



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ,
ФИЛИАЛ В Г.НИЖНЕВАРТОВСКЕ
КАФЕДРА «ИНФОРМАТИКА»



Моделирование систем

Методические указания к выполнению
лабораторных работ для студентов направления
«Информатика и вычислительная техника»

Нижевартовск 2016

*Одобрено
редакционно-издательским советом филиала*

Моделирование систем: методические указания к выполнению лабораторных работ для студентов направления «Информатика и вычислительная техника» / сост. А.В. Ялаев. – Нижневартовск, 2016. – 85 с.

Задания составлены в соответствии с ФГОС-3+ по направлению обучения «Информатика и вычислительная техника» и предназначены для формирования компетенции ПК-3.

Утверждено на заседании кафедры
Протокол №3
«3» ноября 2016 г.

Лабораторная работа №1

Моделирование случайных чисел

1 Цель работы

Изучить методы и алгоритмы моделирования случайных величин.

2 Теоретические сведения

2.1 Основные вероятностные понятия

Случайным опытом или *экспериментом* называется процесс, при котором возможны различные исходы, так что нельзя заранее предсказать, каков будет результат. Величина $X = \{x_i\} = x_1, x_2, \dots, x_n$, представляющая собой результат случайного опыта, называется *случайной величиной*. Непостоянство результата такого опыта может быть связано с наличием случайных ошибок измерений или со статистической природой самой измеряемой величины (например, процесс распада радиоактивного вещества). Будем обозначать отдельные значения, которые принимает случайная величина (не обязательно численные), как x_i , где $i = 1, 2, \dots, n$. Любая функция от x_i будет также случайной величиной.

Под *моделированием* случайной величины X принято понимать процесс получения на ЭВМ её выборочных значений x_1, \dots, x_n .

На практике используются три основных способа генерации случайных чисел:

- *табличный* (файловый) – ввод таблиц равномерно распределённых случайных чисел во внешнюю или оперативную память ЭВМ;
- *аппаратный* (физический) – использование специального приспособления к ЭВМ – "датчика" случайных чисел, формирующего случайные величины путём физического моделирования некоторых случайных процессов (излучения радиоактивных источников, шумов электронных ламп и др.);
- *алгоритмический* (программный) – использование псевдослучайных (квазислучайных) последовательностей, реализуемых программным генератором случайных чисел. *Псевдослучайными* числами называются числа, вырабатываемые ЭВМ рекуррентным способом по специальным алгоритмам, когда каждое последующее число x_i получается из предыдущих в результате применения некоторых арифметических и логических операций. Такая последовательность чисел удовлетворяет известным критериям случайности, хотя входящие в эту последовательность числа зависимы между собой. Одним из недостатков этого метода является периодичность образованных программным способом псевдослучайных чисел, но для ряда задач, не требующих большого количества случайных чисел, длина периода является достаточной.

Первый способ. При решении задачи без применения ЭВМ чаще всего используют *таблицы случайных чисел*. Таблицы получают с помощью специальных приборов (типа рулетки) и заносят в память ЭВМ, используются по мере необходимости.

В таблицах случайных чисел случайные цифры имитируют значения дискретной случайной величины с равномерным распределением:

x_i	0	1	2	3	...	9
p_i	0,1	0,1	0,1	0,1	...	0,1

При составлении таких таблиц выполняется требование, чтобы каждая из этих цифр от 0; 1; ...; 9 встречалась примерно одинаково часто и независимо от других с вероятностью $p_i = 0,1$.

Самая большая из опубликованных таблиц случайных чисел содержит 1 000 000 цифр. Таблицы случайных чисел составить не так просто. Они требуют тщательной проверки с помощью специальных статистических тестов.

Основной недостаток – необходимость в памяти достаточно большой емкости, затрудняющий решение "больших" задач, тем более что преимущество "случайных" таблиц перед "псевдослучайными" числами, получаемыми алгоритмически, никем не было доказано.

Во *втором способе* используются аппаратные датчики, основанные на некоторых физических процессах, случайных по своей природе (шумы в электронных и полупроводниковых приборах, процессы при радиоактивном распаде и т.п.).

Или же при решении задач на ЭВМ для выработки случайных чисел, равномерно распределенных в интервале (0;1), могут применяться *генераторы случайных чисел*. Данные генераторы преобразуют результаты случайного физического процесса в двоичные числа. В качестве случайного физического процесса обычно используют собственные шумы (случайным образом меняющееся напряжение).

Основные недостатки – невозможность повторного получения одной и той же последовательности случайных величин для проверочных расчетов и невозможность гарантировать постоянную надежную работу датчика.

Третий способ. Получение *псевдослучайных чисел* с равномерным законом распределения заключается в выработке псевдослучайных чисел. *Псевдослучайные числа* – это числа, полученные по какой-либо формуле и имитирующие значения случайной величины. Под словом "имитирующие" подразумевается, что эти числа удовлетворяют ряду тестов так, как если бы они были значениями этой случайной величины.

Первый алгоритм для получения псевдослучайных чисел предложил Дж. Нейман. Это так называемый метод середины квадратов, который заключается в следующем:

$$x_0 = 0,9876, x_0^2 = 0,97535376,$$

$$x_1 = 0,5353, x_1^2 = 0,28654609,$$

$$x_2 = 0,6546 \quad \text{и т.д.}$$

Метод средин квадратов фон Неймана является сравнительно бедным источником случайных чисел, т.к. последовательность стремится войти в привычную колею, т.е. короткий цикл повторяющихся элементов.

Назовем достоинства метода псевдослучайных чисел.

1. На получение каждого случайного числа затрачивается несколько простых операций, так что скорость генерирования случайных чисел имеет тот же порядок, что и скорость работы ЭВМ.

2. Малый объем памяти ЭВМ для программирования.

3. Любое из чисел легко воспроизвести.

4. Качество генерируемых случайных чисел достаточно проверить один раз.

Случайные величины бывают дискретные и непрерывные, одномерные (зависящие от одной переменной) или многомерные (зависящие от двух и более переменных).

Дискретной случайной величиной называется такая величина, число возможных значений которой либо конечное, либо бесконечное счетное множество (множество, элементы которого могут быть занумерованы).

Непрерывной случайной величиной называется такая величина, возможные значения которой непрерывно заполняют некоторый интервал (конечный или бесконечный) числовой оси.

Полной характеристикой случайной величины X с вероятностной точки зрения является ее *закон распределения*, т.е. заданная в той или иной форме связь между возможными значениями случайной величины и вероятностями их появления.

Универсальной формой закона распределения (непрерывных и дискретных величин) является *функция распределения вероятностей* – это такая функция $F(x)$, значение которой в точке x равно вероятности (P) того, что при проведении опыта значение случайной величины X окажется меньше, чем x :

$$F(x) = P(X < x).$$

Основные свойства функции распределения вероятностей следующие:

- 1) числовые значения заключены в пределах $0 \leq F(x) \leq 1$;
- 2) если $x_1 \leq x_2$, то $F(x_1) \leq F(x_2)$, т.е. $F(x)$ – неубывающая функция;
- 3) $F(x) \rightarrow 0$ при $x \rightarrow -\infty$, $F(x) \rightarrow 1$ при $x \rightarrow \infty$.

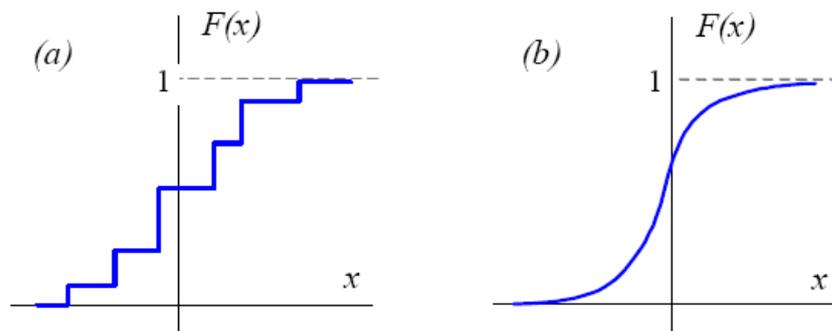


Рисунок 1 – Графическое изображение функции распределения вероятностей

Если случайная величина дискретна, то ее функция распределения представляет собой ступенчатую функцию (рисунок 1, а), а у непрерывных случайных величин функция распределения также непрерывна (рисунок 1, б).

Функцию распределения вероятностей $F(x)$ непрерывной случайной величины можно представить в виде интеграла от некоторой неотрицательной функции $f(x)$:

$$F(x) = \int_{-\infty}^x f(u) du .$$

Функция $f(x)$ называется *плотностью распределения вероятности*. Основные свойства плотности вероятности таковы:

$$1) f(x) = F'(x) = \frac{dF(x)}{dx}; \quad F(x) = \int_{-\infty}^x f(u) du ;$$

$$2) \int_{-\infty}^{\infty} f(u) du = 1;$$

3) плотность вероятности пропорциональна вероятности события ($x \leq X \leq x + dx$).

Кроме закона распределения, случайную величину характеризуют значениями некоторых параметров, определяющих наиболее существенные особенности ее распределения. Наиболее часто используемыми параметрами распределения являются математическое ожидание или среднее значение случайной величины, а также дисперсия случайной величины.

2.1.1 Параметры случайной величины

Основное назначение числовых характеристик случайной величины состоит в том, чтобы в сжатой форме выразить наиболее существенные особенности того или иного распределения.

Математическим ожиданием или *средним значением* дискретной случайной величины называется сумма всех возможных значений x_i случайной величины X , умноженных на соответствующие вероятности:

$$M\{X\} \equiv \bar{x} = \sum_{i=1}^n x_i P(X = x_i) = \sum_{i=1}^n x_i P_i ,$$

$$M\{X^2\} = \sum_{i=1}^n (x_i)^2 P_i,$$

$$(M\{X\})^2 = \left(\sum_{i=1}^n x_i P_i \right)^2.$$

Заметим, что x является не случайной, а определенной, детерминированной величиной.

Так как функция от случайной величины является также случайной величиной, то математическое ожидание функции $Y = H(X)$ определяется следующим образом:

$$M\{H(X)\} = \sum_{i=1}^n H(x_i) P(X = x_i) = \sum_{i=1}^n H(x_i) P_i.$$

Для непрерывных случайных величин будем иметь

$$M\{X\} \equiv \bar{x} = \int_{-\infty}^{\infty} x f(x) dx \quad \text{и} \quad M\{H(X)\} \equiv \bar{x} = \int_{-\infty}^{\infty} H(x) f(x) dx.$$

Важной характеристикой отклонения или разброса случайной величины от ее среднего значения является *дисперсия* случайной величины, определяемая как математическое ожидание квадрата отклонения случайной величины X от своего среднего значения:

$$D\{X\} \equiv \sigma^2(X) = M\{(X - M\{X\})^2\} = M\{(X - \bar{x})^2\} \equiv M\{X^2\} - (M\{X\})^2.$$

Положительный квадратный корень из дисперсии $\sigma = +\sqrt{\sigma^2(X)}$ называется *стандартным* или *среднеквадратичным отклонением*¹. Среднеквадратичное отклонение количественно показывает, насколько сильно значения случайной величины X разбросаны вокруг среднего значения \bar{x} .

Пример 1. В качестве примера дискретной случайной величины рассмотрим числа, выпадающие при бросании игрального кубика.

Пусть N раз бросили игральный кубик и получили $N_1, N_2, N_3, N_4, N_5, N_6$ выпадений значений 1, 2, 3, 4, 5, 6 соответственно. Тогда говорят, что вероятность выпадения какого-нибудь числа i ($i = 1, 2, 3, 4, 5, 6$) приблизительно равна

$$P_i \approx \frac{N_i}{N},$$

т.е. P_i равна доле числа случаев, в которых выпало значение i , от полного числа бросаний. Знак приближенного равенства означает, что если повторить еще N бросаний, то получим, вообще говоря, другое значение N_i . Соотношение для вероятности становится точным в пределе, когда $N \rightarrow \infty$:

$$P_i \approx \lim_{N \rightarrow \infty} \frac{N_i}{N}.$$

В нашем случае, если кубик "честный", вероятности выпадения значений 1, 2, 3, 4, 5, 6 равны $P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = 1/6$.

¹ Часто используют также эквивалентный термин *квадратичное отклонение*.

Очевидно также, что

$$\frac{1}{N} \cdot \sum_{i=1}^6 N_i = \sum_{i=1}^6 P_i = 1.$$

Вычислим математическое ожидание $M\{X\}$ и дисперсию $D\{X\}$ для игрального кубика:

$$M\{X\} = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5,$$

$$D\{X\} = M\{X^2\} - (M\{X\})^2 = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} - (3.5)^2 = 2.917.$$

2.2 Моделирование случайных событий

2.2.1 Получение равномерно распределенных случайных чисел

При моделировании систем на ЭВМ программная имитация случайных воздействий любой сложности сводится к генерированию некоторых стандартных (базовых) процессов и к их последующему функциональному преобразованию. В качестве базового процесса примем последовательность случайных чисел $\{x_i\} = x_1, x_2, \dots, x_n$, представляющих собой реализации независимых, равномерно распределенных на интервале $(0;1)$ случайных величин $\{r_i\} = r_1, r_2, \dots, r_n$.

Непрерывная случайная величина R имеет равномерное распределение на интервале $(a; b)$, если ее функции плотности вероятности (рис. 2, а) и распределения (рис. 2, б) задаются следующим образом:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{если } a \leq x \leq b \\ 0, & \text{в противном случае.} \end{cases} \quad F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}.$$

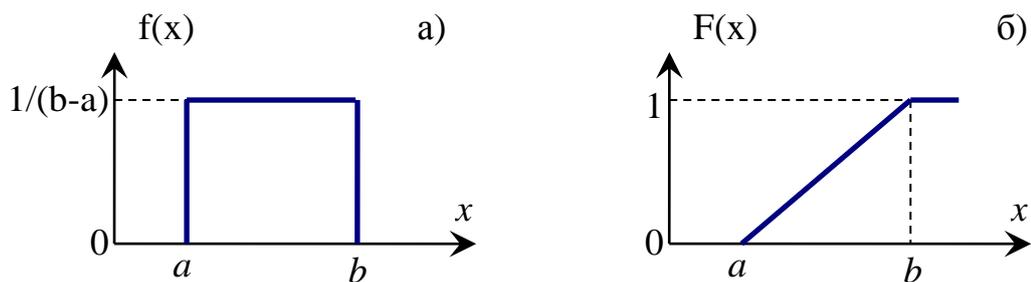


Рисунок 2 – Графическое изображение равномерно распределенной случайной величины X

Легко можно вычислить, что $M\{X\} = \bar{x} = \int_a^b x f(x) dx = \frac{b+a}{2}$, а

$$D\{X\} = M\{(X - M\{X\})^2\} = \int_a^b \left(x - \frac{b+a}{2}\right)^2 f(x) dx = \frac{(b-a)^2}{12}.$$

В частном случае, когда $a = 0$ и $b = 1$, имеем равномерно распределенную на интервале $(0;1)$ случайную величину, для которой математическое ожидание $M(X)=1/2$, а дисперсия $D(X) = 1/12$.

От последовательности случайных чисел, равномерно распределенных в интервале $(0;1)$, нетрудно перейти к последовательности случайных чисел с произвольным заданным законом распределения.

Пример 2. Получение равномерно распределенной в интервале $(0;1)$ случайной величины X может быть осуществлено следующим алгоритмом:

$$x_{i+1} = \{ \pi \cdot x_i \},$$

$$x_0 = 0.1.$$

Знак $\{ \}$ означает, что берется дробная часть произведения. Вычисления дают такую последовательность: $x_0 = 0.1$, $x_1 = 0.415926$, $x_2 = 0.667$, $x_3 = 0.54422$, $x_4 = 0.97175$, $x_5 = 0.28426$ и т.д.

К настоящему времени разработано множество алгоритмов получения псевдослучайных чисел. Наиболее популярным для получения псевдослучайных чисел x_1, x_2, \dots, x_n является **метод вычетов** (мультипликативный датчик), который можно записать в следующей форме:

$$x_{i+1} = \{ M \cdot x_i \}, x_0 = 2^{-m},$$

где M – достаточно большое целое число, фигурные скобки обозначают дробную часть, а m – число двоичных разрядов в мантиссе чисел в ЭВМ.

Методы выбора значений M , x_0 и m разнятся для разных вариантов реализаций данного метода (это своя собственная "наука") и определяют основные свойства датчика случайных чисел (соответствие статистическим критериям, длину периода повторения последовательности и т.п.).

2.2.2 Моделирование нормально распределенных случайных величин

В технике и природе наиболее распространенное распределение случайных чисел – гауссовское или нормальное.

Нормальной (или гауссовской) называется случайная величина X , определенная на всей числовой оси $(-\infty, \infty)$ имеющая плотность распределения вероятности:

$$f(x) = \left(\frac{1}{\sqrt{2\pi D}} \right) \exp \frac{-(x-M\{X\})^2}{2D}.$$

Здесь $D=\sigma^2$ – дисперсия случайной величины, а $M\{X\}$ – математическое ожидание случайной величины X .

Распространенный алгоритм моделирования таких величин основан на *центральной предельной теореме теории вероятностей*.

Рассмотрим алгоритм моделирования нормально распределенных случайных величин.

Известно, что если случайная величина X распределена равномерно в интервале $(0;1)$, то ее математическое ожидание $M(X)=1/2$, а дисперсия $D(X) = 1/12$.

Для практического использования можно считать, что случайная величина $X = \sum_{i=1}^n x_i$ распределена нор-

мально при $n \geq 8$ (n – количество случайных величин) с математическим ожиданием $M\{X\}=n/2$, дисперсией $D\{X\}=n/12$ и среднеквадратичным отклонением $\sigma = \sqrt{D\{X\}} = \sqrt{n/12}$.

Величина $M(X)$ характеризует центр тяжести распределения X и не влияет на форму кривой. Величина σ же характеризует разброс случайной величины X относительно ее среднего значения $M(X)$ (см. рисунок 3).

Так как моделирование любого нормального распределения с параметрами $(M\{X\}, \sigma)$ может быть осуществлено по очевидной формуле $X = M\{X\} + \sigma \cdot R$, где $R = \{r_1, r_2, \dots, r_n\}$ – сгенерированная в интервале $[0;1)$ случайная величина, то нормально распределенные случайные величины с параметрами $(0;1)$ можно вычислять по следующей приближенной формуле:

$$R = \left(\frac{12}{n}\right)^{1/2} \cdot \left(\sum_{i=1}^n r_i - \frac{1}{2} \cdot n\right).$$

При $n = 12$ эта формула заметно упрощается: $R = \sum_{i=1}^{12} r_i - 6$.

Пример 3.

1. Разыграть 100 возможных значений случайной величины X распределенной нормально с параметрами $M\{X\} = 0$ и $\sigma = 1$.

2. Оценить параметры разыгранной случайной величины X .

Решение:

1. Выберем 12 случайных чисел распределенных равномерно в интервале $(0;1)$ из таблицы случайных чисел, либо из компьютера. Сложим эти числа и из суммы вычтем 6, в итоге получим:

$$X_1 = R = \sum_{i=1}^{12} r_i - 6 = (0,10 + 0,09 + \dots + 0,67) - 6 = -0,99.$$

Поступая аналогичным образом, найдем остальные возможные значения X_2, X_3, \dots, X_{100} .

2. Выполнив необходимые расчеты, найдем выборочную среднюю, которая является оценкой $M\{X\}$ и выборочное среднее квадратичное отклонение (выборочная дисперсия), которое является оценкой σ .

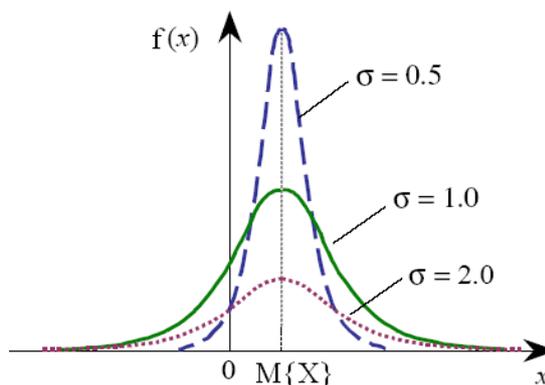


Рисунок 3 - График нормального распределения

Выборочное среднее: $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$.

Выборочная дисперсия: $\sigma_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x}_n)^2$.

Величина \bar{x}_n называется *выборочным средним*, σ_n^2 – *выборочной дисперсией* случайной величины X . Значения \bar{x}_n и σ_n^2 можно принять в качестве оценок математического ожидания $M\{X\}$ и дисперсии $D\{X\}$ величины X , т.е. $M\{X\} \equiv \bar{x} \approx \bar{x}_n$, $D\{X\} \equiv \sigma^2 \approx \sigma_n^2$. Приближенные равенства становятся точными в пределе, когда $n \rightarrow \infty$. *Выборочное среднеквадратичное отклонение* σ_n равно корню квадратному из выборочной дисперсии $\sigma_n = \sqrt{\sigma_n^2}$.

Получим:

$$M\{X\}^* = \bar{x}_n \approx -0.05; \sigma^* = \sqrt{D_n} \approx 1.04.$$

Как видим, оценки удовлетворительны, т.е. $M\{X\}^*$ близко к нулю, а σ^* близко к единице.

Если требуется разыграть значения нормальной ненормированной случайной величины с математическим ожиданием $M\{X\}$ отличным от нуля и σ отличным от единицы, то сначала разыгрывают возможные значения r_i нормированной случайной величины в интервале $[0;1)$, а затем находят искомое значение по формуле

$$X_i = M\{X\} + \sigma \cdot r_i,$$

которая получена из соотношения: $r_i = \frac{X_i - M\{X\}}{\sigma}$.

2.2.3 Моделирование дискретных случайных величин

Рассмотрим дискретную случайную величину X , принимающую n значений x_1, x_2, \dots, x_n с вероятностями P_1, P_2, \dots, P_n . Эта величина задается таблицей распределения

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ P_1 & P_2 & \dots & P_n \end{pmatrix} \quad \sum_{i=1}^n P_i = 1.$$

или

X	x ₁	x ₂	x ₃	...	x _n
P	P ₁	P ₂	P ₃	...	P _n

Для моделирования такой дискретной случайной величины разбивают отрезок $[0;1]$ на n последовательных отрезков $\Delta_1, \Delta_2, \dots, \Delta_n$, длины которых равны соответствующим вероятностям P_1, P_2, \dots, P_n .

Тогда длины отрезков будут равны:

Длина $\Delta_1 = P_1 - 0 = P_1$

Длина $\Delta_2 = (P_1 + P_2) - P_1 = P_2$

.....

Длина $\Delta_n = 1 - (P_1 + P_2 + \dots + P_{n-1}) = P_n$

Видно, что длина частичного интервала с индексом i равна вероятности P с тем же индексом. Длина $\Delta_i = P_i$.

Получают случайную величину R , равномерно распределенную в интервале $(0;1)$. Таким образом, при попадании случайного числа r_i в интервал Δ_i случайная величина X принимает значение x_i с вероятностью P_i .

Теорема: *если каждому случайному числу $r_i(0 \leq r_i < 1)$, которое попало в интервал Δ_i , поставить в соответствие возможное значение x_i , то разыгрываемая величина будет иметь заданный закон распределения.*

Рассмотрим алгоритм моделирования дискретных случайных величин.

1. Нужно разбить интервал $[0;1)$ на n частичных интервалов:

$$\Delta_1 = (0; P_1), \Delta_2 = (P_2; P_1 + P_2), \dots, \Delta_n = (P_1 + P_2 + \dots + P_{n-1}; 1).$$

2. Выбрать (например, из таблицы случайных чисел, или в компьютере) случайное число r_i из интервала $[0;1)$.

Если r_i попало в интервал Δ_i , то моделируемая дискретная случайная величина приняла возможное значение x_i .

Пример 4. Смоделировать 8 значений дискретной случайной величины, заданной таблицей распределения:

X	$X_1 = 3$	$X_2 = 11$	$X_3 = 24$
P	$P_1 = 0,25$	$P_2 = 0,16$	$P_3 = 0,59$

Решение:

1. Разобьем интервал $(0;1)$ точками с координатами $0,25; 0,25+0,16=0,41$ на три частичных интервала:

$$\Delta_1 = [0;0,25), \Delta_2 = [0,25;0,41), \Delta_3 = [0,41;1).$$

2. Сгенерируем с помощью компьютера 8 случайных чисел, например, $r_1 = 0,10; r_2 = 0,37; r_3 = 0,08; r_4 = 0,99; r_5 = 0,12; r_6 = 0,66; r_7 = 0,31; r_8 = 0,85$.

3. Случайное число $r_1 = 0,10$ принадлежит первому частичному интервалу, поэтому разыгрываемая случайная величина приняла возможное значение $x_1 = 3$. Случайное число $r_2 = 0,37$ принадлежит второму частичному интервалу, поэтому разыгрываемая величина приняла возможное значение $x_2 = 11$. Аналогично получим остальные возможные значения дискретной случайной величины X .

Результат: последовательность смоделированных возможных значений дискретной случайной величины X такова: 3; 11; 3; 24; 3; 24; 11; 24.

2.2.4 Моделирование непрерывных случайных величин

Может оказаться так, что алгоритм численного решения указанного уравнения будет достаточно сложным или требовать заметных затрат времени на вычисления. Тогда могут быть использованы другие методы генерирования случайных величин. Среди этих методов отметим **метод исключения**.

Суть метода исключения (или метода Неймана) заключается в следующем.

Пусть случайная величина X определена на конечном интервале $(a;b)$ и плотность ее распределения ограничена, так что $f(x) \leq M$. Тогда, используя пару равномерно распределенных на интервале $(0;1)$ случайных чисел R , осуществляем следующие действия для розыгрыша (моделирования) значения X :

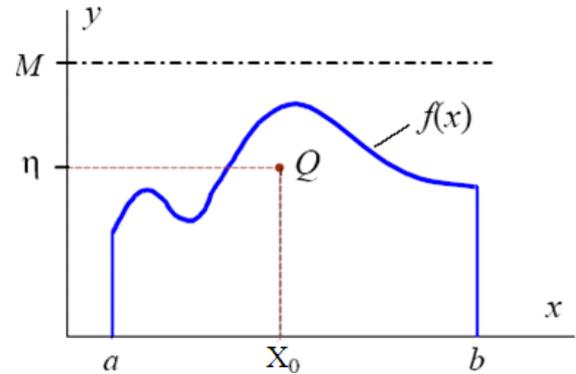


Рисунок 4 – Графическое изображение метода Неймана

1. Разыгрываем два значения r_1 и r_2 случайной величины R и строим случайную точку Q с координатами (см. рисунок 4):

$$X_0 = a + r_1 \cdot (b - a), \quad \eta = r_2 \cdot M.$$

2. Если $\eta > f(X_0)$, то пару значений (r_1, r_2) отбрасываем и переходим к пункту 1; иначе принимаем $X = X_0$.

Таким образом, определяются координаты случайной точки $Q(X_0, \eta)$ и, если точка окажется под кривой $f(x)$, то абсцисса этой точки принимается в качестве значения случайной величины $X = X_0 = a + r_1 \cdot (b - a)$ с плотностью распределения $f(x)$. В противном случае точка отбрасывается, определяются координаты следующей точки, и все повторяется.

Существуют и другие многочисленные способы формирования случайных величин с различными определенными законами распределения.

2.3 Проверка гипотезы о законе распределения методом гистограмм

Пусть в результате эксперимента получено n значений x_1, x_2, \dots, x_n случайной величины X и все они заключены в пределах $a < x_i < b$.

Суть проверки по гистограмме сводится к следующему.

а) Выдвигается гипотеза о равномерности распределения чисел в интервале $(0;1)$.

б) Затем интервал $(a; b)$ разбивается на L (любое число, не слишком большое и не слишком малое) равных подынтервалов длиной Δ_j , тогда при генерации последовательности $\{x_i\}$ каждое из чисел x с вероятностью $p_j = 1/L$, $j = \overline{1, L}$, попадает в один из подынтервалов. Всего в каждый j -й подынтервал попадает ν_j чисел последовательности $\{x_i\}$, $i = \overline{1, n}$, причем $\nu = \sum_{j=1}^L \nu_j$. Относи-

тельная частота попадания случайных чисел последовательности $\{x_i\}$ в каждый из подынтервалов будет равна v_j/n .

в) Над каждым из подынтервалов разбиения строится прямоугольник, площадь которого равна частоте попадания $\{x_i\}$ в этот подынтервал. Высота каждого прямоугольника равна частоте, деленной на Δ_j . Полученную ступенчатую линию называют *гистограммой*.

г) Вид соответствующей гистограммы представлен на рисунке 5, где пунктирная линия соответствует теоретическому значению p_j , а сплошная – экспериментальному v_j/n . Очевидно, что если числа $\{x_i\}$ принадлежат псевдослучайной равномерно распределенной последовательности, то при достаточно больших n экспериментальная гистограмма (ломаная линия на рис. 5) приблизится к теоретической прямой $p_j=1/L$.

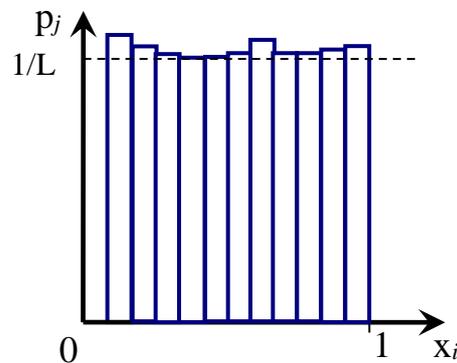


Рисунок 5 – Проверка равномерности последовательности

Гистограмма служит приближением к неизвестной плотности случайной величины X . Площадь гистограммы, заключенная между x_i и x_{i+1} , дает приближенное значение вероятности $P\{x_i < X < x_{i+1}\}$.

3 Порядок выполнения работы

1. Используя метод вычетов, сгенерировать последовательность из 1 000 псевдослучайных чисел, результат вывести на экран.

1.1. Оценить математическое ожидание полученной последовательности, математическое ожидание и выборочную среднюю вывести на экран.

1.2. Оценить дисперсию полученной последовательности, дисперсию и выборочную дисперсию вывести на экран.

1.3. Построить таблицу 1 (количество L подынтервалов не менее 10), частотную таблицу вывести на экран.

Таблица 1 – Частотная таблица

Интервал	Кол-во СВ (частота попаданий), выпавших в данный интервал	Относительная частота попадания
Δ_1	v_1	v_1/n
Δ_2	v_2	v_2/n
...
Δ_L	v_L	v_L/n
	\sum кол-во СВ	

1.4. Проверить гипотезу о законе распределения методом гистограмм, построить гистограмму, вывести ее на экран.

2. Смоделировать дискретную случайную величину, заданную таблицей 2, результат вывести на экран.

2.1. Оценить математическое ожидание полученной дискретной случайной величины, результат вывести на экран.

2.2. Оценить дисперсию полученной дискретной случайной величины, результат вывести на экран.

2.3. Построить частотную таблицу, вывести ее на экран.

2.4. Оценить закон распределения случайной величины по графику частоты появления ее значений в результате экспериментов.

3. Смоделировать методом исключений непрерывную случайную величину с заданной плотностью распределения вероятности (таблица 3).

Функции для графика рассчитываются по формулам $\frac{x-x_1}{x_2-x_1} = \frac{y-y_1}{y_2-y_1}$ или

$y = -kx + b$ (в зависимости от вида графика).

3.1. Оценить математическое ожидание полученной непрерывной случайной величины, результат вывести на экран.

3.2. Оценить дисперсию полученной непрерывной случайной величины, результат вывести на экран.

3.3. Построить частотную таблицу, вывести ее на экран.

3.4. Проверить гипотезу о законе распределения методом гистограмм, построить и вывести на экран гистограмму.

Таблица 2 – Таблица распределений

Вариант	Таблица распределения							
1	x_i	5	7	17	19	21	25	55
	p_i	0.01	0.05	0.3	0.3	0.3	0.02	0.02
2	x_i	1	3	7	10	15	18	23
	p_i	0.1	0.05	0.02	0.05	0.25	0.33	0.2
3	x_i	2	3	5	12	21	33	44
	p_i	0.1	0.15	0.2	0.05	0.02	0.33	0.15
4	x_i	5	8	13	16	21	24	29
	p_i	0.1	0.02	0.25	0.15	0.35	0.03	0.1
5	x_i	2	3	5	8	11	15	20
	p_i	0.1	0.15	0.25	0.05	0.05	0.3	0.1
6	x_i	1	8	17	23	37	42	50
	p_i	0.01	0.15	0.05	0.25	0.5	0.02	0.02
7	x_i	1	4	12	16	25	33	37
	p_i	0.05	0.25	0.25	0.15	0.13	0.1	0.07
8	x_i	1	10	15	23	29	38	42
	p_i	0.02	0.05	0.1	0.28	0.23	0.22	0.1
9	x_i	2	3	7	12	19	23	30
	p_i	0.04	0.15	0.2	0.25	0.2	0.15	0.01
10	x_i	1	5	7	14	21	26	31
	p_i	0.34	0.28	0.16	0.15	0.05	0.01	0.01
11	x_i	3	5	8	14	27	29	35
	p_i	0.02	0.07	0.1	0.19	0.19	0.2	0.23
12	x_i	7	16	28	33	39	46	56
	p_i	0.01	0.05	0.07	0.1	0.17	0.25	0.35
13	x_i	5	6	8	13	19	26	36
	p_i	0.05	0.07	0.2	0.23	0.17	0.23	0.05
14	x_i	3	9	18	23	29	27	45
	p_i	0.05	0.14	0.2	0.22	0.17	0.14	0.08
15	x_i	13	16	28	33	39	47	52
	p_i	0.08	0.14	0.25	0.16	0.25	0.09	0.03
16	x_i	1	6	8	13	19	24	27
	p_i	0.09	0.1	0.21	0.17	0.23	0.15	0.05
17	x_i	4	6	10	14	16	20	24
	p_i	0.04	0.1	0.1	0.27	0.33	0.13	0.03
18	x_i	2	6	12	16	22	26	32
	p_i	0.02	0.14	0.24	0.27	0.2	0.1	0.03
19	x_i	3	6	9	13	19	27	31
	p_i	0.04	0.12	0.22	0.28	0.2	0.1	0.04
20	x_i	1	3	8	11	19	29	33
	p_i	0.02	0.26	0.18	0.32	0.16	0.02	0.04

Таблица 3 – Плотность распределения вероятности

Вариант	Плотность распределения	Вариант	Плотность распределения
1		11	
2		12	
3		13	
4		14	
5		15	

Продолжение таблицы 3

Вариант	Плотность распределения	Вариант	Плотность распределения
6		16	
7		17	
8		18	
9		19	
10		20	

Лабораторная работа № 2

Моделирование случайных чисел с заданным законом распределения

1 Цель работы

Изучение методов и получение навыков составления алгоритмов генерации псевдослучайных чисел с различными законами распределения, наиболее часто используемых при имитационном моделировании сложных систем на ЭВМ.

2 Методические указания

В процессе анализа и проектирования имитационных моделей стохастических систем возникает необходимость задания различных случайных воздействий или имитирования стохастических процессов. Подобные ситуации предопределяют необходимость программной генерации случайных чисел с некоторым законом распределения. Пусть требуется получить (смоделировать) реализацию случайной величины X с плотностью распределения $f(x)$. Данная задача решается путем моделирования случайной величины R , равномерно распределенной на интервале $[0;1)$, и преобразования последовательности случайных чисел r_1, r_2, \dots, r_n в последовательность x_1, x_2, \dots, x_n . В общем случае преобразование можно реализовать с помощью некоторой функции

$$X=\psi(R), \quad (1)$$

связывающей случайные числа с равномерным распределением со случайными числами с заданным законом распределения. Преобразование (1) может быть выполнено различными методами.

2.1 Метод обратных функций

Пусть требуется получить значения случайной величины X , распределенной в интервале $(a;b)$ с плотностью вероятности $f(x)$.

Стандартный метод моделирования основан на том, что интегральная функция распределения $F(y) = \int_a^y f(x)dx$ любой непрерывной случайной величины равномерно распределена в интервале $(0;1)$, т.е. для любой случайной величины X с плотностью распределения $f(x)$ случайная величина равномерно распределена на интервале $(0;1)$.

Тогда случайную величину X с произвольной плотностью распределения $f(x)$ можно рассчитать по следующему алгоритму, графическое решение смотрите на рисунке 1:

1. Необходимо сгенерировать случайную величину r (значение случайной величины R), равномерно распределенную в интервале $(0;1)$.

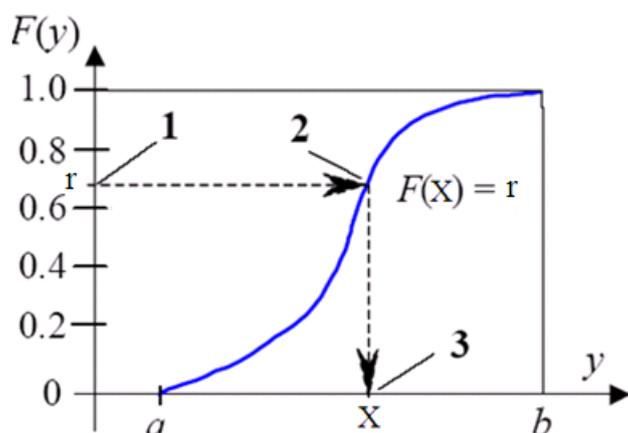


Рисунок 1 – Графическое изображение метода обратных функций

2. Приравнять сгенерированное случайное число известной функции распределения $F(X)$ и получить уравнение

$$F(X) = \int_a^x f(x)dx = r.$$

3. Решая уравнение $X=F^{-1}(r)$, находим искомое значение X .

Такой способ получения случайных величин называется **методом**

обратных функций.

Пример 1. Пусть требуется получить случайную величину X , распределенную в интервале $(a; b)$ с равномерной плотностью:

$$f(x) = \frac{1}{b-a}, \quad a < x < b.$$

Тогда $F(x) = \int_a^x \frac{dx}{b-a} = \frac{x-a}{b-a} = r$ и получаем следующее выражение:

$x = a + r \cdot (b-a)$. Эта формула часто используется для расширения или смещения стандартного интервала $(0;1)$ равномерно распределенной случайной величины до необходимого интервала $(a; b)$.

Следует отметить, что не всегда возможно так легко разрешить получаемые уравнения. Чаще всего аналитическое решение не существует, и тогда приходится прибегать к численному решению уравнения $x = F^{-1}(r)$.

Так выглядит пример теоретически, практически же:

Пример 2. Разыграть 3 возможных значения непрерывной случайной величины X , распределенной равномерно в интервале $(2; 10)$.

Решение:

Функция распределения величины X имеет следующий вид:

$$F(x) = (x-a)/(b-a).$$

По условию, $a = 2, b = 10$, следовательно, $F(x) = (x-2)/8$.

В соответствии с алгоритмом разыгрывания непрерывной случайной величины приравняем $F(x)$ выбранному случайному числу r_i . Получим отсюда:

$$(x_i - 2) / 8 = r_i \Rightarrow x_i = 8r_i + 2$$

Далее в соответствии с алгоритмом выберем три случайных числа, распределенных равномерно в интервале $(0;1)$. Например, $r_1 = 0,11; r_2 = 0,17;$

$r_3 = 0,66.0$

Подставим эти числа в уравнение. Получим соответствующие возможные значения X :

$$x_1=8 \cdot 0,11+2=2,88; x_2=8 \cdot 0,17+2=3,36; x_3=8 \cdot 0,66+2=7,28.$$

Пример 3. Показательное распределение. Метод логарифма¹.

Случайная величина X имеет экспоненциальный (показательный) закон распределения

$$F(X) = \int_0^x f(x)dx = 1 - \exp^{-\frac{x}{\lambda}}, \quad x \geq 0, \lambda > 0;$$

где λ – параметр распределения.

Очевидно, если $r=F(x)=1-\exp^{-\frac{x}{\lambda}}$, то $X=F^{-1}(r)=-\lambda \ln(1-r)$.

Так как $1-r$ имеет то же самое распределение, что и R , то удобнее при нахождении значений случайной величины X пользоваться формулой

$$X = -\lambda \ln(R). \quad (2)$$

Случай, когда $R=0$, должен трактоваться особо; его можно заменять любым подходящим значением \exp , так как вероятность возникновения этого случая крайне мала.

Пример 4. Обратную функцию для нормального распределения нельзя представить в виде простых, легко вычисляемых формул. Поэтому для получения случайных чисел с нормальным распределением используют аппроксимацию:

$$\sqrt{\frac{2}{\pi}} \exp^{-\frac{x}{2}} \approx \frac{2k \cdot \exp^{-kx}}{(1 + \exp^{-kx})^2}, \quad x > 0, \quad (3)$$

где $k = \sqrt{8/\pi}$.

Если воспользоваться (3) для случайной величины с плотностью

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp^{-\frac{x^2}{2}}, \quad -\infty < x < \infty,$$

то получим для положительной полуоси методом обратных функций величину

$$x = \frac{1}{k} \ln \frac{1+R}{1-R}. \quad (4)$$

Если значениям x , полученным с помощью (4), приписывать с вероятностью 0,5 знак “+” и с вероятностью 0,5 знак “-”, то полученная последовательность случайных чисел может рассматриваться как реализация случайных величин $S \cdot X$ с нормальным распределением $N(0;1)$, $-\infty < x < \infty$, где S – случайная величина с распределением

¹ Кнут Д.Э. Искусство программирования. Т.2. : Получисленные алгоритмы / Под общ. ред. Козаченко Ю.В. - 3-е изд. - М.: Вильямс, 2001. - 828с. - Предм.-имен. указ.:с. 801-828. [с. 157]

S	+ 1	- 1
p	0,5	0,5

Пример 5. Метод обратного преобразования можно также использовать, если величина X является *дискретной*. В этом случае функция распределения

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i),$$

где $p(x_i)$ – является вероятностной мерой $p(x_i) = P(X=x_i)$. (Допускается, что величина X может иметь только такие значения x_1, x_2, \dots , для которых $x_1 < x_2 < \dots$.) Тогда алгоритм будет иметь следующий вид.

1. Генерируем число R в интервале $(0;1)$.
2. Определяем положительное наименьшее целое число R_i , для которого $R_i \leq F(x_i)$, и возвращаем $X=x_i$.

Метод обратного преобразования для непрерывных и дискретных величин можно, формально, объединить в одну общую формулу

$$X = \min\{x: F(x) \geq R\} \tag{5}$$

Графически формула (5) представлена на рисунке 2.

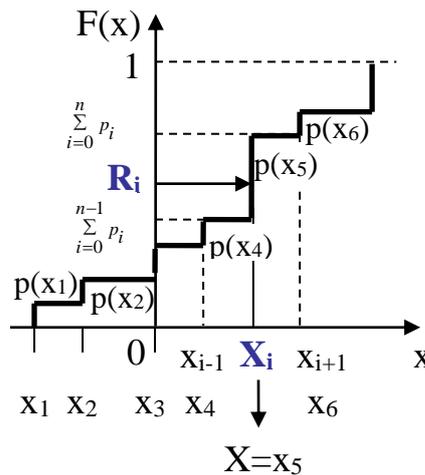


Рисунок 2 – Использование метода обратного преобразования для дискретных случайных величин

Таким образом, если явного выражения для обратной функции получить не удастся, то используются приближенные способы моделирования случайных величин. Они основаны на аппроксимации функции $F(x)$ некоторой другой функцией $G(x)$, обратная функция которой G^{-1} имеет более простое аналитическое выражение и легко вычислима.

Н.П. Бусленко предложен метод кусочно-линейной аппроксимации, при котором функция распределения $F(x)$ заменяется функцией $F^*(x)$, составленной из отрезков прямых. Для этого интеграл изменения функции $F(x)$ разбивается на n подынтервалов, число которых зависит от требуемой точности конечных результатов. На каждом из линейных участков $F^*(x)$, соответ-

ствующих этим подынтервалам, определяется обратная функция. При этом функция распределения на выделенных подынтервалах может быть аппроксимирована не только линейной, но и другой подходящей простой функцией, если требуется более высокая степень точности результатов моделирования.

Метод обратных функций применяется также при моделировании случайной величины X по эмпирической функции ее распределения, построенной по результатам выборки. На рисунке 3 показана реализация случайной величины X по эмпирической функции распределения, являющейся ступенчатой ломаной линией. Случайные числа (реализации случайной величины X) принимаются равными $(x_{k-1}+x_k)/2$, т.е. они соответствуют серединам интервалов $(x_{k-1}; x_k)$ если случайная величина R попадает в интервал $\left(\sum_{i=0}^{k-1} p_i; \sum_{i=0}^k p_i \right)$.

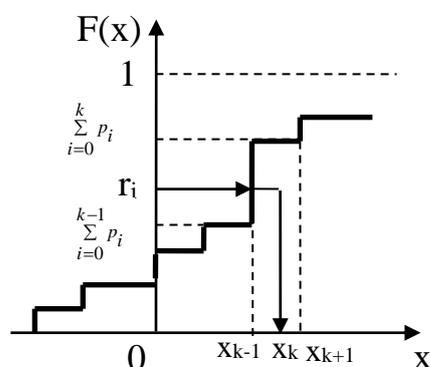


Рисунок 3 – Реализация случайной величины X по эмпирической функции распределения

2.2 Метод отбора

Данный метод предложен Дж. Нейманом и сущность его сводится к тому, что из области задания случайной величины X “отбирается” точка с координатами, являющимися фикциями случайных чисел с равномерным распределением; если эта точка не может быть использована для расчета X , происходит ее “отбрасывание” и “отбирается” новая. Метод отбора применим для получения реализаций только таких случайных величин, закон распределения которых может быть задан с помощью функции плотности.

В рамках метода отбора существует несколько процедур моделирования случайной величины с заданной плотностью распределения.

2.2.1 Метод отбора – процедура № 1

С помощью данной процедуры моделируется одномерная случайная величина X , определенная на интервале плотностью $f(x)$. Вне этого интервала

$f(x)=0$ и, кроме того, плотность распределения ограничена сверху, т.е. $f(x)\leq C$, где C – постоянная.

Процедура № 1 получения значений случайной величины графически показана на рисунке 4 и заключается в следующем:

а) получаем два независимых значения r_1 и r_2 случайной величины R с равномерным распределением на интервале $[0; 1)$;

б) строим точку с координатами (z_1, z_2) , где $z_1=a+r_1(b-a)$; $z_2=r_2c$;

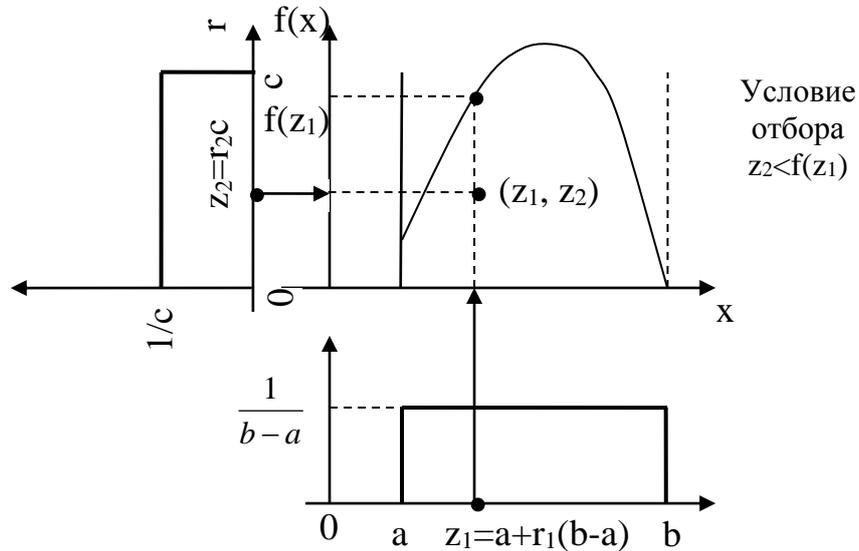


Рисунок 4 – Графическое изображение метода отбора (процедура №1)

в) если $z_2 < f(z_1)$, то полагаем, что случайная величина X приняла значение z_1 ;

если $z_2 \geq f(z_1)$, то точка $z=(z_1, z_2)$ отбрасывается, и вычисления повторяются с получением новой пары случайных чисел (r_n, r_{n+1}) по пункту а.

Эффективностью метода отбора называют вероятность того, что точка $z=(z_1, z_2)$ будет использована для расчета X . В рассмотренной процедуре № 1 эффективность метода характеризуется отношением площади, ограниченной кривой $f(x)$, осью x , прямыми $x=a$ и $x=b$, к площади прямоугольника $C(b-a)$.

2.2.2 Метод отбора – процедура № 2

Пусть случайная величина X имеет распределение с плотностью $f(x)$, которую можно представить в виде

$$f(x) = a_1 f_1(x) g_1(x), \quad (6)$$

где a_1 – постоянная;

$f(x)$ – некоторая известная плотность вероятности, а функция $g_1(x)$ удовлетворяет условию $0 \leq g_1(x) \leq 1$.

Значения случайной величины X можно получить по следующему алгоритму (см. рисунок 5):

- а) моделируем случайную величину Y с плотностью распределения $f_1(x)$;
- б) вычисляем $g_1(Y)$;
- в) моделируем случайную величину R , равномерно распределенную на интервале $[0; 1)$;
- г) если $R < g_1(Y)$, то принимаем $x=Y$. В противном случае полученные числа отклоняются и вычисления повторяются с пункта (а).

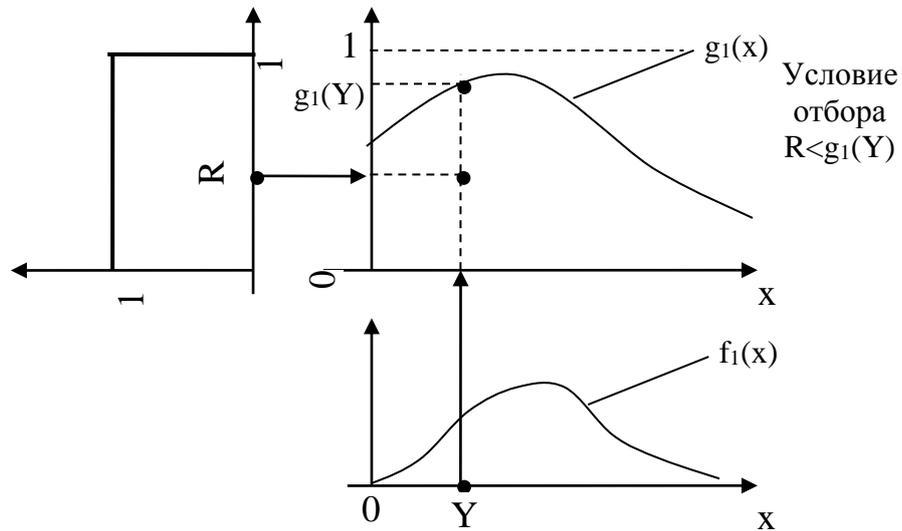


Рисунок 5 – Графическое изображение метода отбора (процедура №2)

2.2.3 Метод отбора – процедура № 3

Если в (б) функция $g_1(x)$ является монотонно неубывающей, то получается следующий алгоритм:

- а) моделируем случайную величину Y с плотностью $f_1(x)$;
- б) моделируем случайную величину γ с функцией распределения $g_1(\gamma)$;
- в) если $\gamma < Y$, то принимаем $x=Y$, в противном случае полученные числа отклоняем и вычисления повторяем с пункта (а).

2.2.4 Метод отбора – процедура № 4

Процедура № 4 определяет процесс моделирования случайной величины X , если плотность $f(x)$ можно представить в виде

$$f(x) = a_1(1 - g_1(x))f_1(x), \tag{7}$$

где $g_1(x) = g(t(x))$, причем g_1 - монотонно неубывающая функция, а g - некоторая известная функция распределения.

Значения случайной величины X тогда можно получить по следующему алгоритму:

- а) моделируем случайную величину Y с плотностью распределения $f_1(x)$;
- б) вычисляем величину $t(Y)$;
- в) моделируем случайную величину γ с функцией распределения g ;
- г) если $\gamma > t(Y)$, то принимаем $x=Y$. В противном случае полученные числа отклоняются и повторяются вычисления с пункта (а).

Пример 6. Для усеченного нормального распределения ($x > 0$) и с учетом аппроксимации (3) можно плотность распределения представить в виде

$$f(x) = \sqrt{\frac{2e}{\pi}} \left[1 - \left(1 - e^{-\frac{(x-1)^2}{2}} \right) \right] e^{-x}, \quad (8)$$

Введя обозначения:

$$a_1 = \sqrt{\frac{2e}{\pi}}; g_1(x) = 1 - e^{-\frac{(x-1)^2}{2}}; f_1(x) = e^{-x};$$

$$t(x) = \frac{(x-1)^2}{2}; g(t) = 1 - e^{-t},$$

можно убедиться в выполнении разложения $f(x)$ в виде (7).

2.3 Метод суперпозиций

Данный метод широко используется для моделирования случайных величин и основан на следующем утверждении.

Пусть условная плотность случайной величины X при $Y=y$ равна $f(x,y)$, причем для случайного параметра известна функция распределения $F(y)$. Тогда плотность распределения X

$$f(x) = \int_{-\infty}^{\infty} f(x,y) dF(y). \quad (9)$$

Таким образом, если $f(x)$ представима в виде (9), то случайную величину X можно моделировать в два этапа:

- а) выбираем значение Y соответственно функции распределения $F(y)$;
- б) выбираем X соответственно плотности $f(x, y)$.

Если Y принимает только целые значения, то

$$f(x) = \sum_i p_i f_1(x) = \sum_i p_i f(x, i), \quad (10)$$

где $p_i = P(Y=i)$.

Пример 7. Пусть $f(x) = \sum_{n=0}^{\infty} a_n x^n, 0 \leq x \leq 1, a_n \geq 0$.

Преобразуем $f(x)$ следующим образом:

$$f(x) = \sum_{n=0}^{\infty} \frac{a_n}{n+1} (n+1)x^n = \sum_{n=0}^{\infty} p_n (n+1)x^n$$

В последнем выражении

$$p(Y=n)=p_n; f_n(x)=f(x, n)=(n+1)x^n, 0 \leq x \leq 1$$

Отсюда получаем алгоритм моделирования:

а) для распределения p_0, p_1, p_2, \dots выбираем Y ;

б) если $Y=n$, то $X=R^{1/(n+1)}$

2.4 Комбинация метода суперпозиций и метода отбора

Выражение (10) позволяет реализовать комбинированный метод, если каждую из плотностей $f_i(x)$ представить в виде:

$$f_i(x) = a_i \omega_i(x) g_i(x) \quad (11)$$

где a_i – постоянные ($\sum_{i=0}^{\infty} a_i < \infty$),

$\omega_i(x)$ – некоторые известные плотности вероятностей,

$g_i(x)$ – функция, которая для каждого x удовлетворяет условию $0 \leq g_i(x) \leq 1$.

В результате для получения последовательности случайных чисел может быть использован следующий алгоритм:

а) реализация случайной величины Y по дискретному распределению $p(Y=i)=p_i, i=0, 1, 2, \dots$;

б) реализация случайной величины X по плотности вероятности $\omega_i(x)$ с параметром i , полученным в пункте (а). Эта часть алгоритма совпадает с методом суперпозиций. Затем следует использование метода отбора на основе представления плотности $f_i(x)$ в виде формулы (11);

в) реализация случайной величины R , равномерно распределенной на интервале $[0; 1)$;

г) если $R < g_i(x)$, то считаем, что полученное в пункте (б) значение является одним из значений случайной величины X с плотностью $f(x)$. В противном случае повторяем все вычисления, начиная с пункта (а).

Комбинация метода суперпозиций и метода отбора позволяет реализовать случайную величину практически с любым законом распределения.

Пример 8. Рассмотрим способ генерирования случайных чисел с нормальным законом распределения $N(0;1)$, основанный на аппроксимации (3). Представим $f(x)$ в виде следующей суммы:

$$f(x) = p_1 \omega_1(x) g_1(x) + p_2 \omega_2(x) g_2(x). \quad (12)$$

Пусть

$$\left\{ \begin{array}{l} p_1 = \sqrt{\frac{2}{\pi}}; \\ \omega_1(x) = \begin{cases} 1, & \text{при } 0 \leq x \leq 1; \\ 0, & \text{при } x < 0, x > 1; \end{cases} \\ g_1(x) = e^{-\frac{x^2}{2}}; \end{array} \right. \quad \left\{ \begin{array}{l} p_2 = \frac{1}{\sqrt{2\pi}}; \\ \omega_2(x) = \begin{cases} 2e^{-2(x-1)}, & \text{при } x > 1; \\ 0, & \text{при } x \leq 1; \end{cases} \\ g_2(x) = e^{-\frac{(x-2)^2}{2}}; \end{array} \right.$$

Для выполнения в (12) условия $\sum p_i = 1$ пронормируем p_1 и p_2 . Вследствие этого в пункте (а) алгоритма реализация случайной величины Y будет выполняться с вероятностью $p_1/(p_1 + p_2) = 2/3$ по плотности $\omega_1(x)$. Причем при $R > g_1(Y)$ полученное значение случайной величины Y будет отбрасываться. С вероятностью $p_2/(p_1 + p_2) = 1/3$ реализация случайной величины Y будет выполняться по плотности $\omega_2(x)$. Исключению будут подвергаться те выработанные числа, которые удовлетворяют условию $R > g_2(Y)$. Реализуя случайную величину S , принимающую значения “+1” и “-1” с вероятностями 0.5, и вычисляя $S \cdot X$, получают в итоге случайные числа, распределенные по $N(0;1)$, $-\infty < X < \infty$.

2.5 Моделирование некоторых специальных распределений

Для некоторых часто встречающихся в приложениях распределений разработаны частные алгоритмы и приемы, которые, тем не менее, не исключают возможности использования общих методов. Решение каждой конкретной задачи требует индивидуального подхода к выбору алгоритма моделирования на ЭВМ. Одним из показателей при этом является эффективность метода, под которой понимают отношение числа полученных случайных чисел требуемого распределения к числу применяемых для этого равномерно распределенных случайных чисел.

2.5.1 Биномиальное распределение

Биномиальным распределением является распределение вероятностей появления m числа событий в n независимых испытаниях, в каждом из которых вероятность появления события постоянна и равна P . Вероятность возможного числа появления события вычисляется по формуле Бернулли

$$p_n(x = m) = C_n^m p^m (1-p)^{n-m}, \quad (13)$$

где $C_n^m = \frac{n!}{m!(n-m)!}$ - число сочетаний из n элементов по m элементов.

Моделирующий алгоритм основан на представлении случайной величины X , подчиненной биномиальному закону распределения, в виде суммы

n независимых случайных величин X_i , каждая из которых имеет распределение

$$\begin{array}{c|cc} X_i & 1 & 0 \\ \hline p_i & p & q=1-p \end{array} \quad (14)$$

Процедура получения значений случайной величины X с распределением вероятностей (13) заключается в следующем:

а) реализуется случайная величина R ;

б) для каждого члена последовательности r_1, r_2, \dots, r_n проверяется выполнение неравенства $r_i < p$ ($i=1, 2, \dots, n$). Если неравенство выполняется, принимается $X_i=1$, в противном случае $X_i=0$.

в) вычисляется сумма $X_1 + X_2 + \dots + X_n$ значений n случайных величин X_i , которая принимается за значение случайной величины $X=S$.

При повторении этой процедуры k раз получаем последовательность значений s_1, s_2, \dots, s_k , которые являются реализацией биномиально распределенной случайной величины.

2.5.2 Распределение Пуассона

Пуассоновское распределение – это распределение случайной величины, которая равна числу событий, происшедших в единицу времени.

Случайная величина X имеет распределение Пуассона, если

$$p(X = m) = \frac{\lambda^m e^{-\lambda}}{m!}, \quad m = 0, 1, 2, \dots, \quad (15)$$

где λ - параметр распределения.

Моделирующий алгоритм основывается на следующем утверждении: если случайные величины X_1, X_2, \dots независимы и все имеют экспоненциальное распределение с математическим ожиданием, равным 1, то неотрицательное целое число n , для которого выполняется неравенство

$$\sum_{i=1}^n x_i \leq \lambda < \sum_{i=1}^{n+1} x_i, \quad (16)$$

имеет распределение Пуассона с параметром λ .

В связи с тем, что $X_i = -\ln r_i$, где r_i - случайная величина R с равномерным распределением на $[0;1)$, условие (16) можно записать в виде

$$\prod_{i=1}^{n+1} r_i \leq e^{-\lambda} < \prod_{i=1}^n r_i, \quad (17)$$

где $\prod_{i=1}^n r_i$ – произведение всех r_i , таких, что значение i – целое и выполняется соотношение $i=1 \dots n$.

На основании (17) можно построить алгоритм получения случайной величины, распределенной по закону Пуассона (15) с параметром λ :

а) реализуются последовательности r_1, r_2, \dots, r_n независимых случайных величин, равномерно распределенных на $[0; 1)$;

б) вычисляются произведения $r_1, r_1r_2, r_1r_2r_3, \dots$ до тех пор, пока не выполнится условие

$$\prod_{i=1}^{n+1} r_i \leq e^{-\lambda} < \prod_{i=1}^n r_i.$$

В качестве значения случайной величины X принимается число n . Если неравенству удовлетворяет первое из равномерно распределенных чисел r_1 , то $X=0$.

Моделирование случайных чисел X , имеющих закон распределения Пуассона (15), может быть реализовано другими способами. С этой целью можно воспользоваться предельной теоремой Пуассона, в соответствии с которой, если p – вероятность наступления события A при одном испытании, то вероятность наступления m событий в N независимых испытаниях при $N \rightarrow \infty$, $p \rightarrow 0$, $Np = \lambda$ асимптотически равна $p(X=m)$.

Выберем достаточно большое N , такое, чтобы $p = \lambda/N < 1$, и будем проводить серии по N независимых испытаний, в каждом из которых событие A происходит с вероятностью p , и будем подсчитывать число случаев u_j фактического наступления события A в серии с номером j . Числа u_j будут приближенно следовать закону Пуассона, причем тем точнее, чем больше N . Практически N должно выбираться таким образом, чтобы $p = 0.1 + 0.2$.

2.5.3 Моделирование нормального распределения

Наиболее часто встречающимся видом распределения является *нормальное*. В связи с этим при моделировании различных явлений возникает потребность иметь в распределении последовательности случайных чисел, отвечающих нормальному закону распределения. Реализация случайной величины с нормальным распределением $N(0;1)$ с помощью классических методов продемонстрирована выше на примере аппроксимации функции (3). Кроме этих методов, разработаны специальные методы, позволяющие получать с большой скоростью достаточно длинные последовательности случайных чисел, отвечающих нормальному закону распределения. На первом этапе выполняют реализацию случайной величины с плотностью $N(0;1)$. При помощи линейного преобразования

$$y_i = \mu + \sigma x_i, \quad i=1, 2, \dots, \quad (18)$$

при любом μ и $\sigma > 0$ можно затем получить последовательность случайных чисел y_1, y_2, \dots , отвечающих распределению $N(\mu; \sigma^2)$ с математическим ожиданием $M(Y) = \mu$ и дисперсией $D(Y) = \sigma^2$.

Один из самых известных методов реализации нормально распределенной случайной величины при использовании ЭВМ основан на *центральной предельной теореме*, в соответствии с которой распределение суммы независимых случайных величин $X_i (i=1, 2, \dots, n)$ приближается к нормаль-

ному при неограниченном увеличении n , если выполняются следующие условия:

а) все величины имеют конечные математические ожидания и дисперсии;

б) ни одна из величин по своему значению резко не отличается от всех остальных.

Согласно этой теореме можно сконструировать алгоритм реализации случайной величины X на основе аппроксимации распределения $N(0;1)$ суммой независимых случайных величин R_1, R_2, \dots, R_n , равномерно распределенных на интервале $[0;1)$. Практика показывает, что при $n=12$ аппроксимация уже довольно удовлетворительна. В результате получаем формулу для вычисления нормально распределенной случайной величины

$$x = \sum_{i=1}^{12} R_i - 6. \quad (19)$$

В другом известном генераторе нормально распределенных случайных величин используется точный обратный метод Бокса и Малера, который дает хорошие результаты, легко программируется и достаточно быстро работает. По этому методу генерируется пара нормированных нормальных чисел ($\mu=0, \sigma=1$) из двух стандартных случайных чисел (R_1 и R_2 на интервале от 0 до 1) с помощью следующих уравнений:

$$\left. \begin{aligned} x_1 &= -2\ln R_1 \cos(2\pi R_2); \\ x_2 &= -2\ln R_1 \sin(2\pi R_2). \end{aligned} \right\} \quad (20)$$

Модификацией этого метода является процедура Марсальи и Брея, в соответствии с которой генерируются два случайных числа R_1 и R_2 . Далее, полагая $V_1 = -1+2R_1$ и $V_2 = -1+2R_2$, вычисляют $S = \sqrt{V_1^2 + V_2^2}$. При $S \geq 1$ начинают цикл снова. При $S < 1$ реализуют отбор, т.е.

$$\left. \begin{aligned} x_1 &= V_1 \sqrt{\frac{-2\ln S}{S}}; \\ x_2 &= V_2 \sqrt{\frac{-2\ln S}{S}}. \end{aligned} \right\} \quad (21)$$

2.5.4 Моделирование (бета-распределения)

Плотность бета-распределения на интервале $(0;1)$ определяется формулой

$$f_{p,m}(x) = \frac{x^{p-1}(1-x)^{m-1}}{B(p,m)}, \quad 0 < x < 1, \quad (22)$$

где $p > 0, m > 0$ – параметры распределения;

$$B(p,m) = \int_0^1 x^{p-1}(1-x)^{m-1} dx - \text{бета-функция.}$$

Если m – целое, а p – нецелое, формула для моделирования случайной величины $X_{p,m}$, имеющей бета-распределение с параметрами p и m , имеет вид

$$X_{p,m} = \prod_{k=1}^m R_k^{1/(p+k-1)} = \exp\left(\sum_{k=1}^m \frac{\ln R_k}{p+k-1}\right), \quad (23)$$

где R_1, R_2, \dots, R_m – случайные числа, равномерно распределенные на интервале $[0;1)$. Формулу (23) можно использовать и в случае, когда m – нецелое, а p – целое, сделав предварительно замену переменных вида: $y=1-x$.

Для моделирования бета-распределения с нецелыми параметрами $p > 0$ и $m > 0$ можно использовать метод суперпозиций, для которого γ – целочисленная величина, имеющая дискретное распределение

$$p(\gamma=k) = p_k, \quad k=0, 1, 2, \dots,$$

$$\text{где } p_k = \frac{[m]!}{B(p,m)} \times \frac{a(a+1)\dots(a+k+1)}{k!(k+p)(k+p+1)\dots(k+p+[m])};$$

$[m]$ – обозначает целую часть m .

С учетом последнего распределения получаем формулу для бета-распределения:

$$X_{p,m} = \exp\left(\sum_{k=1}^{[m]+1} \frac{\ln R_k}{\gamma+p+k-1}\right). \quad (24)$$

Австрийский математик Йонк предложил следующий алгоритм моделирования бета-распределения:

а) выбираются значения R_1 и R_2 , равномерно распределенные на интервале $[0, 1)$;

б) если $R_1^{1/p} + R_2^{1/m} \geq 1$, то повторяется пункт (а) и т.д., иначе

$$X_{p,m} = \frac{R_1^{1/p}}{R_1^{1/p} + R_2^{1/m}}.$$

Метод Йонка основан на отборе (исключении).

2.5.5 Моделирование гамма-распределения

Одним из наиболее полезных видов непрерывных распределений, которым может воспользоваться исследователь при имитационном моделировании, является гамма-распределение

$$f_\gamma(x) = \frac{x^{y-1} e^{-x}}{\Gamma(y)}, \quad x > 0, \quad (25)$$

где $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ – гамма-функция.

Гамма-функция и бета-функция связаны между собой выражением

$$B(p,m) = \frac{\Gamma(p)\Gamma(m)}{\Gamma(p+m)};$$

в частном случае $\Gamma(m+1) = m!$

Правило моделирования гамма-распределения для $\gamma=n>0$ (γ – целое) имеет вид:

$$X = - \sum_{k=1}^n \ln R_k = -\ln \left(\prod_{k=1}^n R_k \right). \quad (26)$$

2.6 Первичная обработка статистических данных

2.6.1 Выборка. Эмпирическая функция распределения. Гистограмма

В математической статистике имеют дело со стохастическими экспериментами, состоящими в проведении n повторных независимых наблюдений над некоторой случайной величиной $X=\{x_i\}=x_1, x_2, \dots, x_n$, имеющей неизвестное распределение вероятностей, т.е. неизвестную функцию распределения $F_x(x) = F(x)$.

В этом случае множество \mathbf{X} возможных значений наблюдаемой случайной величины X называют *генеральной совокупностью*, имеющей функцию распределения $F(x)$.

Числа x_1, \dots, x_n , $x_i \in \mathbf{X}$, $i = \overline{1, n}$, являющиеся результатом n независимых наблюдений над случайной величиной X , называют *выборкой* из генеральной совокупности или *выборочными* (статистическими) данными. Число n называется *объемом* выборки.

В таблице 1 приведены обозначения параметров выборки для выборочных значений.

Таблица 1 – Параметры выборки

Параметр	Обозначение	Определение
Выборочные данные	x_i , где $i = 1, \dots, n$	Наблюдённые значения случайной величины
Объём выборки	n	Количество случайных чисел в выборке

Выборка является исходной информацией для статистического анализа и принятия решений о неизвестных вероятностных характеристиках наблюдаемой случайной величины X . Однако на основе конкретной выборки обосновать качество статистических выводов невозможно. Для этих целей на выборку следует смотреть априори как на случайный вектор (X_1, \dots, X_n) , координаты которого являются независимыми, распределёнными так же как и X , случайными величинами, и который еще не принял конкретного значения в результате эксперимента. Существует несколько способов представления статистических данных. Простейший из них – в виде статистического ряда:

Номер наблюдения	i	1	2	...	n
Результат наблюдения	x_i	x_1	x_2	...	x_n

Если среди выборочных значений имеются совпадающие, то статистический ряд удобнее записывать в виде следующей таблицы 2.

Таблица 2 – Статистический ряд

Выборочные значения	y_i	y_1	y_2	...	y_r
Частоты	m_i	m_1	m_2	...	m_r
Относительные частоты	$p_i^* = m_i/n$	m_1/n	m_2/n	...	m_r/n

Здесь (y_1, \dots, y_r) ($r < n$) – различные значения среди x_1, \dots, x_n ; m_i – частота значения y_i , $p_i^* = m_i/n$ – относительная частота значения y_i . Очевидно, что

$$\sum_{i=1}^r m_i = n, \quad \sum_{i=1}^r p_i^* = 1.$$

Совокупность пар (y_i, p_i^*) , $i = \overline{1, r}$ называют иногда *эмпирическим законом распределения*, а саму таблицу 2 – *таблицей частот*. Выборочные значения x_1, \dots, x_n , упорядоченные по возрастанию, носят название *вариационного ряда*:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

где $x_{(1)} = \min\{x_1, \dots, x_n\}$, $x_{(n)} = \max\{x_1, \dots, x_n\}$.

Величина $R = x_{(n)} - x_{(1)}$ называется *размахом выборки*.

Эмпирической функцией распределения, соответствующей выборке x_1, \dots, x_n , называется функция

$$F_n^* = 1/n \sum_{i=1}^n I(x_i < x) = \frac{1}{n} v_n(x),$$

где $I(A)$ – индикатор множества A , а $v_n(x)$ – число выборочных значений, не превосходящих x . Для каждой выборки x_1, \dots, x_n функция $F_n^*(x)$ является неубывающей и непрерывной слева. Ее график имеет ступенчатый вид:

– если все значения x_1, \dots, x_n различны, то $F_n^*(x) = i/n$ при $x \in [x_{(i)}, x_{(i+1)})$, $x_{(0)} = -\infty$, $x_{(n+1)} = \infty$;

– если y_1, \dots, y_r – различные значения среди x_1, \dots, x_n , то $F_n^*(x) = 1/n \sum_{i: y_i < x} m_i$.

Эмпирическая функция распределения $F_n^*(x)$ служит статистическим аналогом (оценкой) неизвестной функции распределения $F(x)$, которую называют при этом **теоретической**. Если x_1, \dots, x_n – выборка объема n из генеральной совокупности, имеющей непрерывное распределение с неизвестной плотностью вероятностей $f_x(x) = f(x)$, то для получения статистического аналога $f(x)$ следует произвести группировку данных. Она состоит в следующем:

1. По данной выборке x_1, \dots, x_n строят вариационный ряд $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.

2. Промежуток $[x_{(1)}, x_{(n)}]$ разбивают точками $u_0 = x_{(1)}, u_1, \dots, u_L = x_{(n)}$: $u_0 < u_1 < \dots < u_L$ на L непересекающихся интервалов $J_k = [u_{k-1}, u_k)$ (на практике $L \ll n$).
3. Подсчитывают частоты ν_k попадания выборочных значений в k -ый интервал J_k .
4. Полученную информацию заносят в таблицу, которую называют *интервальным статистическим рядом* (таблица 3).

Таблица 3 – Интервальный статистический ряд

Интервалы	J_k	$[u_0, u_1)$	$[u_1, u_2)$...	$[u_{L-1}, u_L]$
Частоты	ν_k	ν_1	ν_2	...	ν_L
Относительные частоты	$\tilde{p}_k^* = \nu_k / n$	ν_1/n	ν_2/n	...	ν_L/n

Очевидно, что $\sum_{k=1}^L \nu_k = n$, $\sum_{k=1}^L p_k^* = 1$. Поэтому совокупность пар

$$(\tilde{u}_k, \tilde{p}_k^*) \quad \text{с} \quad \tilde{u}_k = \frac{1}{2}(u_k + u_{k-1}), \quad k = \overline{1, L}$$

называют иногда эмпирическим законом распределения, полученным по сгруппированным данным. Далее в прямоугольной системе координат на каждом интервале J_k как на основании длины $\Delta u_k = u_k - u_{k-1}$ строят прямоугольник с высотой $h_k = \frac{\nu_k}{n \Delta u_k}$, $k = \overline{1, L}$. Получаемую при этом ступенчатую фигуру

называют *гистограммой*. Поскольку при больших n выполняется $\frac{\nu_k}{n \Delta u_k} \approx f(u_k)$, то верхнюю границу гистограммы можно рассматривать как оценку неизвестной плотности $f(x)$.

Ломаная с вершинами в точках (\tilde{u}_k, h_k) называется *полигоном частот* и для гладких плотностей является более точной оценкой, чем гистограмма.

На практике при группировке данных обычно берут интервалы одинаковой длины $\Delta u = \text{const}$, а число интервалов группировки определяют с помощью так называемого **правила Стаджерса**, согласно которому полагается

$$L = [1 + 3,32 \ln(n)] + 1,$$

или следующими рекомендациями:

при $n \geq 1000$ $L = 11 \dots 15$;

$n \geq 400$ $L = 10$;

$n \geq 200$ $L = 9$;

$100 < n < 200$ критерий применяют в исключительных случаях с числом интервалов $L = 7 \dots 9$.

Если интервалы выбраны одинаковой длины, то ширина их равна $\frac{x_n - x_1}{L}$.

Располагая только сгруппированными данными, можно определить аналог эмпирической функции распределения следующим образом:

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{k: \tilde{u}_k < x} v_k .$$

Статистическим аналогом (оценкой) теоретической числовой характеристики

$$g = Mg(x) = \int_{-\infty}^{\infty} g(x) dF(x) .$$

является *выборочная (эмпирическая) числовая характеристика* g^* , определяемая как среднее арифметическое значений функции $g(x)$ для элементов выборки x_1, \dots, x_n :

$$g^* = \int_{-\infty}^{\infty} g(x) dF_n^*(x) = \frac{1}{n} \sum_{i=1}^n g(x_i) .$$

В частности, k -й выборочный момент есть величина

$$\alpha_k^* = \frac{1}{n} \sum_{i=1}^n x_{ik} .$$

При $k = 1$ величину α_1^* называют выборочным средним и обозначают \bar{x} :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

При $k = 2$ величину μ_2^* называют *выборочной дисперсией* и обозначают s^2 :

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 .$$

Между выборочными начальными и выборочными центральными моментами сохраняются те же соотношения, что и между теоретическими. Например, справедливо равенство

$$s^2 = \alpha_2^* - (\bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 ,$$

являющееся аналогом известного равенства $\mu_2 = DX = \alpha_2 - \alpha_1^2 = M\{X\}^2 - (M\{X\})^2$. Для вычисления выборочных моментов k -го порядка по сгруппированным данным используются формулы:

$$\tilde{\alpha}_k^* = \frac{1}{n} \sum_{i=1}^N \tilde{u}_i^k v_i, \quad \tilde{\mu}_k^* = \frac{1}{n} \sum_{i=1}^N (\tilde{u}_i - \alpha_1^*)^k v_i .$$

В частности, выборочное среднее и выборочная дисперсия по сгруппированным данным определяются с помощью формул

$$\tilde{x} = \frac{1}{n} \sum_{i=1}^N \tilde{u}_i v_i, \quad \tilde{s}^2 = \frac{1}{n} \sum_{i=1}^N (\tilde{u}_i - \tilde{x})^2 v_i .$$

2.6.2 Проверка статистических гипотез

Статистической гипотезой называют любое утверждение о виде или свойствах наблюдаемых в эксперименте случайных величин. Правило, поз-

воляющее по имеющимся статистическим данным (выборке) принять или отклонить выдвинутую гипотезу, называется *статистическим критерием*.

Если формулируется только одна гипотеза H_0 и требуется проверить, согласуются ли статистические данные с этой гипотезой или же они ее опровергают, то критерии, используемые для этого, называются *критериями согласия*.

Если гипотеза H_0 однозначно фиксирует закон распределения наблюдаемых случайных величин, то она называется *простой*, в противном случае – *сложной*.

Пусть относительно наблюдаемой случайной величины X сформулирована некоторая гипотеза H_0 ; x_1, \dots, x_n – выборка объема n , являющаяся реализацией случайного вектора (X_1, \dots, X_n) , координаты которого $X_i, i = \overline{1, n}$ независимы и распределены так же, как X .

Общий метод построения критерия согласия для проверки гипотезы H_0 состоит в следующем. Вначале ищут статистику $T = T(X_1, \dots, X_n)$ (случайную величину!), характеризующую отклонение эмпирического распределения от теоретического, распределение которой в случае справедливости H_0 можно определить (точно или приближенно). Далее задают некоторое положительное малое число α , так что событие с вероятностью α можно считать практически невозможным в данном эксперименте. Затем для заданного α определяют подмножество K_α в множестве $K = \{t: t = T(x_1, \dots, x_n)\}$ возможных значений статистики T , так чтобы $P\{T\{X_1, \dots, X_n\} \in K_\alpha / H_0\} \leq \alpha$. Критерий согласия имеет следующий вид:

- если $t = T(x_1, \dots, x_n)$ – значение статистики $T(X_1, \dots, X_n)$, соответствующее данной выборке x_1, \dots, x_n и $t \in K_\alpha$, то гипотеза H_0 отвергается;
- если $t \notin K_\alpha$, то гипотеза H_0 принимается.

Статистика $T = T(X_1, \dots, X_n)$ называется статистикой критерия; множество K_α – критической областью для гипотезы H_0 , число α – уровнем значимости критерия.

2.6.2.1 Проверка гипотезы о виде распределения

Пусть x_1, \dots, x_n – выборка объема n , представляющая собой результат n независимых наблюдений над случайной величиной X , относительно распределения которой выдвинута простая гипотеза $H_0: F_X(x) = F(x)$. ($F(x)$ – теоретическая функция распределения, соответствующая гипотезе H_0).

Наиболее распространенным критерием проверки этой гипотезы H_0 является критерий χ^2 **Пирсона**. Чтобы воспользоваться критерием χ^2 Пирсона, выборочные данные x_1, \dots, x_n следует предварительно сгруппировать, представив их в виде интервального статистического ряда (см. таблицу 3).

Пусть $J_k = [u_k, u_{k+1}), k = \overline{1, L}$ – интервалы группировки; v_1, \dots, v_L – частоты попадания выборочных значений в интервалы J_1, \dots, J_L соответственно ($v_1 + \dots + v_L = n$). Обозначим p_k теоретическую (соответствующую H_0) веро-

ятность попадания случайной величины X в интервал $J_k : p_k = P\{u_k \leq X < u_{k+1}\} = F(u_{k+1}) - F(u_k), k = \overline{1, L}$, где $F(u_{k+1}), F(u_k)$ – значение теоретической функции распределения соответственно на правой и левой границах k -ого интервала гистограммы, построенной по таблице 3.

При расчетах принимают $F(u_1)=0, F(u_{L+1})=1$.

Статистикой критерия χ^2 является величина
$$\chi_n^2 = \sum_{k=1}^L \frac{(v_k - np_k)^2}{np_k},$$

где L – количество интервалов гистограммы, построенной по таблице 3;
 v_k – количество реализаций СВ, попавших в k -й интервал;
 p_k – вероятность попадания случайной величины в k -й интервал, вычисленная для теоретического закона распределения;
 n – объем выборки (количество случайных чисел в выборке).

Она характеризует отклонение эмпирической функции распределения $F_n^*(x)$ (v_k/n – приращение $F_n^*(x)$ на интервале J_k) от теоретической функции распределения $F(x)$ (p_k – приращение $F(x)$ на том же интервале J_k). Поскольку относительные частоты v_k/n сближаются с вероятностями p_k при $n \rightarrow \infty, k = \overline{1, L}$, то в случае справедливости H_0 значение величины χ_n^2 не должно существенно отличаться от нуля. Поэтому критическая область критерия χ^2 задается в виде $K_\alpha = \{t \geq t_\alpha\}$, где $t = \chi_n^2(x_1, \dots, x_n)$ – значение величины χ_n^2 , вычисленное для заданной выборки, а порог t_α определяется по заданному уровню значимости α так, чтобы $P\{\chi_n^2 \in K_\alpha / H_0\} = \alpha$. Нахождение t_α основано на том факте (известном как теорема Пирсона), что случайная величина χ_n^2 имеет при $n \rightarrow \infty$ предельное распределение хи-квадрат с $(L-1)$ степенью свободы $\chi^2(L-1)$.

На практике предельное распределение $\chi^2(L-1)$ можно использовать с хорошим приближением при $n \geq 50$ и $v_k \geq 5, k = \overline{1, L}$. При выполнении этих условий для заданного уровня значимости α можно положить $t_\alpha = \chi_{1-\alpha, L-1}^2$, где $\chi_{1-\alpha, L-1}^2$ – $(1-\alpha)$ -квантиль распределения $\chi^2(L-1)$.

Таким образом, критерий согласия χ^2 Пирсона состоит в следующем:

1. По таблице 3 строят интервальный статистический ряд.
2. Строится гистограмма.
3. По виду гистограммы формулируется гипотеза о виде закона распределения.
4. Вычисляются теоретические вероятности попадания случайной величины в каждый из интервалов гистограммы по формуле $p_k = F(u_{k+1}) - F(u_k), k = \overline{1, L}$.
5. Вычисляется значение статистики $\chi_n^2(x_1, \dots, x_n) = t$.
6. По таблице распределения для вычисленного значения χ_n^2 и числа степеней свободы $n-s=L-k-1$, где k – количество параметров теоретического закона распределения (для экспоненциального равно 1, для

нормального и Вейбулла – 2), по заданному уровню значимости σ находится по табл. П4 порог $\chi^2_{1-\alpha, L-1}$.

7. Если $t \geq \chi^2_{1-\alpha, L-1}$, то гипотезу H_0 отвергают.

8. Если $t < \chi^2_{1-\alpha, L-1}$, то гипотезу H_0 принимают.

Если случайная величина X дискретная, $x_k, k = \overline{1, L}$ – различные выборочные значения, а $P\{X=x_k\}=p_k$ в случае справедливости H_0 , то всегда можно определить L интервалов, содержащих ровно по одному выборочному значению. Поэтому в данном случае можно сразу считать, что $\nu_k=m_k, k = \overline{1, L}$, где m_k – частота выборочного значения x_k .

3 Порядок выполнения работы

а) Получить явную формулу для моделирования случайной величины с законом распределения, заданным в таблице 4. Если необходимо, найти неизвестные параметры.

б) Моделирование случайной величины $X = \{x_i\} = x_1, x_2, \dots, x_n$ выполнить с помощью последовательности чисел $R = \{r_i\} = r_1, r_2, \dots, r_n$, полученных указанным в таблице 4 способом построения случайных чисел.

в) Вычислить критерий χ^2 Пирсона при $k=16$ или $k=21$ для последовательности случайных чисел $X = \{x_i\} = x_1, x_2, \dots, x_n$ и сделать заключение о соответствии смоделированной величины данному закону распределения.

г) Вывести на экран заполненную *интервальным статистическим рядом* таблицу (см. таблицу 3); гистограмму, построенную по таблице 3; теоретический и практический графики случайных величин, моделируемых по заданному закону распределения; значение критерия χ^2 Пирсона.

Таблица 4 – Варианты заданий

Вариант	Закон распределения	Способ построения
1	Экспоненциальный, $\lambda=2$	Метод обратных функций, формула (2)
2	Нормальный, $N(0;1)$	Метод отбора
3	Экспоненциальный, $\lambda=3$	Метод отбора
4	Распределение Пуассона, формула (15); $\lambda=9$	Алгоритм - формула (17)
5	Нормальный, $(0; 0.81)$, формула (18)	Метод отбора
6	Нормальный, $N(0; 1)$	Аппроксимация распределения; формула (19)
7	Гамма-распределение; формула (25); $\lambda=3$	Формула (26)
8	Логарифмическо-нормальный, $f_y(y) = \frac{1}{y^{\sigma_Y} \sqrt{2\pi}} \times \exp\left[-\frac{(\ln y - m_Y)^2}{2\sigma_Y^2}\right],$ $\sigma_Y = 1; m_Y = 1; y > 0$	Метод отбора
9	Закон Вейбулла, $f(y) = \frac{a}{\sigma^a} y^{\lambda-1} e^{-\left(\frac{y}{\sigma}\right)^a},$ $y \geq 0, a = 3, \sigma = 2$	Метод отбора
10	Бета-распределение, формула (22); $m=2, p=1.5$	Формула (23)

| 11 | Нормальный, $N(0; 1)$ | Аппроксимация; формула (3) |

Продолжение таблицы 4

12	Бета-распределение, формула (22); $m=5, p=2.3$ Дискретное распределение; $k=10$;	Метод Йонка
13	$p_1=0.5; p_5=0.04; p_8=0.95;$ $p_2=0.8; p_6=0.09; p_9=0.97;$ $p_3=0.7; p_7=0.3; p_{10}=0.35.$ $p_4=0.2$;	Формула (5)
14	Усеченное нормальное распределение, формула (8)	Процедура 4
15	Распределение Вейбулла (см. вариант 9); $a=2.5; \sigma=2; g_1(x)=x^{a-1}$	Процедура 3
16	Нормальный, $N(0; 1)$	Комбинация метода суперпозиции и метода отбора, (12)
17	Биномиальный, формула (13); $p=0.3$	См. п. 5.1
18	Закон Вейбулла (см. вариант 9); $a=2; \sigma=3$ Закон Релея:	Процедура 2
19	$f(y) = \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right),$ $y \geq 0; \sigma^2 = 0.5$	Процедура 1
20	Экспоненциальный; $\lambda=0.4$	Метод обратных функций, формула (2)
21	Нормальный; $N(0; 1)$ Дискретное распределение; $k=12$;	Процедура Марсальи и Брея, формула (21)
22	$p_1=0.3; p_5=0.39; p_9=0.9;$ $p_2=0.2; p_6=0.8; p_{10}=0.3;$ $p_3=0.4; p_7=0.89; p_{11}=0.05;$ $p_4=0.38; p_8=0.7; p_{12}=0.03.$	Формула (5)
23	Гамма-распределение; формула (25); $\lambda=4$	Формула (26)
24	Нормальный, $N(2, 1)$	Аппроксимация распределения; формула (19)

4 Содержание отчета

1. Титульный лист.
2. Постановка задачи (вариант, задание).
3. Ход работы:
 - Математическая модель (вывод формулы для моделирования заданной случайной величины).
 - Полученные результаты (таблицы, графики).
 - Описание листинга программы.
4. Отчет распечатать.

Лабораторная работа № 3

Планирование машинных экспериментов

1 Цель работы

Изучение и практическое освоение методов тактического планирования машинных экспериментов, составление алгоритмов определения объема выборки машинного эксперимента с моделями сложных систем, решение задачи обеспечения точности и достоверности результатов моделирования.

2 Методические указания

После того как установлены цели эксперимента и принято решение использовать машинное имитационное моделирование, полезно провести предварительное планирование предстоящего эксперимента. Исследователь должен иметь подробный план эксперимента для целенаправленного и эффективного получения требуемых данных. Плановые ограничения на время и затраты должны быть приведены в соответствие с имеющимися в распоряжении исследователя ресурсами. Чем дороже и сложнее эксперимент, тем большее внимание следует обращать на этот этап.

Обычно используются три типа экспериментов: сравнение средних и дисперсий различных альтернатив; определение важности учета или значимости влияния переменных и ограничений, наложенных на эти переменные; отыскание оптимальных значений на некотором множестве возможных значений переменных.

При ограниченном ресурсе времени моделирования первейшей задачей исследования является получение ответа на вопрос: как много выборочных значений следует взять во время моделирования, чтобы обеспечить достаточную статистическую значимость? Поскольку разбросы выборочных значений случайны, то обусловленная ими некоторая неточность результата эксперимента в значительной мере определяется размером выборки. Задача определения такого размера выборки, который позволяет обеспечить желаемый уровень точности и в то же время минимальную стоимость моделирования, является чрезвычайно важной.

2.1 Начальные условия и их влияние на достижение установившегося режима

Первая проблема при проведении машинного эксперимента заключается в максимальном исключении влияния начального периода работы машинной модели, что в значительной степени определяется начальными условиями запуска модели. При этом типичное поведение оцениваемого параметра \hat{p} при изменении объема N выборки показано на рисунке 1.

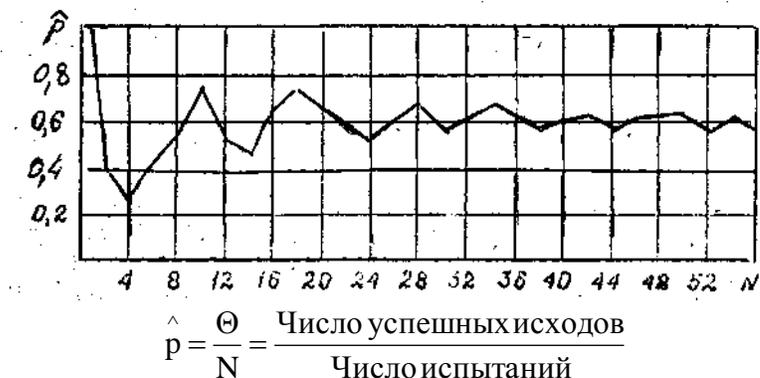


Рисунок 1 – Графическое изображение поведения оцениваемого параметра \hat{p}

Существуют три основные пути уменьшения возможного влияния начального периода на получаемые данные.

1 Использование достаточно длинных вычислительных прогонов, чтобы число данных переходного периода было незначительно по сравнению с числом данных установившегося режима.

2 Исключение из рассмотрения начального периода прогона.

3 Выбор для установившегося состояния такого начального условия, которое ближе к типичному, и тем самым уменьшение переходного периода.

2.2 Оценивание среднего значения выборочной совокупности

Задача состоит в получении оценки неизвестного математического ожидания μ некоторого исследуемого параметра по выборочной средней \bar{x} . Выборочную среднюю \bar{x} можно рассматривать как случайную величину \bar{X} (\bar{x} изменяется от выборки к выборке) и выборочные значения признака x_1, x_2, \dots, x_N – как одинаково распределенные независимые случайные величины X_1, X_2, \dots, X_N с математическим ожиданием μ и среднеквадратическим отклонением σ . Будем использовать предположение о независимости и нормальном распределении откликов модели. Это предположение основано на применении *центральной*

предельной теоремы теории вероятностей, сущность которой состоит в утверждении, что распределение случайной величины, являющейся суммой большого числа независимых величин с одинаковыми распределениями вероятностей, близко к нормальному распределению. В более поздних работах показано, что требования независимости и одинаковой распределенности не являются необходимыми. Часто бывает достаточно, чтобы отклик представлял собой сумму большого числа небольших эффектов. Поэтому переменная отклика сложной имитационной модели, являющаяся результатом действия большого числа случайных переменных, распределена приблизительно нормально. В результате этого, если случайная величина X распределена нормально, то выборочная средняя \bar{X} , найденная по независимым наблюдениям, также распределена нормально с математическим ожиданием

$$M(\bar{X}) = \mu$$

и среднеквадратическим отклонением

$$\sigma(\bar{X}) = \sigma / \sqrt{N}$$

Потребуем, чтобы выполнялось соотношение

$$P(\mu - d \leq \bar{x} \leq \mu + d) < \gamma, \quad (1)$$

где \bar{x} – выборочное значение средней;

γ – вероятность того, что интервал $\mu \pm d$ содержит \bar{x} (надежность, задаваемая исследователем);

d – точность оценки.

Задаваемая надежность γ однозначно связана с уровнем значимости α получаемой оценки \bar{x} формулой

$$\gamma = 1 - \alpha.$$

Если требуется оценить математическое ожидание μ наперед заданной точностью d и надежностью γ (уровнем значимости α), то минимальный объем выборки, который обеспечит эту точность, находим по формуле

$$N = \frac{t^2 \sigma^2}{d^2}, \quad (2)$$

где число t – определяется из равенства $2\Phi(t) = \gamma$ или $\Phi(t) = \gamma/2$ по таблице функции Лапласа.

Если неизвестны точность d и среднеквадратическое отклонение σ , то задачу можно поставить так: каков должен быть размер выборки, чтобы искомая оценка с вероятностью $\gamma = 1 - \alpha$ находилась в пределах $\mu \pm \Delta \sigma$, где $\Delta \sigma$ – некоторая доля σ (например, $\Delta \sigma = \sigma/2; \sigma/4; \sigma/6; \sigma/10; \sigma/20; \dots$).

В таких случаях следует определять дисперсию выхода с данного пробного эксперимента и получить оценку s^2 дисперсии, а затем вычислить полное число необходимых наблюдений. Тогда размер выборки N определится выражением

$$N = \frac{t^2 s^2}{d^2}, \quad (3)$$

где оценка s^2 дисперсии получена по формуле

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2;$$

n – размер выборки пробного эксперимента.

2.3 Применение теоремы Чебышева

Неравенство Чебышева говорит, что при заданном числе k (не меньшем единицы) и произвольной выборке x_1, x_2, \dots, x_N размера N по меньшей мере $1 - 1/k^2$ измерений находятся вблизи среднего значения на расстоянии не более k среднеквадратических отклонений. Это неравенство справедливо для любых распределений. Если не исходить из предположения о нормальном распределении выхода (которое не всегда выполняется с достаточной точностью), можно воспользоваться неравенством Чебышева, которое имеет вид

$$P\left\{|\bar{x} - \mu| > k\sigma\right\} \leq \frac{1}{k^2}. \quad (4)$$

Если задаться условием, при котором оценка \bar{X} должна находиться в интервале $\mu \pm \sigma/m$ ($m=2, 4, 6, 10, \dots$) с вероятностью $\gamma=1-\alpha$, т.е.

$$P\left\{|\bar{x} - \mu| > \frac{\sigma}{m}\right\} \leq \alpha, \quad (5)$$

тогда, используя уравнение (1), получаем

$$P\left\{|\bar{x} - \mu| > \frac{\sqrt{N}}{m} \cdot \frac{\sigma}{\sqrt{N}}\right\} \leq \alpha = \frac{m^2}{N}, \quad (6)$$

так как $k = \frac{\sqrt{N}}{m}$; $\frac{1}{k^2} = \frac{m^2}{N}$.

Отсюда находим требуемый размер выборки машинного эксперимента

$$N = \frac{m^2}{\alpha}. \quad (7)$$

Полученный размер выборки существенно больше того, который оказывается достаточным в случае нормального распределения совокупности. В таблице 1 приведены размеры выборки при различных значениях d при $\alpha=0.05$ ($\gamma=0.95$).

Таблица 1 – Размеры выборки

Отклонение	Необходимый объем выборки N	
	По центральной предельной теореме	По теореме Чебышева
$\sigma/2$	15	80
$\sigma/4$	61	320
$\sigma/6$	138	720
$\sigma/8$	246	1280
$\sigma/10$	384	2000
$\sigma/12$	553	2880
$\sigma/20$	1537	8000

2.4 Оценивание процентных отношений

Во многих случаях целью машинного эксперимента является получение оценки \hat{p} вероятности появления $p=P(A)$ некоторого события A , определяемого состояниями процесса функционирования исследуемой системы. В качестве оценки вероятности p выступает частость $\hat{p} = \frac{\Theta}{N}$, где θ – число положительных исходов.

Представим частость $\hat{p} = \frac{\Theta}{N}$ в виде

$$\hat{p} = \frac{1}{N} \sum_{i=1}^N x_i,$$

где частость в данной реализации из N наблюдений является случайной величиной, принимающей значения $x_i=1$ с вероятностью p и $x_i=0$ с дополнительной вероятностью $1-p$, тогда

$$M(\hat{p}) = p, D(\hat{p}) = \frac{p(1-p)}{N}.$$

По теореме Лапласа частость θ/N при достаточно больших N можно рассматривать как случайную величину, описываемую нормальным законом распределения вероятности с математическим ожиданием $M(\hat{p})$ и дисперсией $D(\hat{p})$. Поэтому точность ε и достоверность γ оценок связаны с количеством N реализаций равенством

$$P\left\{p - \varepsilon < \frac{\Theta}{N} < p + \varepsilon\right\} = \Phi\left(\frac{p + \varepsilon - p}{\sqrt{p(1-p)}} \sqrt{N}\right) - \Phi\left(\frac{p - \varepsilon - p}{\sqrt{p(1-p)}} \sqrt{N}\right) = \gamma,$$

из которого объем выборки, необходимый для получения оценки \hat{p} с точностью ε и достоверностью γ , определим в виде

$$N = \frac{t^2 p(1-p)}{\varepsilon^2}. \quad (9)$$

При тактическом планировании эксперимента, когда значение p неизвестно, проводят предварительное моделирование для произвольного значения p . По результатам предварительного моделирования $p_0 = \theta/n$, а затем по (9) вычисляют, используя вместо p значение p_0 , необходимое количество реализаций N . Такая процедура оценки N может выполняться несколько раз в ходе машинного эксперимента.

При отсутствии возможности получения каких-либо предварительных сведений о вероятности p использование абсолютной точности теряет смысл. В таких случаях целесообразно задавать относительную точность результатов моделирования.

$$\varepsilon_0 = \frac{\varepsilon}{p}.$$

Тогда соотношение (9) примет вид

$$N = \frac{t^2(1-p)}{\varepsilon_0^2 p}. \quad (10)$$

Соотношение (10) для объема выборки статистического моделирования редких событий, выражающееся в том, что для оценивания малых вероятностей p с высокой точностью необходимо очень большое число реализаций. В практических случаях для оценивания вероятностей порядка 10^{-k} целесообразно количество реализаций выбирать равным 10^{k+1} .

2.5 Оценивание дисперсии совокупности

В качестве показателя эффективности системы может выступать дисперсия σ^2 , оценку которой s^2 требуется получить в результате проведения машинных экспериментов с некоторой надежностью $\gamma=1-\alpha$.

Потребуем, чтобы выполнялось соотношение

$$P\{s^2-d \leq \sigma^2 \leq s^2+d\} = \gamma.$$

Преобразуем двойное неравенство

$$s^2-d \leq \sigma^2 \leq s^2+d$$

в равносильное

$$s^2(1-q) \leq \sigma^2 \leq s^2(1+q),$$

где $q=d/s^2$ – число, характеризующее степень близости оценки s^2 к истинной дисперсии σ^2 .

Для удобства пользования введем в рассмотрение χ^2 – статистику

$$\chi^2 = \frac{\sigma^2(N-1)}{\sigma^2},$$

которая распределена по закону χ^2 (хи-квадрат) с $N-1$ степенями свободы.

Плотность распределения χ^2 имеет вид

$$k_N(x) = \begin{cases} \frac{1}{2^{N/2} \Gamma\left(\frac{N}{2}\right)} x^{\frac{N}{2}-1} e^{-\frac{x}{2}}, & x > 0, N \geq 1; \\ 0, & x \leq 0 \end{cases},$$

где $\Gamma(\lambda) = \int_0^\lambda t^{\lambda-1} e^{-t} dt$ – гамма-функция; в частности, $\Gamma(n+1) = n!$.

Распределение нормированной величины $(\chi^2 - N)/\sqrt{2N}$ стремится к стандартному нормальному распределению

$$P\left\{\frac{\chi^2 - N}{\sqrt{2N}} < x\right\} \approx \Phi(x),$$

откуда можно получить уравнение

$$t^2 = \frac{q^2(N-1)}{2}$$

или

$$N = 1 + \frac{2t^2}{q^2},$$

где t – квантиль нормального распределения, значение которого находится из равенства $2\Phi(t)=\lambda$ по таблице функции Лапласа.

2.6 Сравнение двух распределений

При исследовании результатов машинного моделирования часто возникает задача проверки близости распределения отклика модели к некоторому другому распределению. Для проверки близости двух распределений можно воспользоваться критерием Колмогорова-Смирнова. Задаваясь требуемой точностью (абсолютной ошибкой) d при заданном уровне значимости α , можно вычислить необходимый объем выборки по формулам:

$$N = \left(\frac{1.63}{d}\right)^2, \alpha = 0.01; \quad (13)$$

$$N = \left(\frac{1.36}{d}\right)^2, \alpha = 0.05; \quad (14)$$

$$N = \left(\frac{1.22}{d}\right)^2, \alpha = 0.10; \quad (15)$$

2.7 Автокоррелированные данные

Если выборочные значения на выходе модели автокоррелированы, т.е. последующее выборочное значение зависит от предыдущих, то в такой выборке содержится меньше информации, чем в выборке из независимых данных. При этом необходимый размер выборки весьма чувствителен к величине автокорреляции.

Рассмотрим способ оценки автокорреляционной функции в использовании ее при построении оценок параметров, которые в случае коррелированных данных принимают вид:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i; \quad (16)$$

$$\sigma_N^2 = \frac{\sigma_x^2}{N} \left\{ 1 + 2 \sum_{p=1}^m \left(1 - \frac{p}{m+1} \right) \rho_{p,x} \right\}, \quad (17)$$

где σ_x^2 – дисперсия совокупности;

$\rho_{p,x}$ – p -й коэффициент автокорреляции;

m – максимальная глубина рассматриваемой автокорреляции, $p=1, 2, \dots, m$.

Для оценивания коэффициентов корреляции необходимо провести эксперимент с пробной выборкой и применить формулу

$$\rho_{p,x} = \frac{\sum_{i=1}^{N-p} \{(x_i - \bar{x})(x_{i+p} - \bar{x})\}}{s_x^2 (N-1)}; \quad (18)$$

где s_x^2 – дисперсия N измерений в пробном эксперименте.

Минимально необходимый размер выборки можно вычислить по формуле

$$N = \frac{t^2 s_x^2 \left\{ 1 + 2 \sum_{p=1}^m \left(1 - \frac{p}{m+1} \right) \rho_{p,x} \right\}}{(\bar{dx})^2},$$

где \bar{dx} – относительная погрешность истинного среднего.

Максимальную глубину m вычисляемых коэффициентов корреляции выбирают на практике равной примерно 10% от числа N измерений. При выполнении задания по данной лабораторной работе достаточно выбрать $m=5$.

2.8 Стохастические модели

Рассмотренные методы планирования машинных экспериментов применимы для случая стохастических моделей (хотя и могут быть использованы в других случаях, например, при численных алгоритмах решения систем уравнений вычисления определенных интегралов методами Монте-Карло).

Рассмотрим некоторые такие модели, применительно к которым составлены задания по данной лабораторной работе.

Случайное блуждание. Первоначальную формулировку задачи о “случайных блужданиях” предложил Пирсон в 1906 г. Если пьяный делает M шагов равной длины от фонарного столба в произвольных направлениях, то как далеко отойдет он от этого столба? Со времени такой формулировки случайного блуждания (в форме задачи о пьяном пешеходе) модели случайного блуждания получили широкое распространение в технике, физике, биологии и общественных науках (диффузия молекул в газе, броуновское движение коллоидных взвесей в жидкости, моделирование длинных полимерных цепочек, вычисление определенных интегралов, моделирование процессов принятия статистических решений и т.д.).

Одномерное случайное блуждание. Рассмотрим идеализированную одномерную задачу с движением пьяного (пешехода), который начинает двигаться от фонарного столба, расположенного в точке $x=0$ (см. рис. 2). Все шаги имеют одинаковую длину l . Направление каждого шага пешехода не зависит от направления предыдущего. Пешеход делает шаг вправо с вероятностью p и шаг влево с вероятностью $q=1-p$. Основной интересующей исследователя величиной

является вероятность $P_M(x)$ того, что после M шагов пешеход окажется на расстоянии x от фонарного столба.

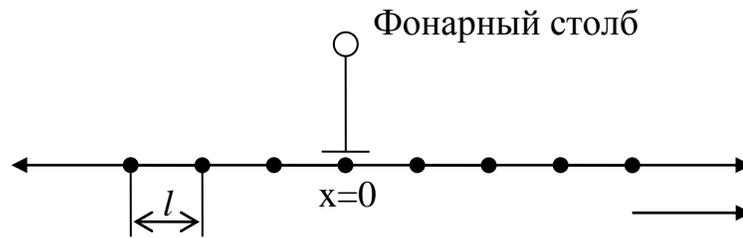


Рисунок 2 – Одномерное случайное блуждание

Исследование задачи одномерного случайного блуждания можно провести аналитически, воспользовавшись теорией вероятности. Полученные при этом результаты могут быть использованы в процессе создания имитационной модели и проверки ее правильности (верификация). Выборочное среднее \bar{x}_n и выборочную дисперсию σ_n^2 можно рассчитать с помощью следующих аналитических выражений:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x}_n)^2. \quad (20)$$

Двумерное случайное блуждание. В моделях случайного блуждания двумерный случай является наиболее важным, т.к. применяется не только к блужданию пьяных или движению молекул. При двумерном случайном блуждании на каждом шаге по времени осуществляется движение случайным образом равновероятно в одном из четырех возможных направлений: на север, юг, восток или запад (см. рис. 3). Движение начинается из начальной точки H , совпадающей с началом координат ($X=0; Y=0$) и заканчивается через M шагов (на рис. 3 показан возможный путь за $M=12$ шагов) в точке K .

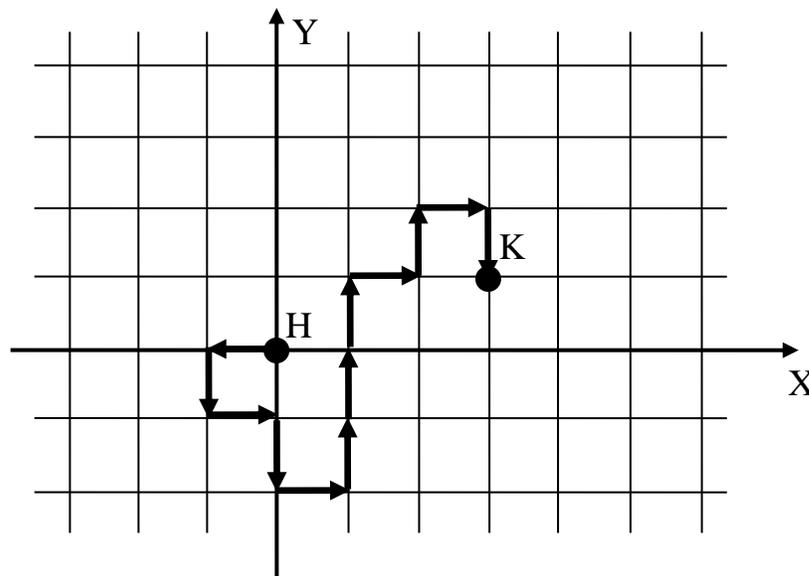


Рисунок 3 – Двумерное случайное блуждание

3 Порядок выполнения работы

а) Каждое задание предполагает разработку программной имитационной модели случайного блуждания, с помощью которой могут быть получены необходимые результаты по двум соседним вариантам.

б) В соответствии с первым вариантом требуется провести пробный эксперимент, результаты которого используются для планирования эксперимента по оценке среднего значения указанного параметра.

в) Во втором варианте выполняется планирование эксперимента для оценки с заданной точностью выборочной дисперсии.

г) В процессе проведения пробного эксперимента строится статистическое распределение исследуемого параметра и определяется целесообразность аппроксимации этого распределения нормальным законом (общая часть для двух вариантов).

1-2. Простое случайное блуждание с поглощающими экранами. Пусть заданы целые числа $a > 0$ и $b > 0$; блуждание прекращается, если $x = b$ или $x = -a$. Составить машинную модель для определения времени блуждания. Для проверки правильности модели вычислить оценку вероятности $\hat{p}(a, b)$ того, что блуждание прекратится возле правого экрана и сравнить с полученной вероятностью.

$$\hat{p}(a, b) = \frac{\left(\frac{q}{p}\right)^a - 1}{\left(\frac{q}{p}\right)^b - 1}. \quad (21)$$

При $q = p = 0.5$ имеем $\hat{p}(a, b) = \frac{a}{b + a}$

3-4. Модель падения дождевой капли. При воздействии случайных порывов резкого ветра падение дождевой капли можно моделировать случайным блужданием на квадратной решетке (см. рис. 4). Движение начинается из узла, расположенного на расстоянии $h = y$ над горизонтальной линией (поверхностью земли). Вероятность p_{\downarrow} шага “вниз” больше вероятности p_{\uparrow} шага “вверх”. Вероятности скачков целесообразно выбрать равными $p_{\downarrow} = 0.5$; $p_{\uparrow} = 0.1$; $p_{\leftarrow} = p_{\rightarrow} = 0.2$.

Определите время τ , за которое капля достигнет горизонтальную прямую (поверхность земли $y = 0$) и функциональную зависимость τ от h (4-6 значений).

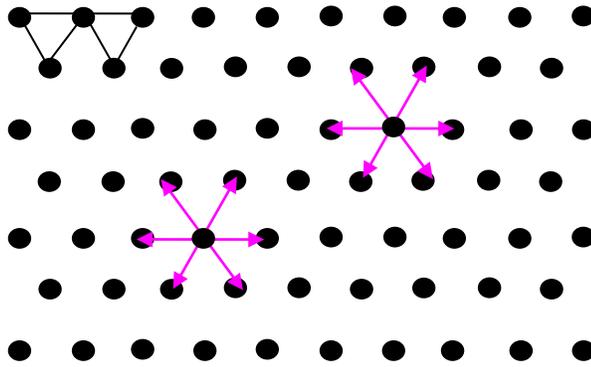


Рисунок 5 – Блуждание на треугольной решетке

11-12. Блуждание на сотах. Составьте имитационную модель случайного блуждания на сотах (см. рис. 6). На каждом шаге по времени пчела движется равновероятно в одном из трех направлений. На какое расстояние удаляется пчела за $M=8$ шагов. Сравните результаты с данными моделирования для заданий 5-6 и 9-10.

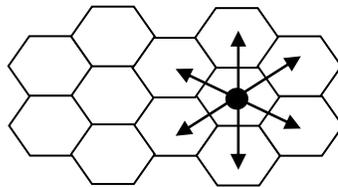


Рисунок 6 – Блуждание на сотах

13-14. Случайное блуждание на трехмерной решетке. Оцените расстояние, на которое удаляется частица, равновероятно блуждающая по трехмерной решетке. Число шагов блуждания $M=10$. Параллельно исследуемому процессу определите удаление от начального состояния отдельно по всем трем составляющим координатам.

15-16. Персистентное случайное блуждание. В персистентном случайном блуждании вероятность перехода или “скачка” зависит от последнего перехода. Рассмотрите одномерное случайное блуждание, в котором шаги совершаются только в ближайшие соседние узлы. Предположим, что сделано $k-1$ шагов. Далее k -й шаг делается в том же направлении с вероятностью α , а шаг в противоположном направлении делается с вероятностью $1-\alpha$. Определите удаление частицы от исходного положения за $M=8$ шагов при $\alpha=0.2$ и $\alpha=0.4$.

17-18. Гибкие полимерные цепочки. Исследование конфигураций полимеров можно проводить с использованием моделей случайных блужданий. Полимер состоит из очень большого числа N ($N \sim 10^3+10^5$) повторяющихся звеньев, или мономеров. Например, повторяющиеся звенья цепочки полиэтилена можно изобразить как $\dots - CH_2 - CH_2 - CH_2 - \dots$. Структура полимеров имеет случай-

ный характер (см. рис. 7) и определяет, свойства вещества, получаемого при смешении полимеров. Одной из фундаментальных (основных) геометрических характеристик при этом является расстояние R_N между концами цепочки, где N – количество мономеров. Определите характеристику R_N для $N=500$, предполагая, что цепочка не может складываться, т.е. случайное блуждание происходит равновероятно в три стороны.

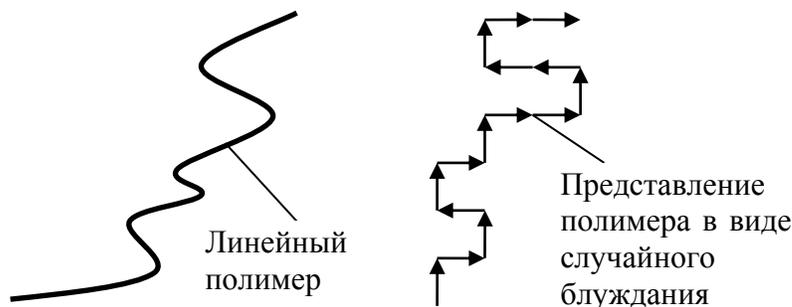


Рисунок 7 – Гибкие полимерные цепочки

19-20. Случайные блуждания с переменным шагом. Рассмотрите одномерное случайное блуждание со всеми допустимыми длинами прыжков. Вероятность того, что длина шага равна j , имеет вид $P(j)=\exp(-j)$. Определите удаление от начального положения после 10 шагов.

Лабораторная работа № 4

Моделирование системы массового обслуживания с одним устройством обслуживания

1 Цель работы

Детально рассмотреть моделирование системы массового обслуживания (СМО) с одним устройством обслуживания. Это банк с одним работником – кассиром.

Реализовать программно, согласно своему варианту, СМО, определить задачи, показать схемы моделируемой системы после возникновения каждого события, описать результаты моделирования, условия останова как способ окончания моделирования.

2 Теоретические сведения

2.1 Постановка задачи

Рассмотрим систему массового обслуживания с одним устройством (рис. 1).

Входные данные:

A_1, A_2, \dots, A_n – поступившие требования в систему (клиенты);

n – количество требований;

S_1, S_2, \dots, S_n – время обслуживания требования (время обслуживания кассиром клиента).

Промежуточные данные:

D_1, D_2, \dots, D_n – задержка требований в очереди (время нахождения клиента в очереди);

$Q(t)$ – число требований в момент времени t ;

$T(n)$ – время, необходимое для наблюдения n -го числа задержек в очереди.

Выходные данные:

$d(n)$ – ожидаемая средняя задержка в очереди для каждого из n -требований (среднее время нахождения клиентов в очереди);

$q(n)$ – ожидаемое среднее число требований в очереди (среднее число клиентов, находящихся в очереди);

$u(n)$ – ожидаемый коэффициент использования устройства.

В этой системе интервалы времени между поступлением требований A_1, A_2, \dots являются *независимыми и одинаково распределенными* случайными величинами, т.е. интервалы времени между поступлением требований имеют

одинаковое распределение вероятностей. Когда требование поступает, а устройство свободно, обслуживание начинается немедленно. Время обслуживания S_1, S_2, \dots следующих требований представлено независимыми одинаково распределенными случайными величинами, на которые не влияют интервалы времени обслуживания. Если при поступлении требования устройство занято, требование становится в очередь. По завершении обслуживания требования устройство выбирает требование из очереди (если такая имеется) по принципу FIFO (First-In, First-Out – «первым пришел – первым обслужен»).

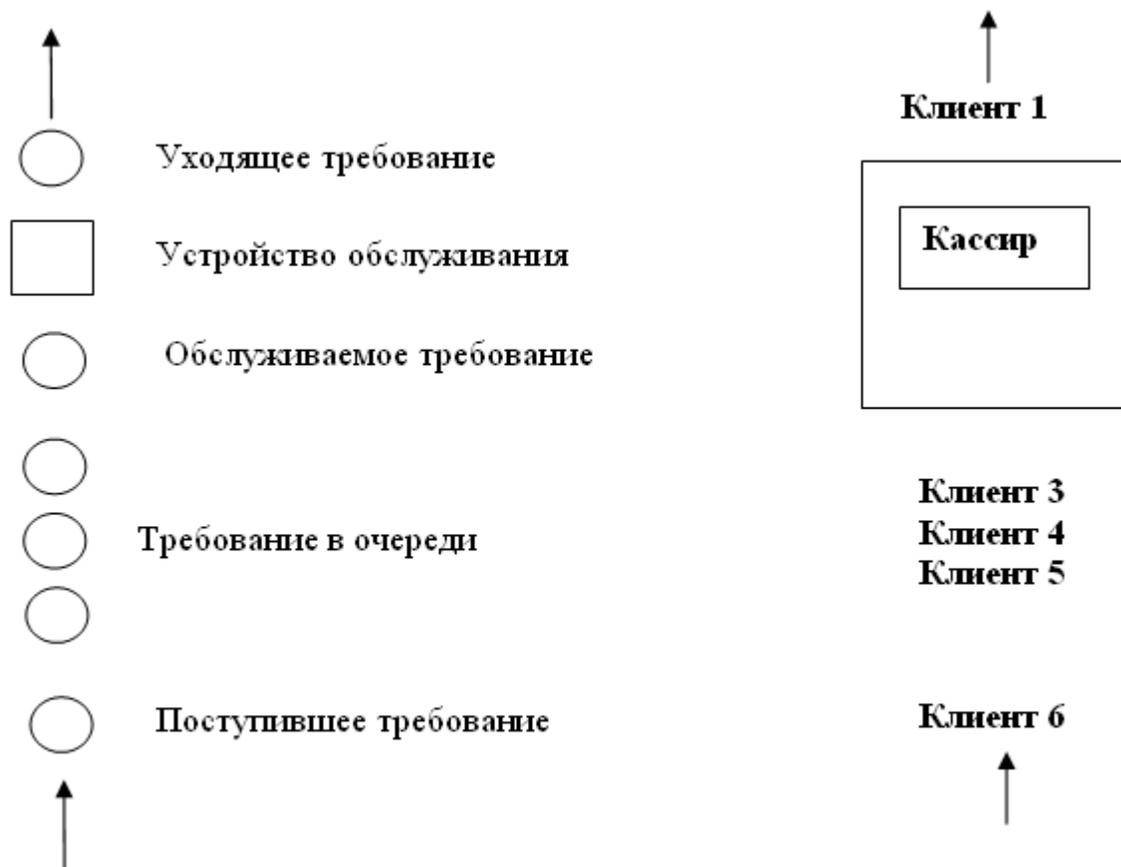


Рисунок 1 – Система массового обслуживания с одним устройством обслуживания

Моделирование начинается с состояния «пустая и незанятая», т.е. когда в системе нет требований, и устройство обслуживания свободно. С момента времени, равного 0, начинается ожидание поступления первого требования. Скорее всего, это произойдет по истечении первого интервала времени между поступлениями A_1 , а не в момент времени, равный 0. Моделирование будет продолжаться до тех пор, пока для определенного числа требований n не истечет время задержки в очереди, то есть моделирование прекратится, когда n -е требование будет обслужено. Следовательно, *время* завершения моделирования является случайной величиной, которая зависит от наблюдаемых значений случайных

переменных, обозначающих время между поступлениями требований и время обслуживания.

Для того чтобы определить *критерии оценки работы системы*, обратимся к трем параметрам их характеристики.

Прежде всего, необходимо найти ожидаемую *среднюю задержку в очереди* для каждого из n требований, завершивших свое ожидание во время моделирования; эта величина обозначается как $d(n)$. Слово «ожидаемая» в определении $d(n)$ означает, что при заданном прогоне имитационной модели средняя задержка, наблюдаемая для n -го числа требований, зависит от полученных наблюдений случайных переменных, обозначающих время между поступлениями требований и время обслуживания. При следующем прогоне имитационной модели (или на второй день – с реальной системой) требования будут прибывать в другие моменты времени, и необходимое время обслуживания также будет иным. Это приведет к получению другого среднего значения задержки для n требований. Следовательно, при прогоне имитационной модели среднее значение задержки требования в очереди также считается случайной величиной. Нам же необходимо установить *ожидаемое значение* случайной величины. Согласно одной интерпретации $d(n)$ является средним огромного (фактически бесконечного) числа средних задержек в очереди n -го числа требований. После одного прогона имитационной модели, при котором зафиксированы задержки требований в очереди D_1, D_2, \dots, D_n , оценкой $d(n)$ является формула

$$\hat{d}(n) = \frac{\sum_{i=1}^n D_i}{n},$$

представляющая среднее время D_i n -го числа задержек, которое наблюдалось при моделировании. Следует отметить, что говоря о времени задержки, не исключается возможность того, что оно будет равно 0 в случае поступления требования в момент, когда система находится в состоянии незанятости (в этой модели точно известно, что $D_i = 0$). Время задержки, равное 0, учитывается в среднем, поскольку большое количество таких случаев, когда время задержки равно 0, указывает на то, что система предоставляет надлежащее обслуживание, и выходные оценки критериев работы должны это отражать. Одна из причин того, что берется среднее значение задержек D_i , а не рассматривается каждая из них по отдельности, состоит в том, что их распределение не будет одинаковым (например, $D_1=0$, но D_2 может иметь положительное значение), и среднее является единым показателем времени задержки всех требований в очереди. $\hat{d}(n)$ – это оценка, основанная на выборке размера «1», поскольку выполняется только один полный прогон имитационной модели. Из элементарной статистики известно, что выборка размера 1 не имеет большой статистической ценности.

Второй необходимый критерий – **ожидаемое среднее число требований в очереди** (без учета уже находящихся на обслуживании), обозначенное

как $q(n)$, где n указывает, что среднее вычисляется за период времени, который необходим для наблюдения n -го числа задержек, определяющих правило останова. Это среднее отличается от среднего времени задержки требований в очереди, поскольку оно берется, скорее, в отношении конкретного (непрерывного) периода времени, нежели в отношении требований (которые дискретны). Таким образом, нужно определить среднее *по времени* число требований в очереди. С этой целью число требований в момент времени t (для любого вещественного числа $t > 0$) обозначается как $Q(t)$, а время, необходимое для наблюдения n -го числа задержек в очереди, как $T(n)$. Тогда для любого времени t между 0 и $T(n)$ значение $Q(t)$ будет неотрицательным целым числом. Далее, пусть p_i - ожидаемая часть (имеющая значение от 0 до 1) времени, когда $Q(t)$ равно i , тогда приемлемым определением $q(n)$ будет

$$q(n) = \sum_{i=0}^{\infty} i p_i.$$

Таким образом, $q(n)$ является взвешенным средним возможных значений i для очереди длиной $Q(t)$, при этом весовые коэффициенты являются ожидаемой частью времени, которое расходуется на каждую из своих возможных длин очереди. Чтобы определить $q(n)$, заменяем значения p_i их оценками и получаем

$$\hat{q}(n) = \sum_{i=0}^{\infty} i \hat{p}_i. \quad (1)$$

где \hat{p}_i является частью времени, измеряемой (а не ожидаемой) в процессе моделирования, когда число требований в очереди равнялось i . Однако для вычисления \hat{q}_i удобнее переписать формулу, используя некоторые геометрические методы. Пусть T_i будет суммарным временем в течение моделирования очереди длиной i , тогда $T(n) = T_0 + T_1 + T_2 + \dots$ и $\hat{p}_i = T_i/T_n$. В результате приведенное выше уравнение (1) можно переписать как

$$\hat{q}(n) = \frac{\sum_{i=0}^{\infty} iT_i}{T(n)}. \quad (2)$$

На рисунке 2 показано возможное течение времени или реализация $Q(t)$ в системе при условии, что $n = 6$ (кол-во временных промежутков). Требования прибывают в моменты времени, равные 0,4; 1,6; 2,1; 3,8; 4,0; 5,6; 5,8 и 7,2. Уход требований (завершение обслуживания) происходит в моменты времени, равные 2,4; 3,1; 3,3; 4,9 и 8,6. В момент времени, когда $T(6) = 8,6$, моделирование завершается.

Обращаясь к рисунку 2, имейте в виду, что $Q(t)$ не учитывает требование, находящееся на обслуживании (если таковое есть), и в период времени между 0,4 и 1,6 очередь будет отсутствовать ($Q(t)=0$), хотя при этом одно требование находится на обслуживании. То же самое касается периодов времени с 3,1 по 3,3, с 3,8 по 4,0 и с 4,9 по 5,6. Но в период времени с 3,3 по 3,8 требования

отсутствуют вообще, и система находится в состоянии незанятости, как и в период времени с 0 по 0,4.

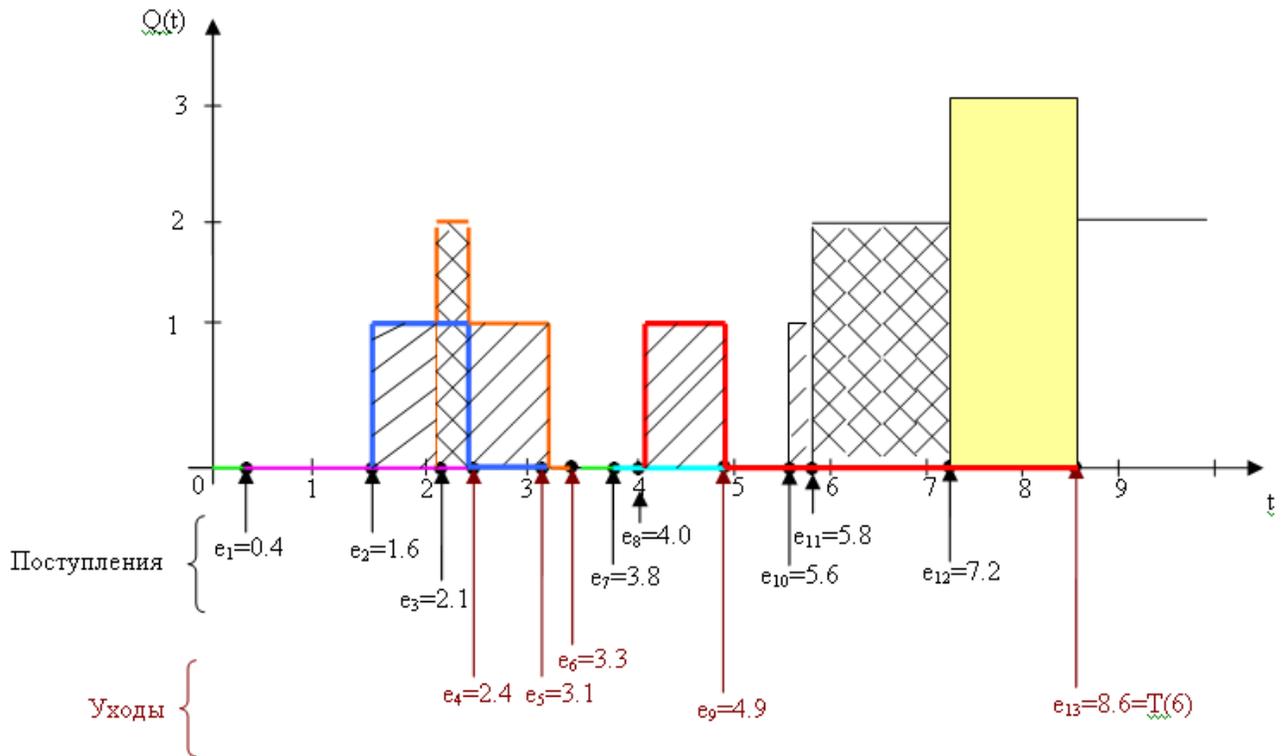


Рисунок 2 – $Q(t)$, время поступления и время ухода в реализации системы массового обслуживания с одним устройством обслуживания

Чтобы вычислить значение $\hat{q}(n)$, вначале необходимо вычислить значения T_i , которые, прибегнув к помощи рисунка 2, можно трактовать как интервалы времени (иногда отделенные друг от друга некоторым промежутком времени), в течение которых $Q(t)$ равно 0, 1, 2 (длина очереди клиентов) и т.д.:

$$T_0 = (1,6 - 0,0) + (4,0 - 3,1) + (5,6 - 4,9) = 3,2;$$

$$T_1 = (2,1 - 1,6) + (3,1 - 2,4) + (4,9 - 4,0) + (5,8 - 5,6) = 2,3;$$

$$T_2 = (2,4 - 2,1) + (7,2 - 5,8) = 1,7;$$

$$T_3 = (8,6 - 7,2) = 1,4.$$

($T_i = 0$, когда $i \geq 4$, поскольку в данной СМО очередь никогда не достигает такой длины. Максимальная длина очереди на рисунке 2 составляет $i=3$.)
Числитель в уравнении (2)

$$\sum_{i=0}^{\infty} iT_i = (0 \times 3,2) + (1 \times 2,3) + (2 \times 1,7) + (3 \times 1,4) = 9,9. \quad (3)$$

Отсюда оценка среднего по времени числа требований в очереди при данном прогоне имитационной модели $\hat{q}(6) = 9,9 / 8,6 = 1,15$. Каждый из не равных нулю членов в левой части уравнения (3) соответствует заштрихованной пло-

щади на рис. 2: $1 \times 2,3$ соответствует площади, заштрихованной диагональными линиями (в четырех прямоугольниках), $2 \times 1,7$ – площади, заштрихованной перекрестными диагональными линиями (в двух прямоугольниках), $3 \times 1,4$ – закрашенной желтым площади (один прямоугольник). Иными словами, суммирование в числителе уравнения (2) представляет накопленную *площадь под кривой $Q(t)$ между началом и концом моделирования*. С учетом того, что площадь под кривой является интегралом, можно написать следующее уравнение:

$$\sum_{i=0}^{\infty} iT_i = \int_0^{T(n)} Q(t) dt,$$

а оценка $q(n)$ может быть выражена как

$$\hat{q}(n) = \frac{\int_0^{T(n)} Q(t) dt}{T(n)} \quad (4)$$

Выражения (4) и (2) эквивалентны для $\hat{q}(n)$, однако (4) предпочтительнее, поскольку интеграл в нем может быть рассчитан как накопление простых площадей прямоугольников, возникающих в течение времени моделирования, что менее удобно при явном суммировании по формуле (2). Более того, выражение (4) предполагает среднее для непрерывного $Q(t)$, так как операцию интегрирования можно только приблизительно считать непрерывным суммированием.

Третий выходной критерий оценки работы системы – **показатель занятости устройства** (занятость кассира). Ожидаемый *коэффициент использования* устройства является отношением времени, когда устройство находится в состоянии занятости, ко времени моделирования (от 0 до $T(n)$) и, следовательно, является числом между 0 и 1. Обозначим его как $u(n)$. При одной реализации моделирования оценкой $u(n)$ является $\hat{u}(n)$, равная измеряемому отношению времени, когда устройство находится в состоянии занятости, ко времени моделирования. Теперь $\hat{u}(n)$ можно вычислить непосредственно при моделировании, отметив моменты времени, когда меняются состояния устройства (со свободного на занятое, и наоборот), а затем выполнив соответствующие действия с вычитанием и делением. Однако такую величину удобнее рассматривать как среднее для непрерывного времени, подобное средней длине очереди, определив функцию занятости как

$$B(t) = \begin{cases} 1, & \text{если устройство занято в момент времени } t; \\ 0, & \text{если устройство свободно в момент времени } t. \end{cases}$$

Таким образом, коэффициент $\hat{u}(n)$ можно выразить как часть времени, когда $B(t)=1$. На рисунке 3 изображен график функции $B(t)$ для реализации процесса моделирования, использованного на рисунке 2. В этом случае получаем

$$\hat{u}(n) = \frac{(3,3 - 0,4) + (8,6 - 3,8)}{8,6} = \frac{7,7}{8,6} = 0,9 \quad (5)$$

Результат указывает на то, что в ходе моделирования 90% времени устройство занято. Но числитель в формуле (5) может также рассматриваться как площадь под функцией $B(t)$, так как высота $B(t)$ всегда равна либо 0, либо 1. Следовательно,

$$\hat{u}(n) = \frac{\int_0^{T(n)} B(t) dt}{T(n)}, \quad (6)$$

Для многих случаев моделирования, включающих разные устройства обслуживания, статистика коэффициента использования дает достаточно информации для решения проблем нехватки ресурсов (коэффициент использования, равный 100%, связан с большими перегрузками показателей продвижения очереди) или их избытка (низкая занятость). Это особенно касается случаев, когда устройствами обслуживания являются такие дорогие приборы, как роботы производственных систем или большие компьютеры (мэйнфреймы), задействованные в операциях по обработке данных.

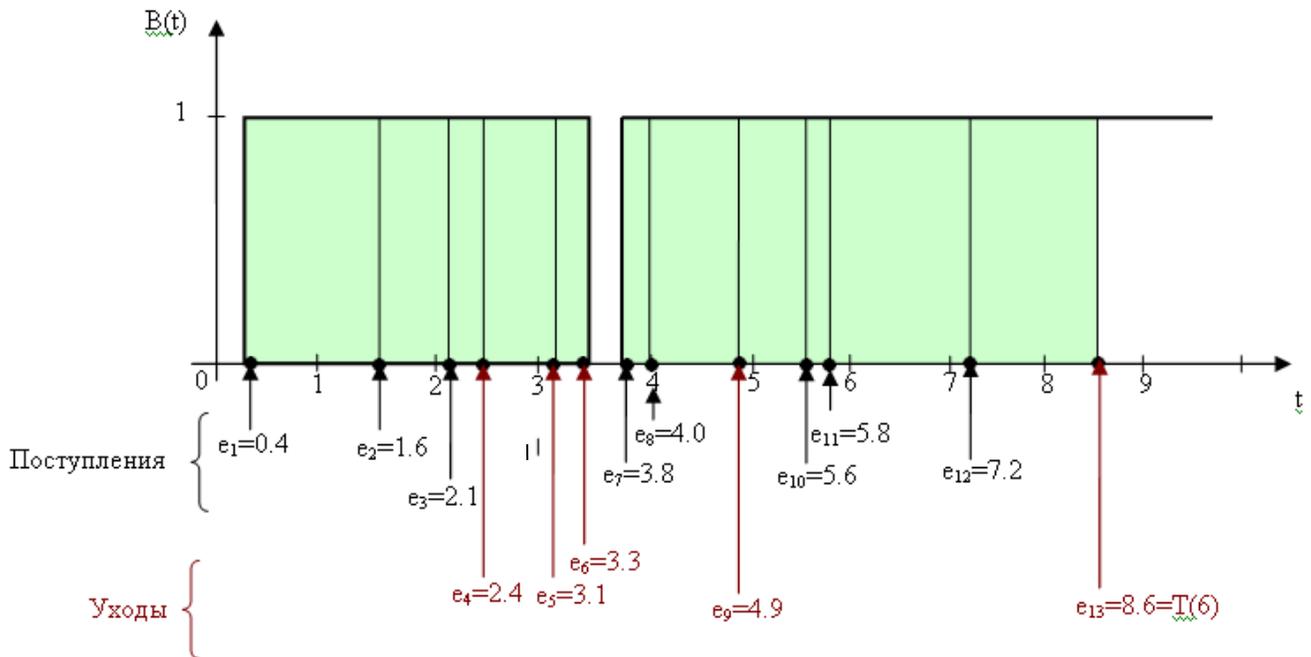


Рисунок 3 – $B(t)$, время поступления и время ухода в реализации системы массового обслуживания с одним устройством обслуживания (та же реализация, что на рисунке 2)

Итак, три критерия оценки работы системы включают:

- среднюю задержку в очереди $\hat{d}(n)$;

- среднее по времени число требований в очереди $\hat{q}(n)$;
- часть времени, когда устройство обслуживания находится в состоянии занятости $\hat{u}(n)$.

Средняя задержка требования в очереди является *статистикой дискретного времени*, т.к. она определяется относительно ряда случайных переменных $\{D_i\}$, которые имеют индекс дискретного времени $i = 1, 2, \dots$

Среднее по времени число требований в очереди и часть времени, когда устройство обслуживания находится в состоянии занятости, – это примеры *статистики непрерывного времени*, поскольку они определяются относительно совокупности случайных переменных $\{Q(t)\}$ и $\{B(t)\}$ соответственно, каждая из которых индексируется по параметру непрерывного времени $t \in [0, \infty)$. Символ \in означает «содержащийся». Таким образом, в этом случае t может быть любым не отрицательным вещественным числом.

Как статистика дискретного времени, так и статистика непрерывного времени часто используются в моделировании, при этом они позволяют получить не только средние показатели.

2.2 Процесс моделирования

Рассмотрим работу СМО с одним устройством обслуживания. Примером будет служить кассовый зал банка с одним работником – кассиром.

Условия постановки задачи:

1. Устройство обслуживания (кассир) – 1 чел.
2. Кассовый зал начинает работать с 8:00 (начальный момент времени $t=0$).
3. $\Delta T_{\text{при}}^{i-1} = 1 \dots 7$ (случайное число от 0 до 7) – время прихода i -го клиента после $i-1$ клиента, выражается в мин.
4. $\Delta T_{\text{обсл}}^i = 1 \dots 5$ (случайное число от 0 до 5) – время обслуживания кассира i -го клиента, выражается в мин.
5. Ограничение на работу кассового зала (*точка останова модели*): количество клиентов – $n=10$ чел.

2.3 Определение событий и переменных в системе

Входные данные:

A_1, A_2, \dots, A_n – поступившие требования в систему (клиенты);

n – количество требований;

$T_{\text{при}}^{i-1}$ – время прихода i -го клиента после $i-1$ клиента.

$T_{\text{обсл}i} = S_1, S_2, \dots, S_n$ – время обслуживания требования (время обслуживания кассиром i -го клиента);

$\Delta T_{\text{пр}}$ – реальное время прихода клиента ($T_{\text{пр}i}^{i-1}$ – время прихода клиента относительно предыдущего клиента).

$\Delta T_{\text{обсл}}$ – реальное время обслуживания клиента.

Промежуточные данные:

$T_{\text{оч}} = \{D_i\} = D_1, D_2, \dots, D_n$ – задержка требований в очереди (время нахождения клиента в очереди);

$D = Q(t)$ – число требований в момент времени t (длина очереди);

$T(n)$ – время, необходимое для наблюдения n -го числа задержек в очереди;

$T_{\text{пр}}$ – время прихода клиента в кассовый зал;

$T_{\text{нач}}$ – время начала обслуживания клиента;

$T_{\text{ух}}$ – время ухода клиента из кассового зала;

$T_{\text{прост}}$ – время простоя кассира.

Выходные данные:

$d(n)$ – ожидаемая средняя задержка в очереди для каждого из n требований (среднее время нахождения клиентов в очереди);

$q(n)$ – ожидаемое среднее число требований в очереди (среднее число клиентов, находящихся в очереди);

$u(n)$ – ожидаемый коэффициент использования устройства.

Приступая к моделированию, предположим, что в системе изначально не содержались никакие требования и устройство обслуживания (кассир) было свободно. Переменные состояния модели устанавливаются в исходную позицию: состояние устройства обслуживания равно 0 (0 используется, чтобы представить состояние незанятости, 1 – состояние занятости устройства, аналогично определению функции $V(t)$), и число требований (клиентов банка) в очереди также равно 0.

Для хранения получаемых данных в процессе моделирования составляется таблица 1, которая изначально пуста, но в ходе моделирования она заполняется сведениями о поведении системы. С помощью данной таблицы строятся графики поступления и обслуживания требований (см. рисунок 4).

Таблица 1 – Сведения о поведении системы

№ клиента	$T_{\text{при}}^{i-1}$	$T_{\text{обсл}}$	$T_{\text{пр}}$ (в час:мин)	$T_{\text{нач}}$ (в час:мин)	$T_{\text{ух}}$ (в час:мин)	Д	$T_{\text{оч}}$ (в мин)	$T_{\text{прост}}$ (в час:мин)
1	1	5	8:01	8:01	8:06	0	0:00	0:01
2	1	4	8:02	8:06	8:10	1	0:04	0:00
3	6	3	8:08	8:10	8:13	1	0:02	0:00
4	6	5	8:14	8:14	8:19	0	0:00	0:01
5	2	3	8:16	8:19	8:22	1	0:03	0:00
6	5	4	8:21	8:22	8:26	1	0:01	0:00
7	3	4	8:24	8:26	8:30	1	0:02	0:00
8	6	4	8:30	8:30	8:34	0	0:00	0:00
9	2	3	8:32	8:34	8:37	1	0:02	0:00
10	5	5	8:37	8:37	8:42	0	0:00	0:00

И так, между моментом времени 0 (8:00) и моментом поступления первого требования (клиента) проходит 1 мин ($T_{\text{при}}^{i-1}$), между поступлением первого и второго требований (клиентов) – 1 мин и т.д. Время, необходимое для обслуживания первого требования, равно 5 мин ($T_{\text{обсл}}$), второго – 4 мин и т.д. В реальном моделировании значения $T_{\text{при}}^{i-1}$ и $T_{\text{обсл}}$ генерируются из соответствующих им распределений вероятностей (по мере необходимости) в процессе моделирования. Приведенные числовые значения $T_{\text{при}}^{i-1}$ и $T_{\text{обсл}}$ выбраны так, чтобы можно было сгенерировать реализацию моделируемой системы, изображенную на рисунке 4, демонстрирующих процессы $Q(t)$ и $B(t)$.

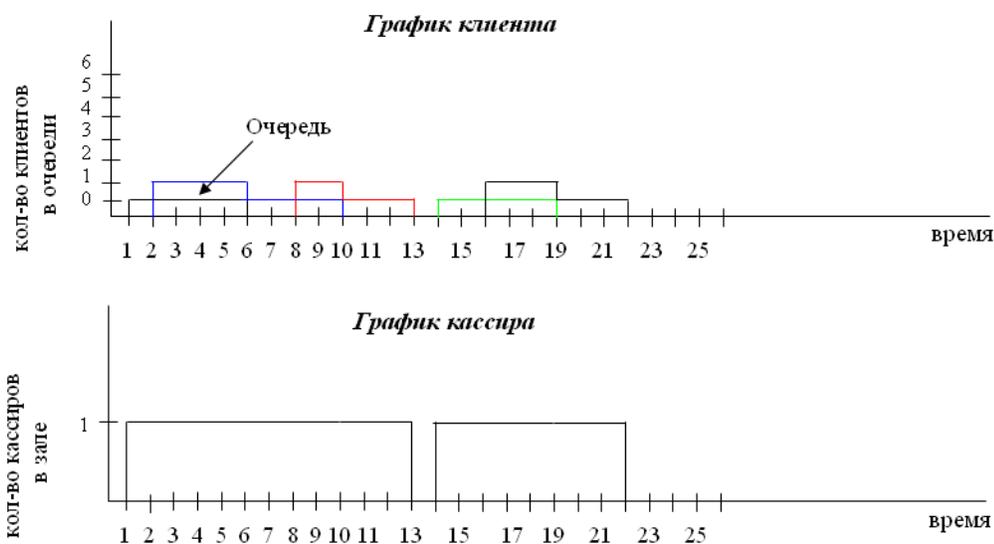


Рисунок 4 – Графики, отражающие время поступления и время ухода требований (клиентов) в реализации СМО банк с одним работником – кассиром

Рассмотрим процессы инициализации и поступлений требований (клиентов) и параллельно заполним таблицу 1.

Время поступления первого требования (клиента) равно $T_{пр1} = 8:00 + T_{пр1}^0$ (1 мин) = 8:01. Поскольку в этот момент требований на обслуживании нет, то данный клиент подходит к кассиру $T_{нач1} = T_{пр1}$. Время обслуживания клиента составляет $T_{обсл1} = 5$ мин, после этого клиент уходит $T_{ух1} = T_{нач1} + T_{обсл1} = 8:01 + 0:05 = 8:06$. Длина очереди составляет 0 клиентов. Времени, потраченного в очереди, нет. Время простоя кассира составляет 1 мин (период с 8:00 до 8:01).

Время поступления второго требования (клиента) равно $T_{пр2} = T_{пр1} + T_{пр2}^1$ (1 мин) = 8:01 + 0:01 = 8:02. Поскольку в этот момент первый клиент у кассира, то данный клиент становится в очередь, до тех пор, пока кассир не освободиться. Время обслуживания второго клиента составляет $T_{обсл2} = 4$ мин, после того как первый клиент ушел $T_{ух1} = 8:06$, второй клиент подходит к кассиру $T_{нач2} = T_{ух1} + T_{обсл2} = 8:06 + 0:04 = 8:10$. Это так называемые *одновременные события*, которые должны происходить в одно и то же модельное время, т.е. событие поступления нового требования происходит сразу за событием окончания обслуживания предыдущего требования. Длина очереди составляет 1 клиент (период с 8:02 до 8:06). Время, потраченное в очереди, составляет $T_{оч} = 0:04$. Время простоя кассира составляет 0 мин. И так до последнего n-го требования заполняется вся таблица.

С уходом требования 10 (десятого клиента) в момент времени 8:42 в системе возникает случай *останова модели*, процесс моделирования останавливается, и подсчитываются статистические данные работы модели.

В процессе моделирования получают следующие статистические данные:

1. Коэффициент занятости устройства обслуживания (долю простоя кассира от общего времени работы).
2. Среднее число требований в очереди (среднее число клиентов банка, стоящих в очереди).
3. Количество клиентов, посетивших кассовый зал за время его работы.
4. Длительность моделируемого периода (время работы кассового зала).
5. Максимальное и минимальное время нахождения клиента в очереди (Ответ выдать в час:мин).
6. Максимальная длина очереди (кол-во человек).
7. Среднее время нахождения клиента в очереди (от общего количества) (Ответ выдать в час:мин).
8. Максимальное и минимальное время работы устройства обслуживания (кассира) без перерыва (Ответ выдать в час:мин).
9. Среднее время пребывания клиента в зале.

2.5 Порядок выполнения работы

Вариант	Задание
1	<p>Программно реализовать работу кассового зала банка и показать графически состояние очереди клиентов и занятость кассира.</p> <p>Условия постановки задачи:</p> <ol style="list-style-type: none"> 1. Кассир – 1 чел. 2. Кассовый зал начинает работать с 8:00. 3. $\Delta T_{\text{при}}^{i-1}$ – моделируется по экспоненциальному распределению со средним временем 1 мин для интервалов между поступлениями. 4. $\Delta T_{\text{обсл}}$ – моделируется по экспоненциальному распределению со средним временем 0.5 мин для времени обслуживания. 5. Ограничения на работу кассового зала: время работы зала с 8:00 до 12:00. <p>Необходимо рассчитать:</p> <ol style="list-style-type: none"> 1. Коэффициент занятости устройства обслуживания (долю простоя кассира от общего времени работы). 2. Среднее число требований в очереди (среднее число клиентов банка, стоящих в очереди). 3. Количество клиентов, посетивших кассовый зал за время его работы. 4. Максимальное и минимальное время нахождения клиента в очереди (Ответ выдать в час:мин). 5. Максимальная длина очереди (кол-во человек). 6. Среднее время нахождения клиента в очереди (от общего количества) (Ответ выдать в час:мин). 7. Максимальное и минимальное время работы устройства обслуживания (кассира) без перерыва (Ответ выдать в час:мин). 8. Среднее время пребывания клиента в зале.
2	<p>Программно реализовать работу кассового зала банка и показать графически состояние очереди клиентов и занятость кассира.</p> <p>Условия постановки задачи:</p> <ol style="list-style-type: none"> 1. Кассир – 1 чел. 2. Кассовый зал начинает работать с 8:00. 3. $\Delta T_{\text{при}}^{i-1}$ – моделируется по экспоненциальному распределению со средним временем 1 мин для интервалов между поступлениями.

	<p>4. $\Delta T_{\text{обсл}}$ – моделируется по экспоненциальному распределению со средним временем 0.5 мин для времени обслуживания.</p> <p>5. Ограничения на работу кассового зала: количество клиентов – $n=1000$ чел.</p> <p>Необходимо рассчитать:</p> <ol style="list-style-type: none"> 1. Длительность моделируемого периода (время работы кассового зала). Ответ выдать в час:мин. 2. Коэффициент занятости устройства обслуживания (долю простоя кассира от общего времени работы). Ответ выдать в процентах. 3. Среднее число требований в очереди (среднее число клиентов банка, стоящих в очереди). 4. Максимальная и минимальная задержка требования в очереди (максимальное и минимальное время нахождения клиента в очереди). Ответ выдать в час:мин. 5. Максимальная длина очереди (максимальное кол-во клиентов, стоящих в очереди). 6. Среднее время нахождения клиента в очереди (от общего количества времени, проведенного клиентом в банке). Ответ выдать в час:мин. 7. Максимальное и минимальное время работы устройства обслуживания (кассира) без перерыва (Ответ выдать в час:мин). 8. Среднее по времени число требований в системе (среднее по времени число клиентов в зале).
3	<p>Предположим, что в СМО с одним устройством, описанной в вариантах 1 и 2, заведение открывается в 9:00 и закрывается в 17:00, но его работа продолжается до тех пор, пока не будет обслужен последний клиент, находившийся в системе в 17:00 (на обслуживании или в очереди). Напишите программу так, чтобы она отражала данное условие останова модели, и оцените те же самые показатели системы.</p>
4	<p>Обратитесь к задачам вариантов 1 и 2. Предположим, что для устройства обычно требуется 30-минутный перерыв в работе, который делается в первый момент времени после 12:00, когда в системе нет клиентов. Однако если перерыва в работе устройства не было до 13:00, он осуществляется по завершении обслуживания клиента, который находился на обслуживании в 13:00. (Допустим, что все клиенты, находившиеся в очереди в 13:00, будут ожидать окончания перерыва.) Если клиент прибывает во время перерыва, он может уйти немедленно,</p>

не дожидаясь обслуживания, такая ситуация называется *неприсоединением к очереди*. Вероятность того, присоединится ли такой клиент к очереди, будет зависеть от промежутка времени, остающегося до завершения перерыва. (Время завершения перерыва сообщается.) В частности, клиент, прибывший во время перерыва, может не присоединиться к очереди со следующей вероятностью

Время, оставшееся до возобновления работы устройства, мин	Вероятность неприсоединения клиента к очереди
[20, 30)	0,75
[10, 20)	0,50
[0, 10)	0,25

Оцените те же показатели системы, что и раньше. (Обратите внимание, что устройство обслуживания не занято во время перерыва, а среднее число в очереди вычисляется с учетом данных за время перерыва.) Кроме того, оцените ожидаемое число клиентов, не присоединившихся к очереди.

5 Рассмотрите СМО с одним устройством обслуживания, описанную в вариантах 1 и 2. Предположим, что время обслуживания требования известно в момент его поступления. По завершении обслуживания одного требования устройство переходит к обслуживанию требования в очереди (если таковые имеются), время обслуживания которого является наименьшим. Выполняйте моделирование до тех пор, пока не будет завершена задержка 1000 требований, и оцените ожидаемую среднюю задержку в очереди, ожидаемое среднее по времени число требований в очереди и ожидаемую часть требований, задержка которых превышает 1 мин. (Такая дисциплина обслуживания называется *самой короткой работой в первую очередь*.)

6 Круглосуточно работающая заправочная станция с одной колонкой может иметь два типа клиентов. Полицейские машины, которые прибывают через каждые 30 мин; первая машина прибывает в момент времени, равный 15 мин. Обычные (не полицейские) машины имеют экспоненциально распределенные интервалы между прибытиями со средним значением 5,6 мин; первая обычная машина прибывает в момент времени 0 (8:00). Время обслуживания для всех машин экспоненциально распределено со средним значением 4,8 мин. Машина, прибывшая, когда колонка свободна, сразу попадает на обслуживание.

	<p>Обычная машина, прибывшая, когда колонка занята, становится в конец единственной очереди. Полицейская машина, прибывшая, когда колонка занята, становится в начало очереди, перед любой обычной машиной. (Если в начале очереди уже есть полицейские машины, допустим, что новоприбывшая полицейская машина все равно становится перед ними в самом начале очереди.) Изначально система свободна, в ней нет машин. Моделирование должно выполняться до тех пор, пока не завершится задержка в очереди 500 машин (любого типа). Оцените ожидаемую среднюю задержку в очереди отдельно для каждого типа машин, а также ожидаемое среднее по времени число машин (каждого типа) в очереди и ожидаемый коэффициент использования колонки.</p>
7	<p>Круглосуточно работающая заправочная станция с одной колонкой может иметь два типа клиентов. Полицейские машины, которые прибывают через каждые 30 мин; первая машина прибывает в момент времени, равный 15 мин. Обычные (не полицейские) машины имеют экспоненциально распределенные интервалы между прибытиями со средним значением 5,6 мин; первая обычная машина прибывает в момент времени 0 (8:00). Время обслуживания для всех машин экспоненциально распределено со средним значением 4,8 мин. Машина, прибывшая, когда колонка свободна, сразу попадает на обслуживание. Обычная машина, прибывшая, когда колонка занята, становится в конец единственной очереди. Полицейская машина, прибывшая, когда колонка занята, становится в начало очереди, перед любой обычной машиной. (Если в начале очереди уже есть полицейские машины, допустим, что новоприбывшая полицейская машина становится за последней в очереди полицейской машиной.) Изначально система свободна, в ней нет машин. Моделирование должно выполняться до тех пор, пока не завершится задержка в очереди 500 машин (любого типа). Оцените ожидаемую среднюю задержку в очереди отдельно для каждого типа машин, а также ожидаемое среднее по времени число машин (каждого типа) в очереди и ожидаемый коэффициент использования колонки.</p>
7	<p>Программно реализовать работу кассового зала банка и показать графически состояние очереди клиентов и занятость кассира.</p> <p>Условия постановки задачи:</p> <ol style="list-style-type: none"> 1. Кассир – 1 чел. 2. Кассовый зал начинает работать с 9:00. 3. $\Delta T_{\text{при}}^{i-1}$ – моделируется по экспоненциальному распределению со средним временем 3 мин для интервалов между поступлениями.

	<p>4. $\Delta T_{\text{обсл}}$ – моделируется по показательному распределению со средним временем 0.7 мин для времени обслуживания.</p> <p>5. Ограничения на работу кассового зала: время работы зала с 9:00 до 16:00.</p> <p style="text-align: center;">Необходимо рассчитать:</p> <ol style="list-style-type: none"> 1. Коэффициент занятости устройства обслуживания (долю простоя кассира от общего времени работы). 2. Среднее число требований в очереди (среднее число клиентов банка, стоящих в очереди). 3. Количество клиентов, посетивших кассовый зал за время его работы. 4. Максимальное и минимальное время нахождения клиента в очереди (Ответ выдать в час:мин). 5. Максимальная длина очереди (кол-во человек). 6. Среднее время нахождения клиента в очереди (от общего количества) (Ответ выдать в час:мин). 7. Максимальное и минимальное время работы устройства обслуживания (кассира) без перерыва (Ответ выдать в час:мин). 8. Среднее время пребывания клиента в зале.
8	<p>Программно реализовать работу кассового зала банка и показать графически состояние очереди клиентов и занятость кассира.</p> <p style="text-align: center;">Условия постановки задачи:</p> <ol style="list-style-type: none"> 1. Кассир – 1 чел. 2. Кассовый зал начинает работать с 9:00. 3. $\Delta T_{\text{при}}^{i-1}$ – моделируется по бета-распределению распределению со средним временем 1,5 мин для интервалов между поступлениями. 4. $\Delta T_{\text{обсл}}$ – моделируется по экспоненциальному распределению со средним временем 0.3 мин для времени обслуживания. 5. Ограничения на работу кассового зала: количество клиентов – $n=1000$ чел. <p style="text-align: center;">Необходимо рассчитать:</p> <ol style="list-style-type: none"> 1. Длительность моделируемого периода (время работы кассового зала). Ответ выдать в час:мин. 2. Коэффициент занятости устройства обслуживания (долю простоя кассира от общего времени работы). Ответ выдать в процентах. 3. Среднее число требований в очереди (среднее число клиентов банка, стоящих в очереди).

- | |
|---|
| <ol style="list-style-type: none">4. Максимальная и минимальная задержка требования в очереди (максимальное и минимальное время нахождения клиента в очереди). Ответ выдать в час:мин.5. Максимальная длина очереди (максимальное кол-во клиентов, стоящих в очереди).6. Среднее время нахождения клиента в очереди (от общего количества времени, проведенного клиентом в банке). Ответ выдать в час:мин.7. Максимальное и минимальное время работы устройства обслуживания (кассира) без перерыва (Ответ выдать в час:мин).8. Среднее по времени число требований в системе (среднее по времени число клиентов в зале). |
|---|

2.6 Содержание работы и отчета

1. Титульный лист (для отчета).
2. Цель и порядок работы.
3. Задание.
4. Описание объекта исследования.
5. Описание входных, выходных параметров, внутренних параметров для расчета (но не являющихся выходными параметрами).
6. Описание входных и выходных переменных вашей программы.
7. Программный вывод на экран следующей информации:
 - таблица;
 - график клиента;
 - график кассира;
 - данные расчетов.
8. Вывод:
 - как работает СМО;
 - нужно ли еще одно устройство обслуживания?

Лабораторная работа № 5

Моделирование системы управления запасами

1 Цель работы

Изучить методы моделирования стратегий осуществления заказов в системе управления запасами. Рассмотреть модели управления запасами, т.к. многие элементы нашей модели характерны для действующих систем управления запасами.

2 Методические указания

Возникновение теории управления запасами можно связать с работами Ф. Эджуорта и Ф. Харриса, появившимися в конце XIX – начале XX вв., в которых исследовалась простая оптимизационная модель определения экономического размера партии поставки для складской системы с постоянным равномерным расходом и периодическим поступлением хранимого продукта.

Запасами называется любой ресурс на складе, который используется для удовлетворения будущих нужд. Примерами запасов могут служить полуфабрикаты, готовые изделия, материалы, различные товары, а также такие специфические товары, как денежная наличность, находящаяся в хранилище. Большинство организаций имеют примерно один тип системы планирования и контроля запасов. В банке используются методы контроля за количеством наличности, в больнице применяются методы контроля поставки различных медицинских препаратов.

Простейшая схема системы управления запасами выглядит следующим образом (см. рисунок 1):

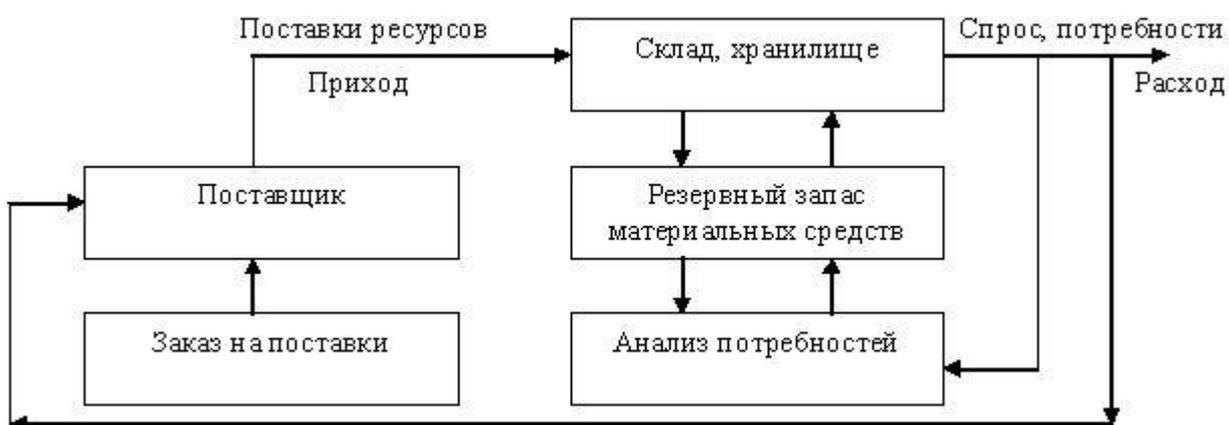


Рисунок 1 – Система управления запасами

Существуют причины, побуждающие организации создавать запасы:

- 1) дискретность поставок при непрерывном потреблении;
- 2) упущенная прибыль;

3) случайные колебания:

- в спросе за период между поставками;
- в объеме поставок;
- в длительности интервала между поставками;

4) предполагаемые изменения конъюнктуры:

- сезонность спроса;
- сезонность производства;
- ожидаемое повышение цен.

Имеются также причины, побуждающие предприятия стремиться к минимизации запасов на складе:

- 1) плата за физическое хранение запаса;
- 2) потери в количестве запаса;
- 3) моральный износ продукта.

Рассмотрим определяющие понятия теории управления запасами.

Издержки выполнения заказа (издержки заказа) – накладные расходы, связанные с приобретением / реализацией заказа и независящие от размера заказа (см. рисунок 2).

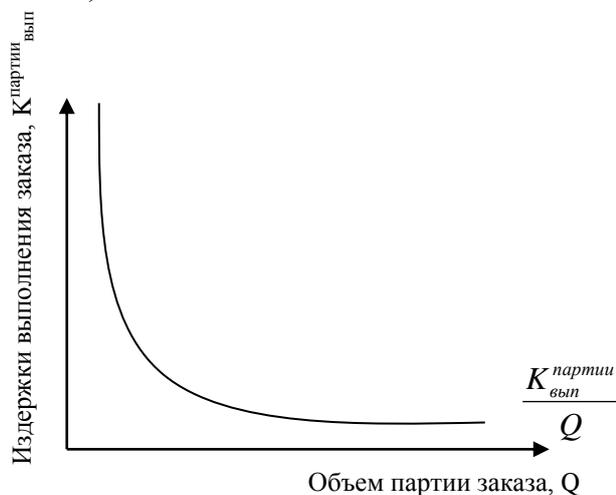


Рисунок 2 – Графическое изображение издержек заказа

Приобретение или изготовление продукции, помимо стоимости самих товаров, связано с еще некоторыми затратами. Так, при изготовлении собственной продукции для создания товарных запасов возникают издержки на наладку, пуск и остановку оборудования, на составление документации и т.д. Закупка производственных и товарных запасов сопряжена с возникновением административных расходов на совершение заказа, процедуру приемки материальных ценностей, обработку документации, транспортировку заказа.

Поскольку издержки выполнения заказа не зависят от размера заказываемой партии, то издержки выполнения заказа в расчете на единицу заказываемых материальных ресурсов $K_{\text{вып}}^{\text{ед}}$ (ден. ед./ нат. ед.) изменяются обратно пропорционально размеру партии.

Издержки выполнения заказов при постоянном их размере за период времени $K_{\text{вып}}$ (ден. ед.) определяются по формуле:

$$K = K_{\text{вып}} = K_{\text{вып}}^{\text{партии}} \times n = K_{\text{вып}}^{\text{партии}} \times \frac{D}{Q},$$

где $K_{\text{вып}}^{\text{партии}}$ – издержки выполнения одного заказа (ден. ед.);

n – количество поставок за период времени;

D – потребность в запасе данного вида (нат. ед.);

Q – размер заказа (нат. ед.).

Издержки хранения (затраты на хранение) – расходы, связанные с физическим содержанием товаров на складе, плюс возможные проценты на капитал, вложенный в запасы (см. рисунок 3).

Издержки хранения запасов включают:

- стоимость капитала, замороженного в запасах (зависит от текущей ставки процента);
- стоимость хранения запасов, включая занимаемое пространство, оборудование, труд обслуживающего персонала (выражается, как правило, также в процентах от стоимости хранящихся материалов, чаще всего составляет 5–10% в год);
- стоимость потерь из запаса (случайные поломки, незапланированное превышение срока хранения, устаревание запасов, в том числе моральное, “усушка – утриска”, а также кража запасов, выражаются потери из запаса также в процентах от стоимости хранящихся материальных ресурсов).



Рисунок 3 – Графическое изображение издержек хранения

Затраты на хранение запасов, таким образом, выражаются или в абсолютных единицах, или в процентах от закупочной цены и связываются с определенным промежутком времени, в большинстве случаев составляют 15-30% в год.

Упущенная прибыль – издержки, связанные с неудовлетворительным спросом, возникающим в результате отсутствия продукта на складе.

Совокупные издержки за период представляют собой сумму издержек

заказа, издержек хранения и упущенного дохода. Иногда к ним прибавляются издержки на покупку товаров.

Срок выполнения заказов – срок между заказом и его выполнением.

Точка восстановления (точка заказа) – фиксированный уровень запаса, при снижении до которого подается заказ на поставку следующей партии.

2.1 Краткая характеристика моделей управления запасами

а) *Модель оптимального размера заказа.*

Предпосылки:

- 1) темп спроса на товар известен и постоянен;
- 2) получение заказа мгновенно;
- 3) отсутствуют количественные скидки при закупке больших партий товара;
- 4) единственные меняющиеся параметры – издержки заказа и хранения;
- 5) исключается дефицит в случае своевременного заказа.

Исходные данные: темп спроса; издержки заказа и хранения.

Результат: оптимальный размер заказа; время между заказами и их количество за период.

б) *Модель оптимального размера заказа в предположении, что получение заказа не мгновенно.*

Следовательно, нужно найти объем запасов, при котором необходимо делать новый заказ.

Исходные данные: темп спроса; издержки заказа и хранения; время выполнения заказа.

Результат: оптимальный размер заказа; время между заказами; точка восстановления запаса.

в) *Модель оптимального размера заказа в предположении, что допускается дефицит продукта и связанная с ним упущенная прибыль.*

Необходимо найти точку восстановления.

Исходные данные: темп спроса; издержки заказа и хранения; упущенная прибыль.

Результат: оптимальный размер заказа; время между заказами; точка восстановления запаса.

г) *Модель с учетом производства* (в сочетании с условиями а–в).

Необходимо рассматривать уровень ежедневного производства и уровень ежедневного спроса.

Исходные данные: темп спроса; издержки заказа и хранения; упущенная прибыль; темп производства.

Результат: оптимальный уровень запасов (точка восстановления).

д) *Модель с количественными скидками.* Появляется возможность количественных скидок в зависимости от размера заказа. Рассматривается зависимость издержек хранения от цены товара. Оптимальный уровень заказа определяется исходя из условия минимизации общих издержек для каждого вида скидок.

Рассмотрим каждую модель более детально.

Модель а наиболее экономичного размера заказа. Заказ, пополняющий запасы, поступает как одна партия. Уровень запасов убывает с постоянной интенсивностью пока не достигнет нуля. В этой точке поступает заказ, размер которого равен Q , и уровень запасов восстанавливается до максимального значения. При этом оптимальным решением задачи будет тот размер заказа, при котором минимизируются общие издержки за период (рисунок 4).

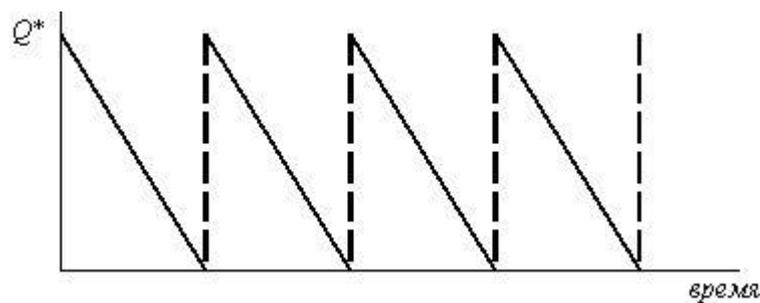


Рисунок 4 – Модель оптимального размера заказа

Пусть Q – размер заказа;

T – протяженность периода планирования;

D – величина спроса за период планирования;

d – величина спроса в единицу времени;

K – издержки заказа;

H – удельные издержки за период;

h – удельные издержки хранения в единицу времени;

i – издержки хранения запасов (% от);

K_3 – стоимость единицы запасов (ден. ед. / нат. ед.).

Тогда:

– совокупные издержки заказа $= \frac{D}{Q} K$;

– удельные издержки за период $H = K_3 \cdot i$;

– совокупные издержки хранения $= \frac{Q}{2} H$;

– величина спроса в единицу времени $d = \frac{D}{T}$;

– удельные издержки хранения в единицу времени $h = \frac{H}{T}$;

– оптимальный размер заказа $Q^* = \sqrt{\frac{2dK}{h}} = \sqrt{\frac{2DK}{H}}$;

- оптимальное число заказов за период $N = \frac{D}{Q^*}$;
- время цикла (оптимальное время между заказам) $t^* = \frac{Q^*}{d} = \frac{T}{N}$.

Модель б. Введем предположение о том, что заказ может быть получен не мгновенно, а с течением времени. Тогда нам необходимо заранее делать заказ, чтобы в нужное время иметь достаточное количество товара на складе. Следовательно, нам необходимо найти тот уровень запасов, при котором делается новый заказ. Этот уровень называется точкой восстановления R . Пусть L – время выполнения заказа. Тогда $R =$ величина спроса в единицу времени, умноженная на время выполнения заказа $= d \cdot L$. Другие характеристики системы определяются так же, как и в модели *a*. Модель иллюстрируется рисунком 5.

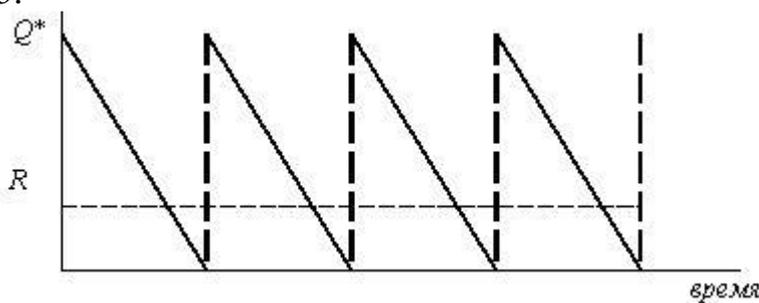


Рисунок 5 – Модель оптимального размера заказа в предположении, что получение заказа не мгновенно

Пример 1. Андрей Удачливый является торговым агентом компании TOYOTA и занимается продажей последней модели этой марки автомобиля. Годовой спрос оценивается в 4 000 ед. Цена каждого автомобиля равна 90 тыс. руб., а годовые издержки хранения составляют 10% от цены самого автомобиля. Андрей произвел анализ издержек заказа и понял, что средние издержки заказа составляют 25 тыс. руб. на заказ. Время выполнения заказа равно восьми дням. В течение этого времени ежедневный спрос на автомобили равен 20.

- Чему равен оптимальный размер заказа?
- Чему равна точка восстановления?
- Каковы совокупные издержки?
- Каково оптимальное количество заказов в год?
- Каково оптимальное время между двумя заказами, если предположить, что количество рабочих дней в году равно 200?

Исходные данные

- величина спроса за год $D = 4\ 000$;
- издержки заказа $K = 25$;
- издержки хранения $h = \frac{9}{200} = 0,045$;

- цена за единицу $K_3 = 90$;
- время выполнения заказа $L = 8$;
- ежедневный спрос $d = 20$;
- число рабочих дней $T = 200$.

Решение

- оптимальный размер заказа $Q^* = \sqrt{\frac{2 \cdot 20 \cdot 25}{0,045}} \approx 149$;
- точка восстановления $R = 8 \cdot 20 - 149 = 160 - 149 = 11$;
- число заказов за год $N = \frac{D}{Q^*} = 26,83$;
- совокупные издержки = совокупные издержки заказа + совокупные издержки хранения $C = \frac{4000}{149} \cdot 25 + \frac{149}{2} \cdot 9 = 1341,64$;
- стоимость продаж = 360 000;
- число дней между заказами $t = 7,45$.

Модель в оптимального размера заказа в предположении, что допускается дефицит продукта и связанная с ним упущенная прибыль (см. рисунок б).

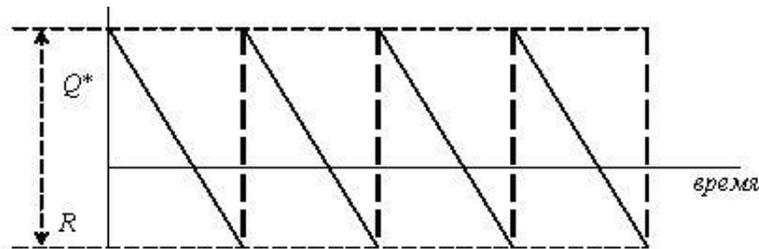


Рисунок б – Модель оптимального размера заказа в предположении, что допускается дефицит продукта и связанная с ним упущенная прибыль

Пусть p – упущенная прибыль в единицу времени, возникающая в результате дефицита одной единицы продукта;

P – упущенная прибыль за период, возникающая в результате дефицита одной единицы продукта.

Тогда:

- оптимальный размер заказа $Q^* = \sqrt{\frac{2dK}{h}} \cdot \sqrt{\frac{p+h}{p}} = \sqrt{\frac{2DK}{H}} \cdot \sqrt{\frac{P+H}{P}}$;
- максимальный размер запаса $S^* = \sqrt{\frac{2dK}{h}} \cdot \sqrt{\frac{p}{p+h}} = \sqrt{\frac{2DK}{H}} \cdot \sqrt{\frac{P}{P+H}}$;
- максимальный дефицит $R = Q^* - S^*$.

Модель 2 производства и распределения. В предыдущей модели допускалось, что пополнение запаса происходит единовременно. Но в некоторых случаях, особенно в промышленном производстве, для комплектования пар-

тии товаров требуется значительное время и производство товаров для пополнения запасов происходит одновременно с удовлетворением спроса. Такой случай показан на рисунке 7.

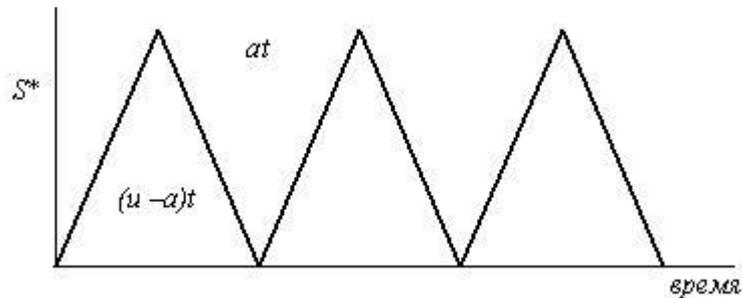


Рисунок 7 – Модель с учетом производства

Спрос и производство являются частью цикла восстановления запасов.

Пусть u – уровень производства в единицу времени;

K – фиксированные издержки хранения.

Тогда:

– совокупные издержки хранения = (средний уровень запасов) \times H =
 $= \frac{Q}{2} \left(1 - \frac{d}{u}\right) \cdot H$;

– средний уровень запасов = (максимальный уровень запасов)/2;

– максимальный уровень запасов = $u \cdot t - d \cdot t = Q \left(1 - \frac{d}{u}\right)$;

– время выполнения заказа $t = \frac{Q}{u}$;

– издержки заказа = $\frac{D}{Q} K$;

– оптимальный размер заказа $Q^* = \sqrt{\frac{2dK}{h} \left(1 - \frac{d}{u}\right)} = \sqrt{\frac{2DK}{H} \left(1 - \frac{d}{u}\right)}$;

– максимальный уровень запасов $S^* = Q^* \left(1 - \frac{d}{u}\right)$.

Модель d с количественными скидками. Для увеличения объема продаж компании часто предлагают количественные скидки своим покупателям.

Количественная скидка – сокращенная цена на товар в случае покупки большого количества этого товара. Типичные примеры количественных скидок приведены в таблице 1.

Таблица 1 – Примеры количественных скидок

Варианты скидок	1	2	3
Количество, при котором делается скидка	от 0 до 999	от 1000 до 1999	от 2000 и выше
Размер скидки, %	0	3	5

Цена со скидкой	5	4,8	4,75
-----------------	---	-----	------

Пусть I – доля издержек хранения в цене продукта K_3 .

Тогда:

– $h=I \cdot K_3$;

– оптимальный размер заказа $Q^* = \sqrt{\frac{2dK}{IK_3}}$.

Пример 2. Рассмотрим пример, объясняющий принцип принятия решения в условиях скидки. Магазин "Медвежонок" продает игрушечные гоночные машинки. Эта фирма имеет таблицу скидок на машинки в случае покупок их в определенном количестве (см. таблицу 1). Издержки заказа составляют 49 тыс. руб. Годовой спрос на машинки равен 5000. годовые издержки хранения в отношении к цене составляют 20%, или 0,2. Необходимо найти размер заказа, минимизирующий общие издержки.

Решение

Рассчитаем оптимальный размер заказа для каждого вида скидок, т.е. $Q1^*$, $Q2^*$ и $Q3^*$, и получим $Q1^* = 700$; $Q2^* = 714$; $Q3^*=718$.

Так как $Q1^*$ – величина между 0 и 999, то ее можно оставить прежней. $Q2^*$ меньше количества, необходимого для получения скидки, следовательно, его значение необходимо принять равным 1 000 единиц. Аналогично $Q3^*$ берем равным 2 000 единиц. Получим $Q1^* = 700$; $Q2^* = 1 000$; $Q3^* = 2 000$.

Далее необходимо рассчитать общие издержки для каждого размера заказа и вида скидок, а затем выбрать наименьшее значение.

Рассмотрим следующую таблицу 2.

Таблица 2 – Расчет данных

Варианты скидок	1	2	3
Цена	5	4,8	4,75
Размер заказа	700	1 000	2 000
Цена на товар за год	25 000	24 000	23 750
Годовые издержки заказа	350	245	122,5
Годовые издержки хранения	350	480	950
Общие годовые издержки	25 700	24 725	24 822,5

Выберем тот размер заказа, который минимизирует общие годовые издержки. Из таблицы видно, что заказ в размере 1 000 игрушечных гоночных машинок будет минимизировать совокупные издержки.

3 Порядок выполнения работы

Вариант 1

Менеджер приобретает в течение года 1500 телевизоров для розничной продажи в своем магазине. Издержки хранения каждого телевизора равны 45 тыс. руб. в год. Издержки заказа – 150 тыс. руб. Количество рабочих дней в году равно 300, время выполнения заказа – 6 дней.

Необходимо найти:

- оптимальный запас заказа;
- годовые издержки заказа;
- точку восстановления запаса.

Вариант 2

Менеджер продает 400 водяных кроватей в год, причем издержки хранения равны 1 тыс. руб. за кровать в день и издержки заказа – 40 тыс. руб. Количество рабочих дней равно 250 и время выполнения заказа – 6 дней.

- Каков оптимальный размер заказа?
- Чему равна точка восстановления запаса?
- Каков оптимальный размер заказа, если издержка хранения равны 1,5 тыс. руб.?

Вариант 3

Владелец маленькой компании, которая выпускает электрические ножи, может производить 150 ножей в день. Дневной спрос на ножи примерно равен 40. Фиксированные издержки производства равны 100 тыс. руб., издержки хранения – 8 тыс. руб. за нож в год. Каков максимальный заказ следует иметь на складе?

Вариант 4

Компания закупает у завода-изготовителя лобовые стекла грузовых автомобилей для розничной продажи. В год, за 200 рабочих дней, реализуется около 10 000 стекол. Издержки заказа для компании составляют 400 тыс. руб., ежедневные издержки хранения одного стекла – 6 тыс. руб.

- Чему равен оптимальный размер заказа?
- Каковы минимальные годовые совокупные издержки?

Вариант 5

Годовой заказ на тостер равен 3 000 единиц, или 10 в день. Издержки заказа равны 25 тыс. руб., издержки хранения – 0,4 тыс. руб. в день. Так как тостер является очень популярным среди покупателей, то в случае отсутствия товара покупатель обычно согласен подождать. Пока не подойдет следующий заказ. Однако издержки, связанные с дефицитом, равны 0,75 тыс. руб. за тостер в день.

- Сколько тостеров будет заказывать менеджер.

- Каков минимальный дефицит?
- Чему равны совокупные издержки?

Вариант 6

Магазин пользуется популярностью у покупателей благодаря широкому ассортименту экологически чистых продуктов. Большинство покупателей не отказываются от услуг магазина даже в том случае, когда интересующий их товар отсутствует в продаже. Они оставляют заказ на товар и ждут, когда поступит новая партия.

Сыр – не самый популярный из всего набора товаров, но администратор магазина регулярно заказывает этот продукт. Годовой спрос на сыр составляет 500 головок. Издержки заказа – 40 тыс. руб. за заказ. Издержки хранения – 5 тыс. руб. в год. Упущенная прибыль вследствие дефицита составляет 100 тыс. руб. в год на одну головку сыра.

- Сколько головок сыра следует заказывать, чтобы не допускать дефицита и иметь при этом минимальные общие издержки?
- Сколько сыра следует заказывать, если допустить возможность дефицита?
- Чему равна точка восстановления запаса, если время выполнения заказа 10 дней и число рабочих дней в году 250?
- Чему равен максимальный размер дефицита?

Вариант 7

Компания предлагает следующие скидки для линолеума размером 2×3 м.

Размер заказа	9 кусков или менее	10-50 кусков	50 кусков и более
Цена 1 куска	18 тыс. руб.	17,5 тыс. руб.	17,25 тыс. руб.

Магазин заказывает у компании линолеум. Издержки заказа равны 45 тыс. руб. Годовые издержки хранения равны 50% от цены. Годовой спрос на линолеум в магазине составляет 100 кусков. Какое количество необходимо приобрести?

Вариант 8

Мебельный салон продает в год около 1 000 спальных гарнитуров по цене 50 тыс. руб. Размещение одного заказа на поставку гарнитуров обходится в 40 тыс. руб. Годовая стоимость хранения гарнитура составляет 25% его цены. Салон может получать 3%-ную скидку у поставщика, если размер заказа составит не менее 200 гарнитуров. Следует ли салону заказывать 200 или более гарнитуров и пользоваться скидкой?

Вариант 9

Допустим, что товар в запасах *скоропортящийся*, и его срок хранения равномерно распределен между 1,5 и 2,5 мес. То есть если товар имеет срок хранения l мес., тогда по прошествии l мес. после помещения его в хранили-

ще он портится. (Различные товары в заказе, получаемом от поставщика, будут иметь различные сроки хранения.) Компания обнаруживает, что товар испортился только непосредственно при проверке перед продажей. Если установлено, что товар испорчен, его выбрасывают, и проверяют следующую единицу товара. Предположим, что товары в запасах проверяются в порядке FIFO. Повторите девять прогонов имитационной модели и после этого определите те же расходы, что и раньше. Вычислите также часть товаров, выброшенных по причине негодности. Для дополнительных сведений см. вариант 6.

Литература:

1. Охорзин, В.А. Прикладная математика в системе MATHCAD. [Электронный ресурс] — Электрон. дан. — СПб.: Лань, 2009. — 352 с. — Режим доступа: <http://e.lanbook.com/book/294>
2. Бахвалов, Л.А. Моделирование систем. [Электронный ресурс] — Электрон. дан. — М.: Горная книга, 2006. — 295 с. — Режим доступа: <http://e.lanbook.com/book/3511>
3. Голубева, Н.В. Математическое моделирование систем и процессов. [Электронный ресурс] — Электрон. дан. — СПб.: Лань, 2013. — 192 с. — Режим доступа: <http://e.lanbook.com/book/4862>
4. Горлач, Б.А. Исследование операций. [Электронный ресурс] — Электрон. дан. — СПб.: Лань, 2013. — 448 с. — Режим доступа: <http://e.lanbook.com/book/4865>
5. Чепурницкий, В.С. Исследование операций на основе стандартных программ. [Электронный ресурс] / В.С. Чепурницкий, А.В. Чесноков. — Электрон. дан. — М.: Горная книга, 2002. — 121 с. — Режим доступа: <http://e.lanbook.com/book/3542>