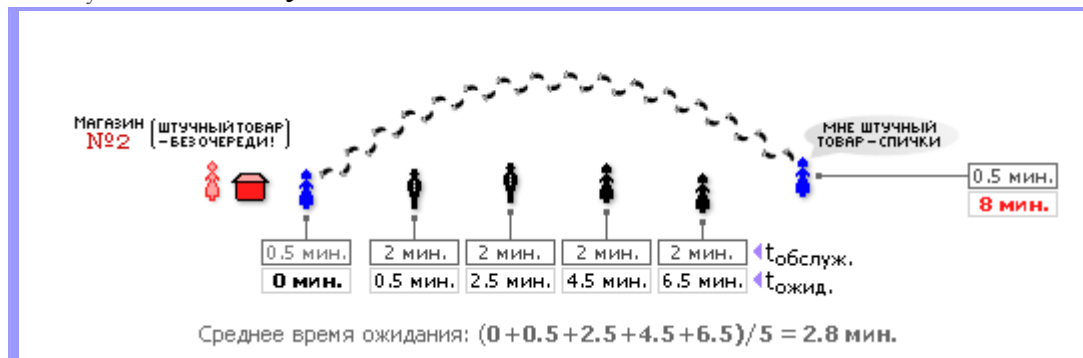


Рис. 2.19. Организация очереди по дисциплине SF

Как видно из обоих рисунков, последний (пятый) покупатель собирается приобрести штучный товар, поэтому время его обслуживания невелико – 0.5 минут. Если этот покупатель придет в магазин № 1, он будет вынужден выстоять в очереди целых 8 минут, в то время как в магазине № 2 его обслужат сразу же, вне очереди. Таким образом, среднее время обслуживания каждого из покупателей в магазине с дисциплиной обслуживания FIFO составит 4 минуты, а в магазине с дисциплиной обслуживания КВ – лишь 2.8 минуты. А экономия времени составит: $(1 - 2.8 / 4) \cdot 100\% = 30$ процентов! Итак, 30% сэкономленного для общества времени – и это лишь за счет правильного выбора дисциплины обслуживания.

Специалист по системам должен хорошо понимать ресурсы производительности и эффективности проектируемых им систем, скрытые в оптимизации параметров, структур и дисциплинах обслуживания. Моделирование помогает выявить эти скрытые резервы.

При анализе результатов моделирования важно также указать интересы и степень их выполнения. Различают интересы клиента и интересы владельца системы. Заметим, что эти интересы совпадают не всегда. Судить о результатах работы СМО можно по показателям. Наиболее популярные из них:

- вероятность обслуживания клиента системой;
- пропускная способность системы;
- вероятность отказа клиенту в обслуживании;
- вероятность занятости каждого из канала и всех вместе;
- среднее время занятости каждого канала;
- вероятность занятости всех каналов;

- среднее количество занятых каналов;
- вероятность простоя каждого канала;
- вероятность простоя всей системы;
- среднее количество заявок, стоящих в очереди;
- среднее время ожидания заявки в очереди;
- среднее время обслуживания заявки;
- среднее время нахождения заявки в системе.

Судить о качестве полученной системы нужно по совокупности значений показателей. При анализе результатов моделирования (показателей) важно также обращать внимание на интересы клиента и интересы владельца системы, то есть минимизировать или максимизировать надо тот или иной показатель, а также на степень их выполнения. Заметим, что чаще всего интересы клиента и владельца между собой не совпадают или совпадают не всегда. Показатели будем обозначать далее $H = \{h_1, h_2, \dots\}$.

Параметрами СМО могут быть: интенсивность потока заявок, интенсивность потока обслуживания, среднее время, в течение которого заявка готова ожидать обслуживания в очереди, количество каналов обслуживания, дисциплина обслуживания и так далее. Параметры – это то, что влияет на показатели системы. Параметры будем обозначать далее как $R = \{r_1, r_2, \dots\}$.

Пример. Автозаправочная станция (АЗС).

1. Постановка задачи. На рис. 2.20 приведен план АЗС. Рассмотрим метод моделирования СМО на ее примере и план ее исследования. Водители, проезжая по дороге мимо АЗС по дороге, могут захотеть заправить свой автомобиль. Хотят обслужиться (заправить машину бензином) не все автомобилисты подряд; допустим, что из всего потока машин на заправку в среднем заезжает 5 машин в час.



Рис. 2.20. План моделируемой АЗС

На АЗС две одинаковые колонки, статистическая производительность каждой из которых известна. Первая колонка в среднем обслуживает 1 машину в час, вторая в среднем – 3 машины в час. Владелец АЗС заасфальтиро-

вал для машин место, где они могут ожидать обслуживания. Если колонки заняты, то на этом месте могут ожидать обслуживания другие машины, но не более двух одновременно. Очередь будем считать общей. Как только одна из колонок освободится, то первая машина из очереди может занять ее место на колонке (при этом вторая машина продвигается на первое место в очереди). Если появляется третья машина, а все места (их два) в очереди заняты, то ей отказывают в обслуживании, так как стоять на дороге запрещено (см. дорожные знаки около АЗС). Такая машина уезжает прочь из системы навсегда и как потенциальный клиент является потерянной для владельца АЗС. Можно усложнить задачу, рассмотрев кассу (еще один канал обслуживания, куда надо попасть после обслуживания в одной из колонок) и очередь к ней и так далее. Но в простейшем варианте очевидно, что пути движения потоков заявок по СМО можно изобразить в виде эквивалентной схемы, а добавив значения и обозначения характеристик каждого элемента СМО, получаем окончательно схему, изображенную на **рис. 2.21**.

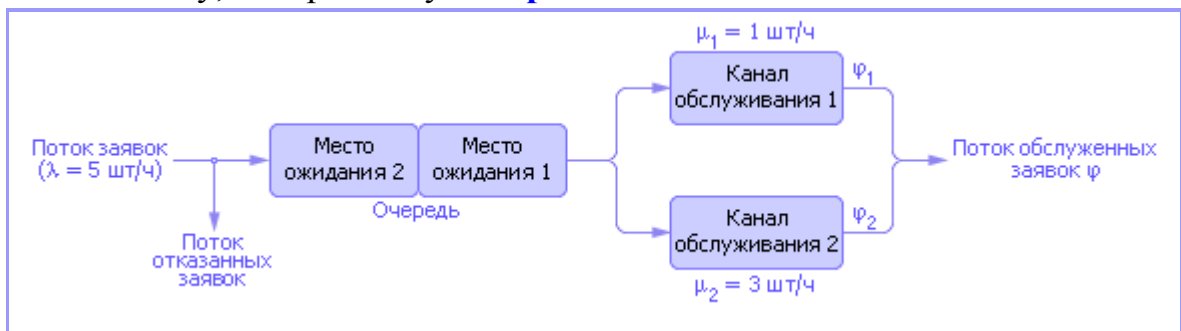


Рис. 2.21. Эквивалентная схема объекта моделирования

2. Метод исследования СМО. Применим в нашем примере принцип **последовательной проводки**. Его идея заключается в том, что заявку проводят через всю систему от входа до выхода, и только после этого берутся за моделирование следующей заявки.

Для наглядности построим временную диаграмму работы СМО, отражая на каждой линейке (ось времени t) состояние отдельного элемента системы. Временных линеек проводится столько, сколько имеется различных мест в СМО, потоков. В нашем примере их 7 (поток заявок, поток ожидания на первом месте в очереди, поток ожидания на втором месте в очереди, поток обслуживания в канале 1, поток обслуживания в канале 2, поток обслуженных системой заявок, поток отказанных заявок).

Для генерации времени прихода заявок используем формулу вычисления интервала между моментами прихода двух случайных событий:

$$t = -\frac{1}{\lambda} \cdot \ln(r)$$

В этой формуле величина потока λ должна быть задана (до этого она должна быть определена экспериментально на объекте как статистическое среднее), r – случайное равномерно распределенное число от 0 до 1 из ГСЧ или [таблицы](#), в которой случайные числа нужно брать подряд.

Задача. Сгенерируйте поток из 10 случайных событий с интенсивностью появления событий 5 шт/час.

Решение задачи. Возьмем случайные числа, равномерно распределенные в интервале от 0 до 1, и их натуральные логарифмы (см. табл. 2.6).

Формула пуассоновского потока определяет **расстояние между двумя случайными событиями** следующим образом: $t = -\text{Ln}(r_{pp})/\lambda$. Тогда, учитывая, что $\lambda = 5$, имеем расстояния между двумя случайными соседними событиями: 0.68, 0.21, 0.31, 0.12 часа. То есть события наступают: первое – в момент времени $t = 0$, второе – в момент времени $t = 0.68$, третье – в момент времени $t = 0.89$, четвертое – в момент времени $t = 1.20$, пятое – в момент времени $t = 1.32$ и так далее. События – приход заявок отразим на первой линейке (см. **рис. 2.22**).

Таблица 2.6.
Фрагмент таблицы случайных чисел и их логарифмов

$r_{pp}[0; 1]$	$\ln(r_{pp}[0; 1])$
0.0333	-3.4022
0.3557	-1.0337
0.2172	-1.5269
0.5370	-0.6218

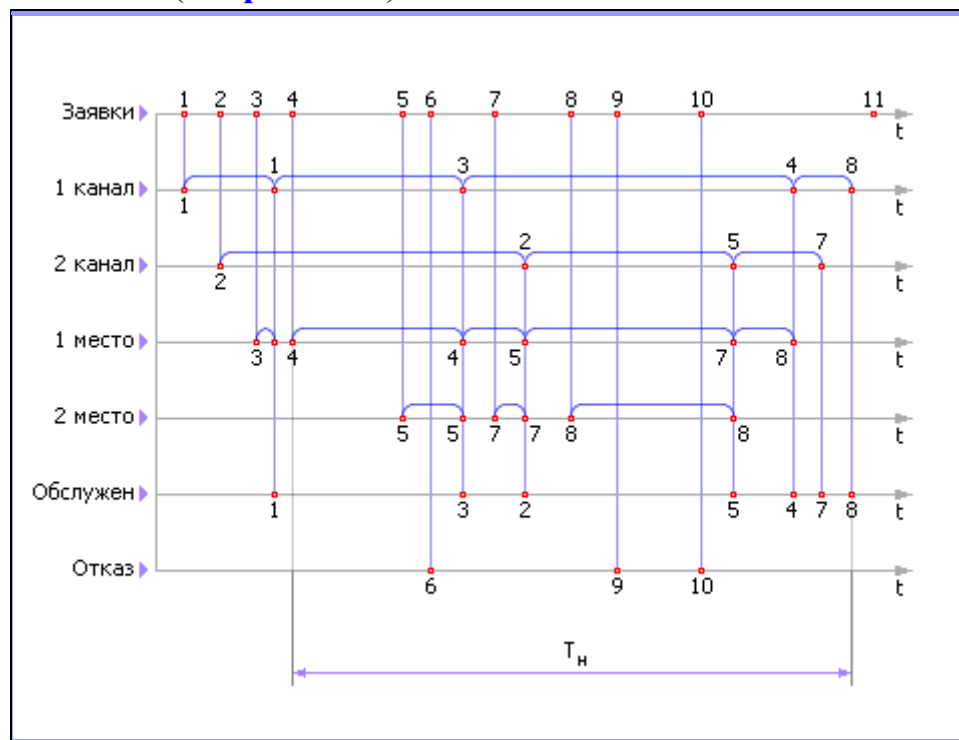


Рис. 2.22. Временная диаграмма работы СМО

Берется первая заявка и, так как в этот момент каналы свободны, устанавливается на обслуживание в первый канал. Заявка 1 переносится на линейку «1 канал». Время обслуживания в канале тоже случайное и вычисляется по аналогичной формуле:

$$t = -\frac{1}{\lambda} \cdot \ln(r)$$

где роль интенсивности играет величина потока обслуживания μ_1 или μ_2 , в зависимости от того, какой канал обслуживает заявку. Находим на диаграмме момент окончания обслуживания, откладывая сгенерированное время обслуживания от момента начала обслуживания, и опускаем заявку на линейку «Обслуженные».

Заявка прошла в СМО весь путь. Теперь можно, согласно принципу последовательной проводки заявок, также проимитировать путь второй заявки.

Если в некоторый момент окажется, что оба канала заняты, то следует установить заявку в очередь. На **рис. 2.22** это заявка с номером 3. Заметим, что по условиям задачи в очереди в отличие от каналов заявки находятся не случайное время, а ожидают, когда освободится какой-то из каналов. После освобождения канала заявка поднимается на линейку соответствующего канала и там организуется ее обслуживание.

Если все места в очереди в момент, когда придет очередная заявка, будут заняты, то заявку следует отправить на линейку «Отказанные». На **рис. 2.22** это заявка с номером 6.

Процедуру имитации обслуживания заявок продолжают некоторое время наблюдения T_n . Чем больше это время, тем точнее в дальнейшем будут результаты моделирования. Реально для простых систем выбирают T_n , равное 50–100 и более часов, хотя иногда лучше мерить эту величину количеством рассмотренных заявок.

Анализ временной диаграммы

Анализ проведем на уже рассмотренном примере.

Сначала нужно дождаться установившегося режима. Откидываем первые четыре заявки как нехарактерные, протекающие во время процесса установления работы системы. Измеряем время наблюдения, допустим, что в нашем примере оно составит $T_n = 5$ часов. Подсчитываем из диаграммы количество обслуженных заявок $N_{\text{обс.}}$, времена простоя и другие величины. В результате можем вычислить показатели, характеризующие качество работы СМО.

1. Вероятность обслуживания: $P_{\text{обс.}} = N_{\text{обс.}}/N = 5/7 = 0.714$. Для расчета вероятности обслуживания заявки в системе достаточно разделить число заявок, которым удалось обслужиться за время T_n (см. линейку «Обслуженные») $N_{\text{обс.}}$, на число заявок N , которые хотели обслужиться за это же время. Как и раньше вероятность экспериментально определяем отношением свершившихся событий к общему числу событий, которые могли совершиться!

2. Пропускная способность системы: $A = N_{\text{обс.}}/T_n = 5/5 = 1.0$ [шт/час]. Для расчета пропускной способности системы достаточно разделить число обслуженных заявок $N_{\text{обс.}}$ на время T_n , за которое произошло это обслуживание (см. линейку «Обслуженные»).

3. Вероятность отказа: $P_{\text{отк.}} = N_{\text{отк.}}/N = 2/7 = 0.286$. Для расчета вероятности отказа заявке в обслуживании достаточно разделить число заявок $N_{\text{отк.}}$, которым отказали за время T_n (см. линейку «Отказанные»), на число заявок N , которые хотели обслужиться за это же время, то есть поступили в систему.

Обратите внимание. $P_{\text{отк.}} + P_{\text{обс.}}$ в теории должно быть равно 1. На самом деле экспериментально получилось, что $P_{\text{отк.}} + P_{\text{обс.}} = 0.714 + 0.286 = 1.0$. Эта неточность объясняется тем, что время наблюдения T_n мало и статистика накоплена недостаточная для получения точного ответа. Погрешность это показателя сейчас составляет 14%!

4. Вероятность занятости одного канала: $P_1 = T_{\text{зан.}}/T_{\text{н}} = 0.05/5 = 0.01$, где $T_{\text{зан.}}$ — время занятости только одного канала (первого или второго). Измерениям подлежат временные отрезки, на которых происходят определенные события. Например, на диаграмме ищутся такие отрезки, во время которых заняты или первый или второй канал. В данном примере есть один такой отрезок в конце диаграммы длиной 0.05 часа. Доля этого отрезка в общем времени рассмотрения ($T_{\text{н}} = 5$ часов) определяется делением и составляет искомую вероятность занятости.

5. Вероятность занятости двух каналов: $P_2 = T_{\text{зан.}}/T_{\text{н}} = 4.95/5 = 0.99$. На диаграмме ищутся такие отрезки, во время которых одновременно заняты и первый, и второй канал. В данном примере таких отрезков четыре, их сумма равна 4.95 часа. Доля продолжительности этих события в общем времени рассмотрения ($T_{\text{н}} = 5$ часов) определяется делением и составляет искомую вероятность занятости.

6. Среднее количество занятых каналов: $N_{\text{ск}} = 0 \cdot P_0 + 1 \cdot P_1 + 2 \cdot P_2 = 0.01 + 2 \cdot 0.99 = 1.99$. Чтобы подсчитать, сколько каналов занято в системе в среднем, достаточно знать долю (вероятность занятости одного канала) и умножить на вес этой доли (один канал), знать долю (вероятность занятости двух каналов) и умножить на вес этой доли (два канала) и так далее. Полученная цифра 1.99 говорит о том, что из возможных двух каналов в среднем загружено 1.99 канала. Это высокий показатель загрузки, 99.5%, система хорошо использует ресурс.

7. Вероятность простоя хотя бы одного канала: $P^*_1 = T_{\text{простоя1}}/T_{\text{н}} = 0.05/5 = 0.01$.

8. Вероятность простоя двух каналов одновременно:

$$P^*_2 = T_{\text{простоя2}}/T_{\text{н}} = 0.$$

9. Вероятность простоя всей системы: $P^*_c = T_{\text{простоя сист.}}/T_{\text{н}} = 0$.

10. Среднее количество заявок в очереди:

$N_{\text{сз}} = 0 \cdot P_{03} + 1 \cdot P_{13} + 2 \cdot P_{23} = 0.34 + 2 \cdot 0.64 = 1.62$ [шт]. Чтобы определить среднее количество заявок в очереди, надо определить отдельно вероятность того, что в очереди будет одна заявка P_{13} , вероятность того, в очереди будет стоять две заявки P_{23} и т. д. и снова с соответствующими весами их сложить.

11. Вероятность того, что в очереди будет одна заявка: $P_{13} = T_{13}/T_{\text{н}} = 1.7/5 = 0.34$ (всего на диаграмме четырех таких отрезка, в сумме дающих 1.7 часа).

12. Вероятность того, в очереди будет стоять одновременно две заявки: $P_{23} = T_{23}/T_{\text{н}} = 3.2/5 = 0.64$ (всего на диаграмме три таких отрезка, в сумме дающих 3.25 часа).

13. Среднее время ожидания заявки в очереди:

$$T_{\text{ож}} = \frac{\sum_{i=1}^N T_{\text{ож}i}}{N}$$

(Сложить все временные интервалы, в течение которых какая-либо заявка находилась в очереди, и разделить на количество заявок). На временной диа-

грамме таких заявок 4.

14. Среднее время обслуживания заявки:

$$T_{\text{ср.обсл.}} = \frac{\sum_{i=1}^N T_{\text{обсл.}i}}{N}$$

(Сложить все временные интервалы, в течение которых какая-либо заявка находилась на обслуживании в каком-либо канале, и разделить на количество заявок).

15. Среднее время нахождения заявки в системе: $T_{\text{ср. сист.}} = T_{\text{ср. ож.}} + T_{\text{ср. обсл.}}$

16. Среднее количество заявок в системе:

$$N_{\text{ср.}} = \frac{\sum_{i=1}^K (N_{2i} + N_{3i} + N_{4i} + N_{5i})}{K}$$

Разобьем интервал наблюдения, например, на десятиминутки. Получится на пяти часах K подынтервалов (в нашем случае $K = 30$). В каждом подынтервале определим по временной диаграмме, сколько заявок в этот момент находится в системе. Смотреть надо на 2, 3, 4 и 5-ю линейки – какие из них заняты в данный момент. Затем сумму K слагаемых усреднить.

Далее следует оценить точность каждого из полученных результатов. То есть ответить на вопрос: насколько мы можем доверять этим значениям?

Если точность не является удовлетворительной, то следует увеличить время эксперимента и тем самым улучшить статистику. Можно сделать и по-другому. Снова несколько раз запустить эксперимент на время $T_{\text{н.}}$. А впоследствии усреднить значения этих экспериментов. И снова проверить результаты на критерий точности. Эту процедуру следует повторять до тех пор, пока не будет достигнута требуемая точность.

Далее следует составить таблицу результатов и оценить значения каждого из них с точки зрения клиента и владельца СМО. Таблица должна иметь примерно такой вид, какой показан в табл. 2.7.

Синтез СМО

Мы проделали анализ существующей системы. Это дало возможность увидеть ее недостатки и определить направления улучшения ее качества. Но остаются непонятными ответы на конкретные вопросы, что именно надо сделать – увеличивать количество каналов или увеличивать их пропускную способность, или увеличивать количество мест в очереди, и, если увеличивать, то насколько? Есть и такие вопросы, что лучше – создать 3 канала с производительностью 5 шт/час или один с производительностью 15 шт/час?

Чтобы оценить чувствительность каждого показателя к изменению значения определенного параметра, поступают следующим образом. Фиксируют все параметры кроме одного, выбранного. Затем снимают значение всех показателей при нескольких значениях этого выбранного параметра. Конечно, приходится повторять снова и снова процедуру имитации и усреднять показатели при каждом значении параметра, оценивать точность. Но в

результате получаются надежные статистические зависимости характеристик (показателей) от параметра.

Таблица 2.7. Показатели СМО

Показатель	Формула	Значение	Интересы владельца СМО	Интересы клиента СМО
Вероятность обслуживания	$P_{\text{обс.}} = N_{\text{обс.}}/N$	0.714	Вероятность обслуживания мала, многие клиенты уходят из системы неудовлетворенными, их деньги для владельца потеряны. Это «минус». Рекомендация: увеличить вероятность обслуживания.	Вероятность обслуживания мала, каждый третий клиент не может обслужиться. Это «минус». Рекомендация: увеличить вероятность обслуживания.
Среднее количество заявок в очереди	$N_{\text{сз}} = 0 \cdot P_{03} + 1 \cdot P_{13} + 2 \cdot P_{23}$	1.62	Все места в очереди используются достаточно эффективно. Вложения на организацию очереди окупают затраты. Это «плюс». Клиенты, которые долго стоят в очереди, могут уйти, не дождавшись обслуживания. Клиенты, простаивая, могут нанести ущерб системе. Много отказов, потерянных клиентов. Это «минусы». Рекомендация: увеличить пропускную	Очередь практически все время вся забита. Клиенту приходится стоять в очереди, прежде чем он попадет на обслуживание. Клиент может не попасть даже в очередь. Это «минус». Рекомендация: увеличить пропускную способность, увеличить число мест в очереди.

Общий итог:	В интересах владельца: а) увеличить пропускную; б) увеличить число	Клиенты заинтересованы в увеличении пропускной способно-
-------------	--	--

Например, для 12 показателей нашего примера можно получить 12 зависимостей от одного параметра: зависимость вероятности отказов $P_{отк.}$ от количества мест в очереди (КМО), зависимость пропускной способности A от количества мест в очереди, и так далее (см. [рис. 2.23](#)).

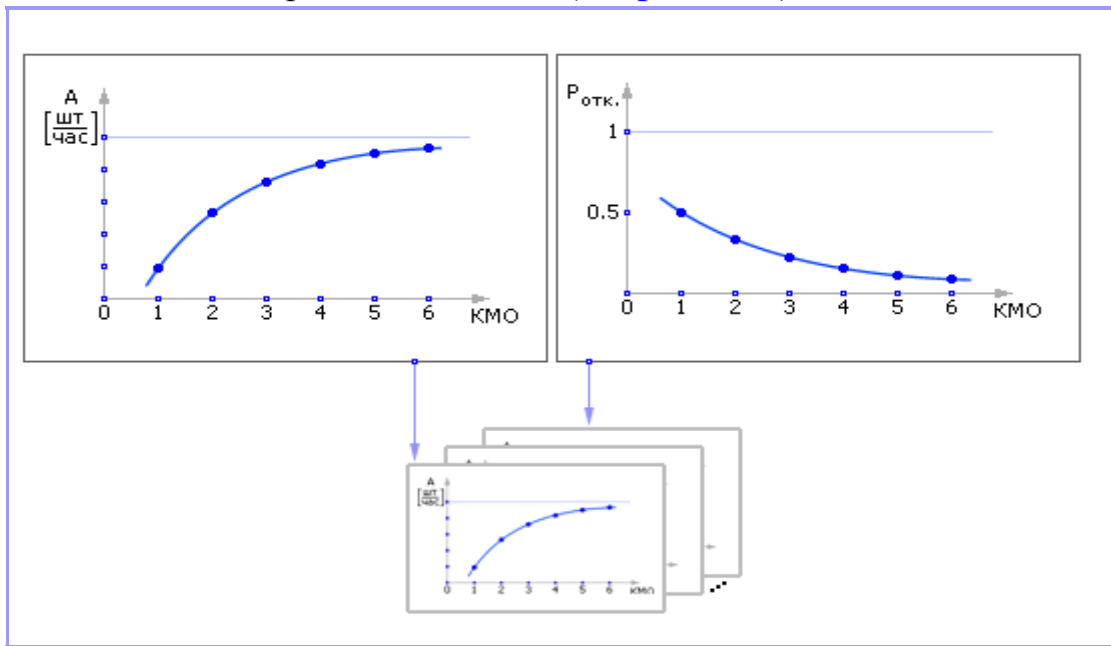


Рис. 2.23. Зависимости показателей от параметров СМО

Затем так же можно снять еще 12 зависимостей показателей P от другого параметра R , зафиксировав остальные параметры. И так далее. Образуется своеобразная матрица зависимостей показателей P от параметров R , по которой можно провести дополнительный анализ о перспективах движения (улучшения показателей) в ту или иную сторону. Наклон кривых хорошо показывает чувствительность, эффект от движения по определенному показателю. В математике эту матрицу называют **якобианом J** , в которой роль наклона кривых играют значения производных $\Delta P_i / \Delta R_j$, см. [рис. 2.24](#).



Рис. 2.24. Якобиан – матрица чувствительностей показателей в зависимости от изменения параметров СМО

Напомним, что производная связана геометрически с углом наклона касательной к зависимости. Если показателей 12, а параметров, например, 5, то матрица имеет размерность 12 x 5. Каждый элемент матрицы – кривая, зависимость *i*-го показателя от *j*-го параметра. Каждая точка кривой – среднее значение показателя на достаточно представительном отрезке T_n или усреднено по нескольким экспериментам.

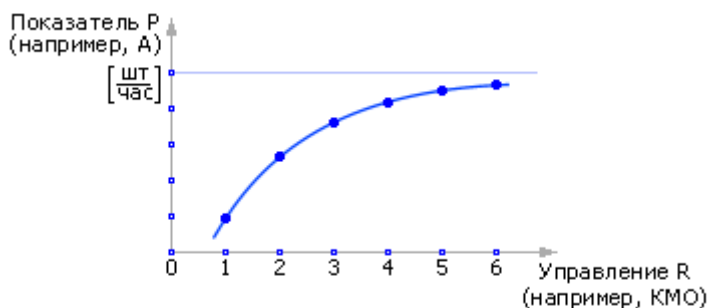
Следует понимать, что кривые снимались в предположении того, что все параметры кроме одного в процессе их снятия были неизменны. Если бы все параметры меняли значения, то кривые были бы другими. Но так не делают, так как получится полная неразбериха и зависимостей не будет видно.

Поэтому, если на основании рассмотрения снятых кривых принимается решение, что некоторый параметр будет в СМО изменен, то все кривые для новой точки, в которой опять будет исследоваться вопрос о том, какой параметр следует изменить, чтобы улучшить показатели, **следует снимать заново**.

Так шаг за шагом можно попытаться улучшить качество системы. Но пока эта методика не может ответить на ряд вопросов. Дело в том, что, во-первых, если кривые монотонно растут, то возникает вопрос, где же все-таки следует остановиться. Во-вторых, могут возникать противоречия, один показатель может улучшаться при изменении выбранного параметра, в то время как другой будет одновременно ухудшаться. В-третьих, ряд параметров сложно выразить численно, например, изменение дисциплины обслуживания, изменение направлений потоков, изменение топологии СМО. Как принять решение, каким должно быть все-таки значение параметра, если с его ростом

показатель все время монотонно улучшается? Вряд ли значение бесконечности устроит инженера.

Параметр *R* – управле-



ние, это то, что находится в распоряжении владельца СМО (например, возможность заасфальтировать площадку и тем самым увеличить количество мест в очереди, поставить дополнительные каналы, увеличить поток заявок за счет увеличения затрат на рекламу и так далее). Меняя управление, можно влиять на значение показателя P , цель, критерий (вероятность отказов, пропускную способность, среднее время обслуживания и так далее). Из **рис. 2.25** видно, что если увеличивать управление R , то можно добиться всегда улучшения показателя P . Но очевидно, что любое управление связано с затратами Z . И чем больше прилагают усилия для управления, тем больше значение управляющего параметра, тем больше затраты. Обычно затраты на управление растут линейно: $Z = C_1 \cdot R$. Однако бывают случаи, когда они растут экспоненциально, иногда – обратно экспоненте (скидки за опт) и т. д.

В любом случае ясно, что когда-то вложение все новых затрат просто перестанет себя окупать. Например, эффект от заасфальтированной площадки размером в 1 км^2 вряд ли окупит затраты владельца бензоколонки в Урюпинске, там просто не будет столько желающих заправиться бензином. Иными словами показатель P в сложных системах не может расти бесконечно. Рано или поздно его рост замедляется. А

Рис. 2.25. Зависимость показателя P от управляемого параметра R (пример)

затраты Z растут (см. **рис. 2.26**).

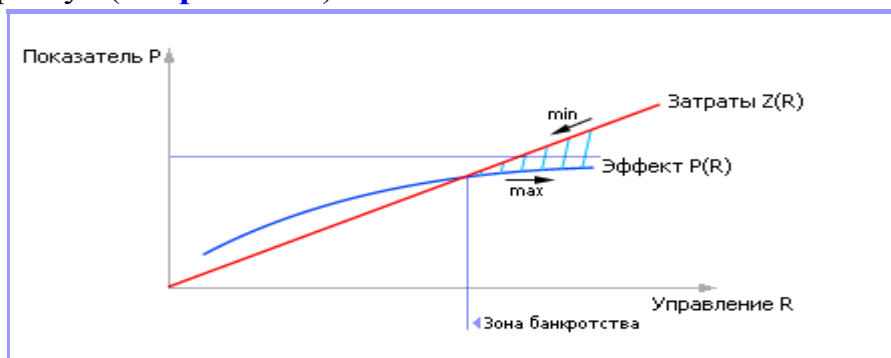


Рис. 2.26. Зависимости эффекта от применения показателя P и затрат Z на его получение как функции управляемого параметра R

Из **рис. 2.27** видно, что при назначении цены C_1 за единицу затрат R и цены C_2 за единицу показателя P , эти кривые можно сложить. Кривые складываются, если их требуется одновременно минимизировать или максимизировать. Если одна кривая подлежит максимизации, а другая минимизации, то следует найти их разность, например по точкам. Тогда результирующая кривая (см. **рис. 2.28**), учитывающая и эффект от управления и затраты на это, будет иметь экстремум. Значение параметра R , доставляющего экстремум функции, и есть решение задачи синтеза.

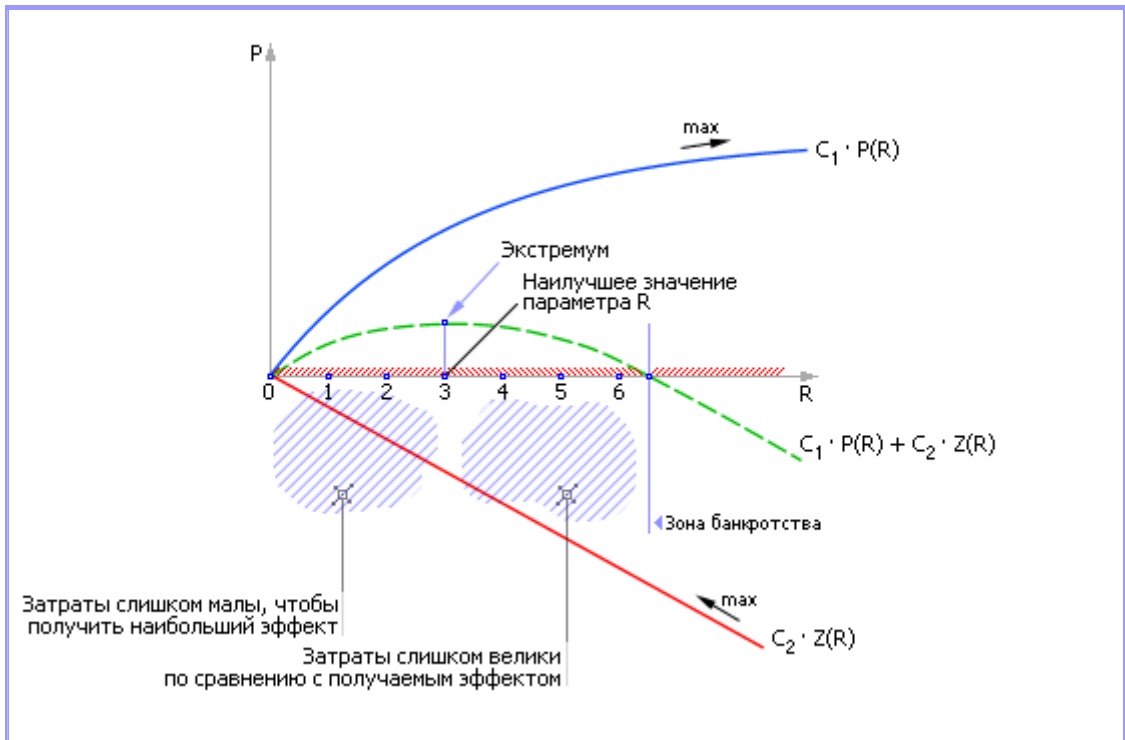


Рис. 2.27. Суммарная зависимость эффекта от применения показателя P и затрат Z на его получение как функции управляемого параметра R

Кроме управления R и показателя P в системах действует возмущение. Возмущения обозначим как $D = \{d_1, d_2, \dots\}$, см. [рис. 2.28](#). Возмущение – это входное воздействие, которое, в отличие от управляющего параметра, не зависит от воли владельца системы. Например, низкие температуры на улице, конкуренция снижают, к сожалению, поток клиентов, поломки оборудования досадно снижают производительность системы. И управлять этими величинами непосредственно владелец системы не может. Обычно возмущение действует «назло» владельцу, снижая эффект P от управляющих усилий R . Это происходит потому, что, в общем случае, система создается для достижения целей, недостижимых самих по себе в природе. Человек, организуя систему, всегда надеется посредством ее достичь некоторой цели P . На это он затрачивает усилия R , идя наперекор природе. Система – организация доступных человеку, изученных им природных компонент для достижения некоторой новой цели, недостижимой ранее другими способами.

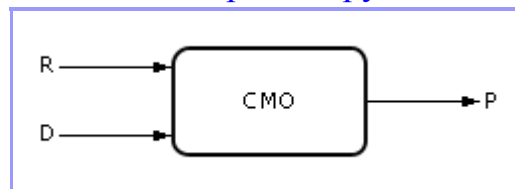


Рис. 2.28. Условное обозначение изучаемой системы, на которую воздействуют управляющие воздействия R и возмущения D

Итак, если мы снимем зависимость показателя P от управления R еще раз (как показано на [рис. 2.25](#)), но в условиях появившегося возмущения D , то, возможно, характер кривой изменится. Скорее всего, показатель будет при одинаковых значениях управлений находиться ниже, так как возмущение

носит «противный» характер, снижая показатели системы (см. [рис. 2.29](#)). Система, предоставленная сама себе, без усилий управляющего характера, перестает обеспечивать цель, для достижения которой она была создана. Если, как и ранее, построить зависимость затрат, соотнести ее с зависимостью показателя от параметра управления, то найденная точка экстремума сместится (см. [рис. 2.30](#)) по сравнению со случаем «возмущение = 0» (см. [рис. 2.27](#)).

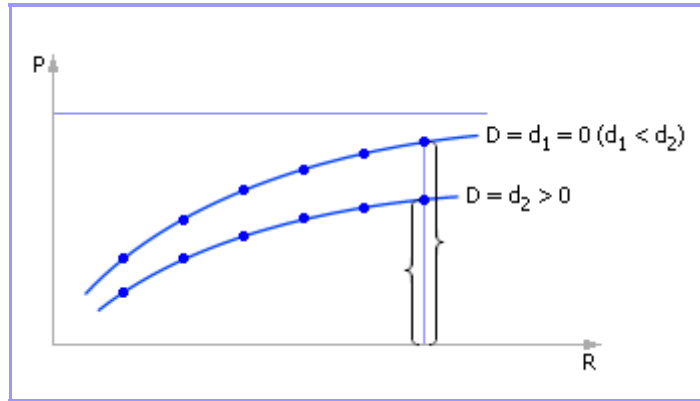


Рис. 2.29. Зависимость показателя P от управляющего параметра R при различных значениях действующих на систему возмущений D

Если снова увеличить возмущение, то кривые изменятся ([рис. 2.29](#)) и, как следствие, снова изменится положение точки экстремума ([рис. 2.30](#)).

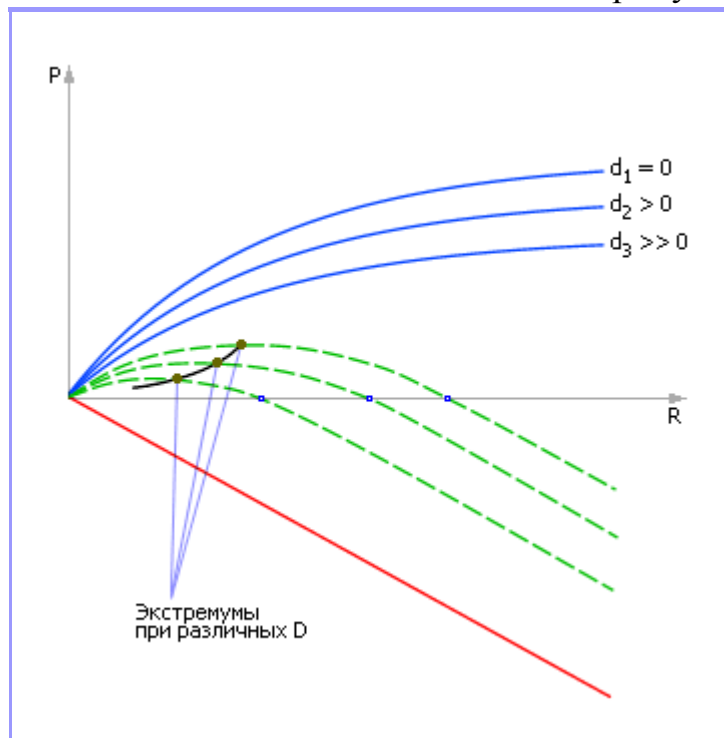


Рис. 2.30. Нахождение точки экстремума на суммарной зависимости при различных значениях действующего возмущающего фактора D

В конечном итоге, все найденные положения точек экстремума переносятся на новый график, где образуют зависимость Показателя P от Управляющего параметра R при изменении Возмущений D (см. [рис. 2.31](#)).

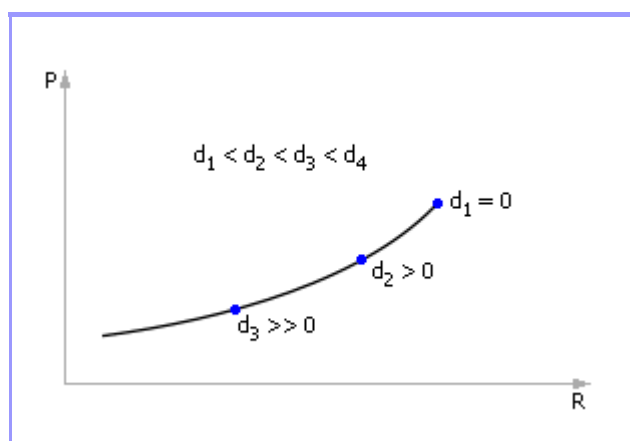


Рис. 2.31. Зависимость показателя P от управляющего параметра R при изменении значений возмущений D (кривая состоит только из точек экстремумов)

Обратите внимание, что на самом деле на этом графике могут быть и другие рабочие точки (график пронизан как бы семействами кривых), но нанесенные нами точки задают такие координаты управляющего параметра, при которых при заданных возмущениях (!) достигается наибольшее из возможных значение показателя P .

Этот график (см. **рис. 2.31**) связывает Показатель P , Управление (ресурс) R и Возмущение D в сложных системах, указывая, как действовать наилучшим образом ЛПР (лицу, принимающему решение) в условиях возникших возмущений. Теперь пользователь может, зная реальную обстановку на объекте (значение возмущения), быстро по графику определить, какое управляющее воздействие на объект необходимо, чтобы обеспечить наилучшее значение интересующего его показателя.

Заметьте, если управляющее воздействие будет меньше оптимального, то суммарный эффект снизится, возникнет ситуация недополученной прибыли. Если управляющее воздействие будет больше оптимального, то эффект **также** снизится, так как заплатить за очередное увеличение управляющих усилий надо будет по величине больше, чем та, которую вы получите в результате ее использования (ситуация банкротства).

Примечание. В тексте лекции мы использовали слова «управление» и «ресурс», то есть считали, что $R = U$. Следует пояснить, что управление действительно играет роль некоторой ограниченной ценности для владельца системы. То есть всегда является ценным для него ресурсом, за который всегда приходится платить, и которого всегда не хватает. Действительно, если бы эта величина не была ограничена, то мы бы могли достигать за счет бесконечной величины управлений бесконечно больших значений целей, а вот бесконечно больших результатов в природе явно не наблюдается.

Иногда различают собственно управление U и ресурс R , называя ресурсом некоторый запас, то есть границу возможного значения управляющего воздействия. В этом случае понятия ресурс и управление не совпадают: $U < R$. Иногда различают предельное значение управления $U \leq R$ и интегральный ресурс $\int U dt \leq R$.

на начало лекции
на содержание документа