Министерство образования и науки РФ
ФГБОУ ВО «ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к выполнению лабораторных работ 1-4 по дисциплине «Математическое моделирование на ЭВМ» для студентов, обучающихся по направлению подготовки бакалавров 01.03.02 – «Прикладная математика и информатика»

Махачкала 2021

Методические указания к выполнению практических работ 1-4 по дисциплине «Математическое моделирование на ЭВМ» для студентов, обучающихся по направлению подготовки бакалавров 01.03.02 – «Прикладная математика и информатика» - Махачкала: ДГТУ, 2021г. -30 с.

Методические указания содержат описания практических работ по следующим темам: «Математические модели и их виды» , «Моделирование случайных чисел», «Моделирование случайных чисел с заданным законом распределения», «Линейная и квадратичная аппроксимация статистических данных»; примеры заданий с описанием технологии выполнения; задания к выполнению практических работ.

Данные методические указания посвящены практической части обучения по дисциплине «Математическое моделирование на ЭВМ». Изучение данной дисциплины базируется на знаниях, полученных студентами в процессе изучения высшей математики, программирования и физики. Изучение дисциплины закладывает основу для освоения в дальнейшем курсов по специальности, выполнения самостоятельных исследований и анализа результатов в рамках учебно-исследовательской и дипломной работ.

Составители:

Л.М.Гаджимахадова - ст. преподаватель кафедры прикладной математики и информатики,

Исабекова Т.И. - зав. кафедрой ПМ и И, к.ф.-м.н., доцент,

Алиосманова О.А.- ст. преподаватель кафедры прикладной математики и информатики.

Рецензенты:
Декан факультета информатики и информационных
технологий ДГУ, д.т.н., профессор С.А..Ахмедов

доцент кафедры ПОВТиАС,
д.т.н. А.Г. Мустафаев

Печатается согласно постановлению Ученого совета Дагестанского государственного технического университета
«_____»_____ 2021г. (протокол № _____)

Практическая работа №1

Математические модели и их виды
Теоретический материал

Существенно важным в теории математического моделирования является постоянное согласование всех аспектов построения модели с задачами и целями исследования. Поэтому сосредоточим внимание на некоторых существенных для исследований особенностях **механических** систем и процессов. Во-первых, факторы, определяющие такие объекты, характеризуются, как измеримые величины – параметры. Во-вторых, в основе таких моделей лежат уравнения, описывающие фундаментальные законы природы (механики), не нуждающиеся в пересмотре и уточнении. Даже готовые частные модели отдельных явлений, используемые при составлении более общих, хорошо сформулированы и описаны с точки зрения условий и областей применения. В-третьих, наибольшую трудность при разработке моделей механических систем и процессов представляет описание недостоверно известных характеристик объекта, как функциональных, так и числовых. В-четвертых, современные требования к таким моделям приводят к необходимости учета множества факторов, влияющих на поведение объекта, не только таких, которые связаны известными законами природы. Все эти особенности приводят к тому, что модели механических систем и процессов относятся в основном к классу математических.

Математические модели основываются на математическом описании объекта. В математическое описание, прежде всего, входят, и это естественно, взаимосвязи параметров объекта, что характеризует его особенности функционирования. Такие связи могут представляться в виде:

- вектор-функций $\mathbf{y} = \mathbf{f}(\mathbf{x},t)$,
- неявных функций $\mathbf{F}(\mathbf{y},\mathbf{x},t) = 0$, – обыкновенных дифференциальных уравнений

$$\mathbf{F}(\mathbf{x},\mathbf{x}',\mathbf{x}'',...,\mathbf{x}^{(m)},t) = 0,$$

- дифференциальных уравнений с частными производными

$$\mathbf{F}y,\mathbf{x},t, \ \frac{y}{x}, \ \frac{y}{t}, \ ... \ 0$$

- вычислительного алгоритма,
- вероятностного (стохастического) описания.

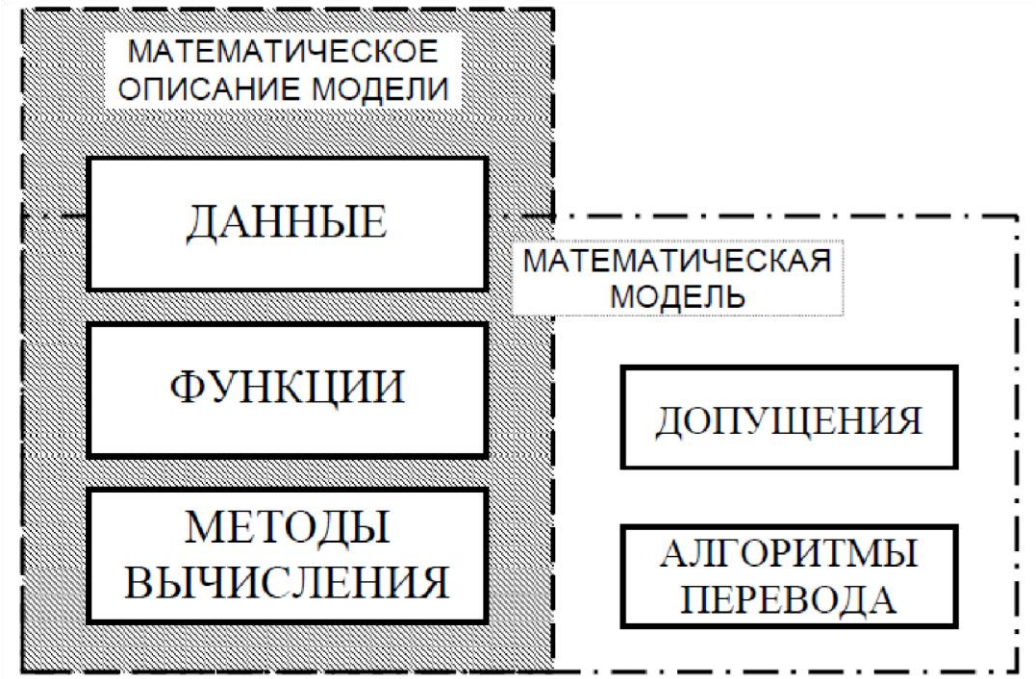
Первые четыре из указанных видов носят обобщающее название: *аналитических зависимостей*.

Математическое описание включает в себя не только взаимосвязь элементов и параметров объекта (**законы и закономерности**), но и полный набор числовых и функциональных **данных** объекта (характеристики; начальные, граничные, конечные условия; ограничения), а также **методы вычисления** выходных параметров модели. Т.е. под математическим описанием понимается **полная совокупность** данных, функций и методов вычисления, позволяющая **получать результат**.

Со своей стороны в *математическую модель* может не входить часть *математического описания* (чаще всего некоторые исходные данные), но помимо него

должны присутствовать описания всех **допущений**, использованных для ее построения, а также **алгоритмы перевода** исходных и выходных данных с модели на оригинал и обратно.

Рис.1



В качестве дополнения к классификации *математические модели* в зависимости от природы объекта, решаемых задач и применяемых **методов** могут различаться следующими видами:

- расчетные (формулы, таблицы, алгоритмы, графики, номограммы);
- соответственные (например, модель в аэродинамической трубе и реальный полет самолета в атмосфере);
- подобные (одинаковые математические описания **И** пропорциональные соответствующие параметры);
- линейные или нелинейные (описываемые функциями, которые содержат основные параметры только в степени 0 и 1, или любыми видами функций),
- стационарные или нестационарные (независящие или зависящие от времени),
- непрерывные или дискретные,
- детерминированные или стохастические (точные, однозначные или вероятностные: модели массового обслуживания, имитационные и др.),
- четкие или нечеткие (примеры нечетких множеств: около 10; глубоко или мелко; хорошо или плохо).

Понятие математических моделей объединяет чрезвычайно широкий круг моделей разнообразного вида. Используемая в аэродинамике с 2уа

аппроксимация поляры летательного аппарата c_{xa} сха0 может рассматриваться как математическая модель. Это – простейший пример. Но математическими моделями

называются и сложные вычислительные комплексы с многочисленным программным обеспечением для моделирования процессов развития экономики.

Исторически сложилось так, что под *математической моделью* иногда подразумевается только один особый вид моделей, содержащих сугубо однозначное прямое математическое описание в виде аналитических зависимостей или вычислительных алгоритмов – т.е. детерминированная математическая модель, с помощью которой при одних и тех же исходных данных можно получить только один и тот же результат. Наибольшее распространение получили детерминированные модели, устанавливающие связь с параметрами оригинала при помощи коэффициентов пропорциональности, всех одновременно равных **единице**. Используемое такой моделью математическое описание естественно рассматривать как описание непосредственно оригинала – и это верно: у модели и оригинала в этом случае существует одно общее математическое описание. В силу такой кажущейся простоты неискушенный инженер воспринимает и модель уже не как модель, а как оригинал (!). На самом деле такая математическая модель является все же моделью со всеми условностями, абстракциями, предположениями, упрощениями, положенными в ее основу. Возникает желание "упростить" процесс добротного моделирования, что в принципе невозможно, так как модель или соответствует оригиналу, или ее нет вообще. Пренебрежительное отношение к этому провоцирует множество ошибочных выводов в прикладных исследованиях, и полученные результаты не согласуются с реальностью.

Задание

Привести и прокомментировать примеры видов взаимосвязей параметров объекта в его математическом описании. Привести примеры видов математических моделей.

Практическая работа №2

Моделирование случайных чисел

Теоретический материал

Случайным опытом или экспериментом называется процесс, при котором возможны различные исходы, так что нельзя заранее предсказать, каков будет результат. Величина $\mathbf{X}=\{x_i\}=x_1, x_2, \ldots, x_n$, 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 $i = 1, 2, \ldots, n$. Любая функция от x_i будет также случайной величиной.

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

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:

Табл.1

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:

$x_0 = 0{,}9876, x_0^2 = 0{,}97\underline{5353}76,$

$x_1 = 0{,}5353, x_1^2 = 0{,}28\underline{6546}09,$

$x_2 = 0{,}6546$ 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 $0 \le F(x) \le 1$;

2) если $x_1 \le x_2$, то $F(x_1) \le F(x_2)$, т.е. $F(x)$ − неубывающая функция;

3) $F(x) \to 0$ при $x \to -\infty$, $F(x) \to 1$ при $x \to \infty$.

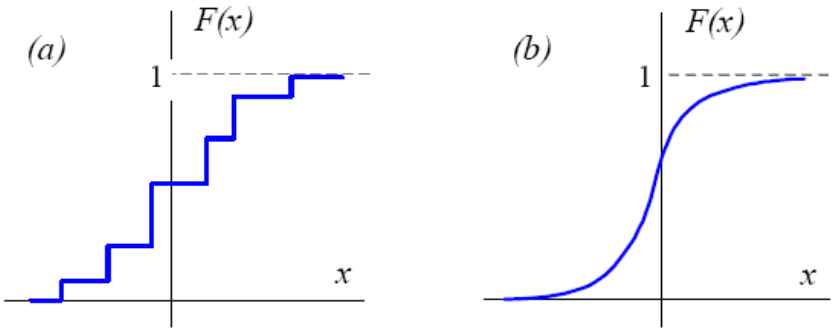


Рисунок 2 – Графическое изображение функции распределения вероятностей

Если случайная величина дискретна, то ее функция распределения представляет собой ступенчатую функцию (рисунок 2, *a*), а у непрерывных случайных величин функция распределения также непрерывна (рисунок 2, *b*).

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) $f(x) = F'(x) = \dfrac{dF(x)}{dx}$; $F(x) = \int\limits_{-\infty}^{x} f(u)du$;

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

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

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.

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:

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

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

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

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:

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

Для непрерывных случайных величин будем иметь

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

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:

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

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 x.

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

[1] Часто используют также эквивалентный термин ***квадратичное отклонение.***

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) приближенно 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 $N \to \infty$:

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

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

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:

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

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

Моделирование случайных событий

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 $\{x_i\} = x_1, x_2, \ldots, x_n$, представляющих собой реализации независимых, равномерно распределенных на интервале (0;1) случайных величин $\{r_i\} = r_1, r_2, \ldots, r_n$.

Непрерывная случайная величина R имеет равномерное распределение на интервале (a; b), если ее функции плотности вероятности (рис. 3, а) и распределения (рис. 3, б) задаются следующим образом:

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

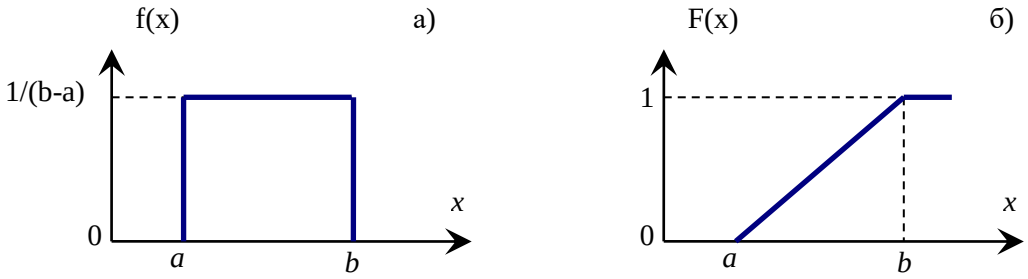


Рисунок 3 – Графическое изображение равномерно распределенной случайной величины X

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

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

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. 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,..., 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.).

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

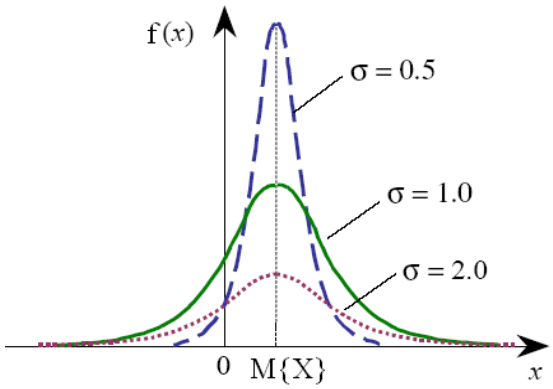


Рисунок 4 - График нормального распределения

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 4).

Так как моделирование любого нормального распределения с параметрами (M{X}, σ) может быть осуществлено по очевидной формуле X= M{X}+σ·R, где R={r_1, r_2, …, 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.

1. Разыграть 100 возможных значений случайной величины *X* распределенной нормально с параметрами M{X} = 0 и σ = 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:

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

Поступая аналогичным образом, найдем остальные возможные значения X_2, X_3, …, 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.

Выборочное среднее: $\overline{x_n} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

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

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→∞. *Выборочное среднеквадратичное отклонение* σ_n равно корню квадратному из выборочной дисперсии $\sigma_n = \sqrt{\sigma_n^2}$.

Получим:

$$M\{X\}^{\bullet} = \overline{x_n} \approx -0.05;\ \sigma^{\bullet} = \sqrt{D_n} \approx 1.04 .$$

Как видим, оценки удовлетворительны, т.е. $M\{X\}^{\bullet}$ близко к нулю, а $\sigma^{\bullet}$ 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: $r_i = \frac{X_i - M\{X\}}{\sigma}$.

Моделирование дискретных случайных величин

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

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

Или

Табл.2

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 Δ_1, Δ_2,..., Δ_n, длины которых равны соответствующим вероятностям P_1, P_2, ..., 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* $r_i(0 \le r_i < 1)$*, 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:

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

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

Табл.3

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.

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 $f(x) \le M$. Тогда, используя пару равномерно распределенных на интервале (0;1) случайных чисел R, осуществляем следующие действия для

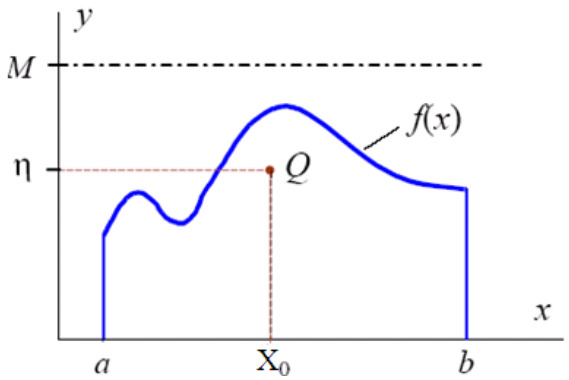


Рисунок 5 – Графическое изображение метода Неймана

розыгрыша (моделирования) значения X:

1. Разыгрываем два значения r_1 и r_2 случайной величины R и строим случайную точку *Q* с координатами (см. рисунок 5):

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

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.

Проверка гипотезы о законе распределения методом гистограмм

Пусть в результате эксперимента получено n значений $x_1, x_2, \ldots, x_n$ 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 $\nu=\sum_{j=1}^{L}\nu_j$. Относительная частота попадания случайных чисел последовательности $\{x_i\}$ в каждый из подынтервалов будет равна ν_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 ν_j/n. Очевидно, что если числа $\{x_i\}$ принадлежат псевдослучайной равномерно распределенной последовательности, то при достаточно больших n экспериментальная гистограмма (ломаная линия на рис. 6) приблизится к теоретической прямой $p_j=1/L$.

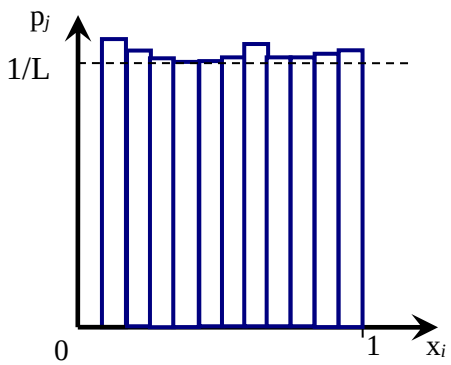


Рисунок 6 – Проверка равномерности последовательности

Гистограмма служит приближением к неизвестной плотности случайной величины *X*.

Площадь гистограммы, заключенная между x_i и x_{i+1}, дает приближенное значение вероятности $P\{x_i<X<x_i+1\}$.

Задание

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.

Табл. 4 – Частотная таблица

Интервал	Кол-во СВ (частота попаданий), выпавших в данный интервал	Относительная частота попадания
Δ_1	ν_1	ν_1/n
Δ_2	ν_2	ν_2/n
…	…	…
Δ_L	ν_L	ν_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.

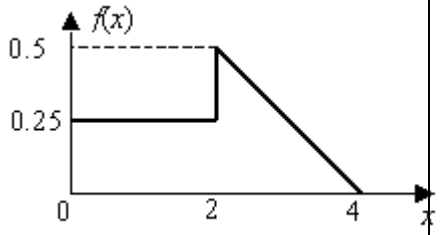
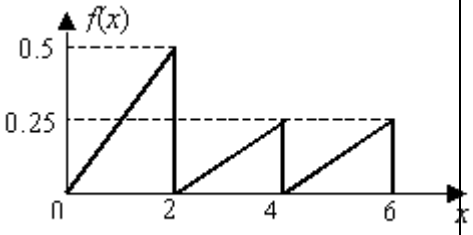
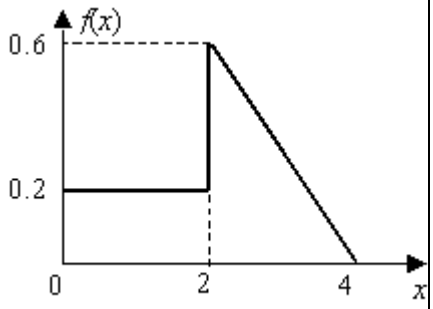
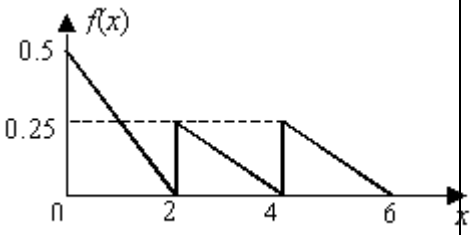
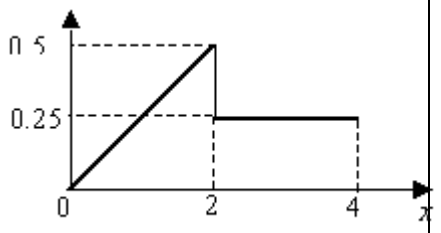
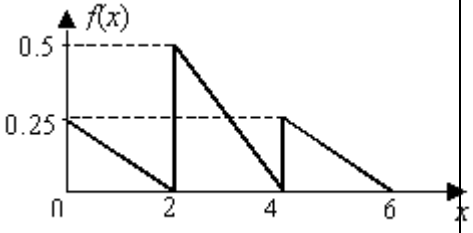
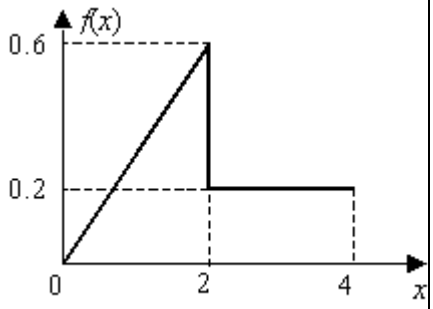
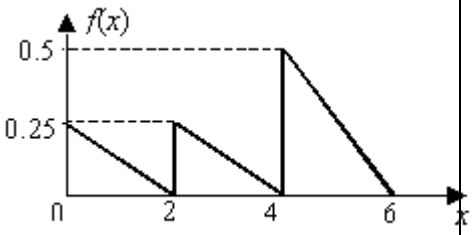
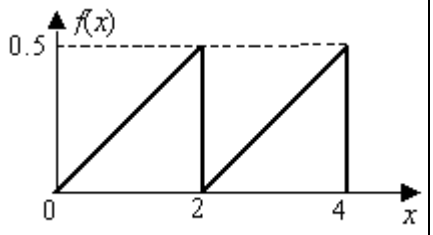
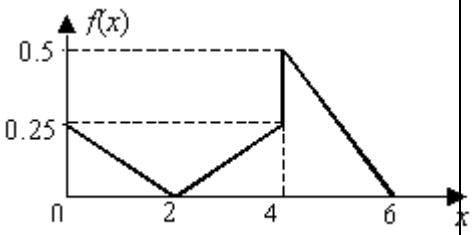
3.3. Построить частотную таблицу, вывести ее на экран.

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

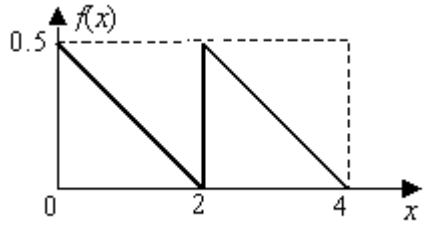
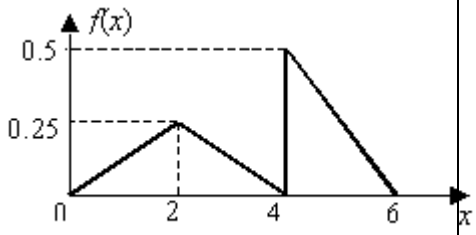
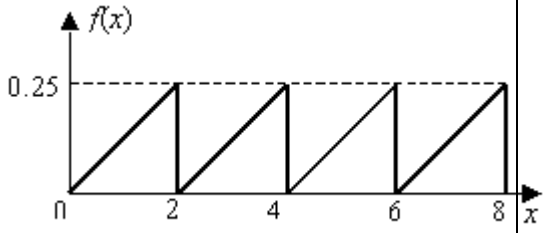
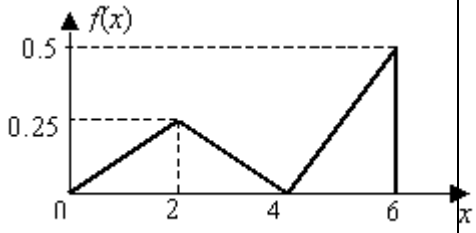
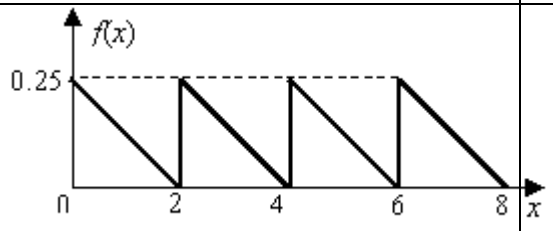
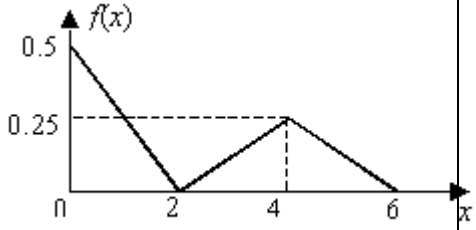
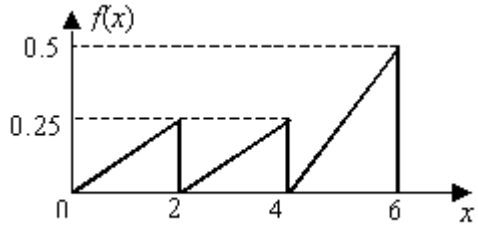
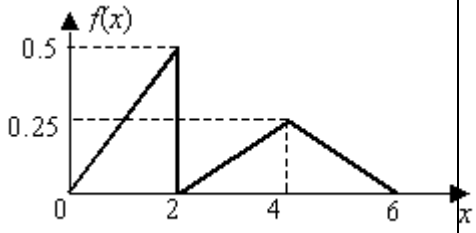
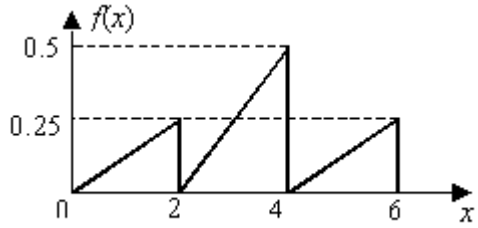
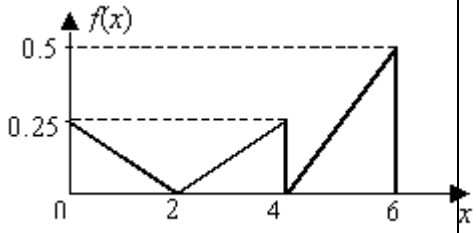
Таблица 5 – Таблица распределений

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 6 – Плотность распределения вероятности

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

Продолжение таблицы 6

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

Практическая работа № 3

Моделирование случайных чисел с заданным законом распределения

Теоретический материал

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 $\mathbf{r_1, r_2, \ldots, r_n}$ в последовательность $\mathbf{x_1, x_2, \ldots, x_n}$. В общем случае преобразование можно реализовать с помощью некоторой функции

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

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

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 7:

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

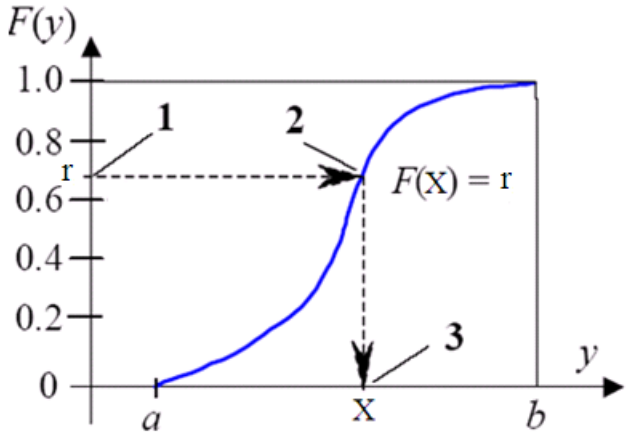


Рисунок 7 – Графическое изображение метода обратных функций

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·(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:

$$x_1=8\cdot0{,}11+2=2{,}88;\ x_2=8\cdot0{,}17+2=3{,}36;\ x_3=8\cdot0{,}66+2=7{,}28.$$

Пример 3. *Показательное распределение. Метод логарифма*[2].

Случайная величина X имеет экспоненциальный (показательный) закон распределения

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

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

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

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:

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

где $k=\sqrt{8/\pi}$.

Если воспользоваться (3) для случайной величины с плотностью

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

то получим для положительной полуоси методом обратных функций величину

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

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·X с нормальным распределением N (0;1) , $-\infty<x<\infty$, где S – случайная величина с распределением

[2] 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

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

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

1. Генерируем число R в интервале (0;1).

2. Определяем положительное наименьшее целое число R_i, для которого $R_i \le F(x_i)$, и возвращаем $X=x_i$.

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

$$X=\min\{x: F(x)\ge R\} \qquad (5)$$

Графически формула (5) представлена на рисунке 8.

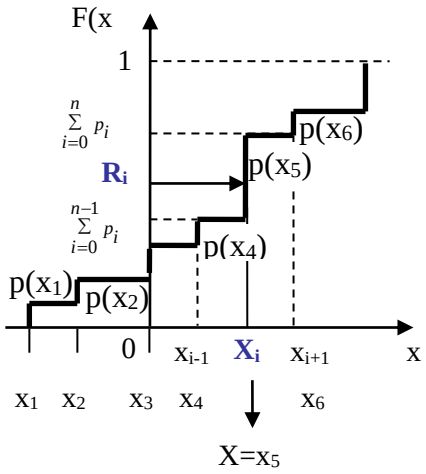


Рисунок 8 – Использование метода обратного преобразования для дискретных случайных величин

Таким образом, если явного выражения для обратной функции получить не удается, 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)*, соответствующих 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 9 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 $\left(\sum_{i=0}^{k-1} p_i;\ \sum_{i=0}^{k} p_i\right)$.

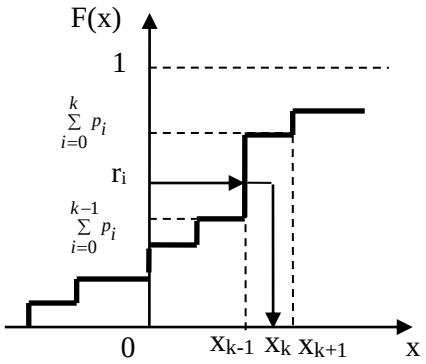


Рис. 9 – Реализация случайной величины X по эмпирической функции распределения

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.

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. $f(x)\le C$, 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 $z_1=a+r_1(b-a)$; $z_2=r_2c$;

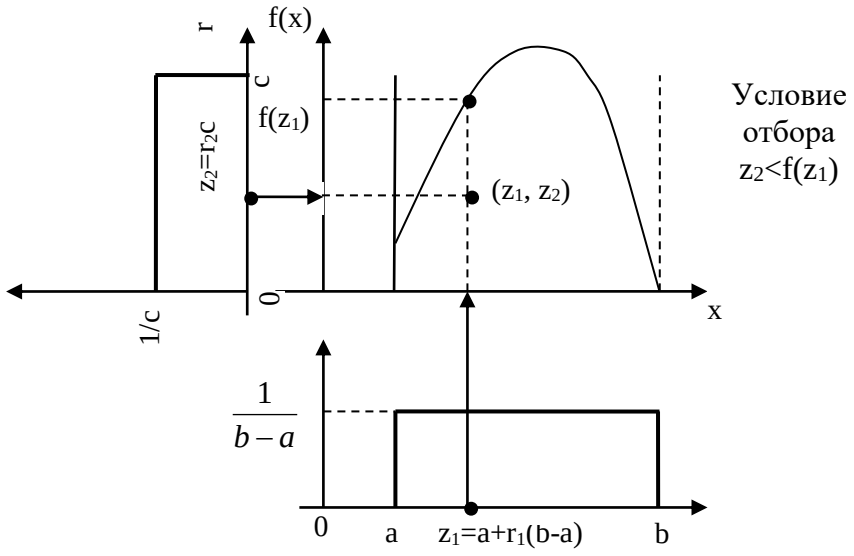


Рис.10– Графическое изображение метода отбора (процедура №1)

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).

Метод отбора – процедура № 2

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

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

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 11):

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).

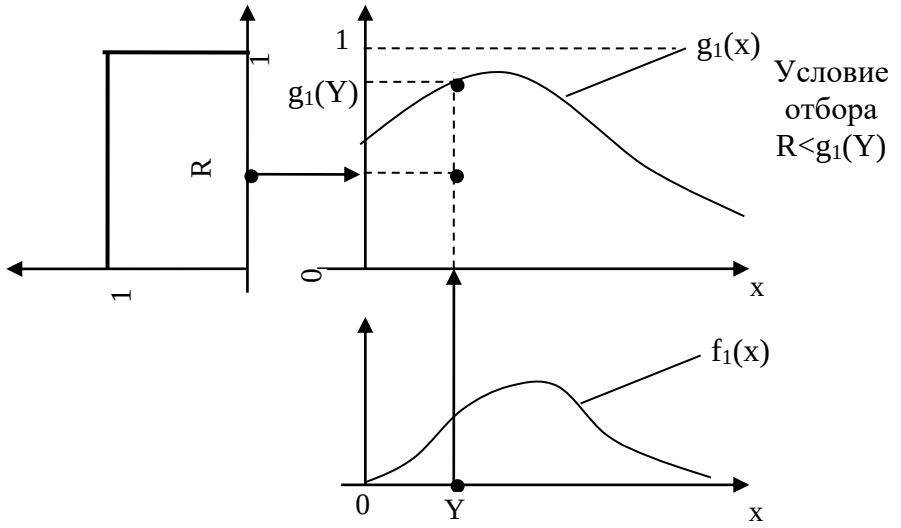


Рис.11 – Графическое изображение метода отбора (процедура №2)

Метод отбора – процедура № 3

Если в (6) 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).

Метод отбора – процедура № 4

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

$$f(x)=a_1(1-g_1(x))f_1(x), \qquad (7)$$

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

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

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).

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

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

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

$$f(x) = \sum_i p_i f_i(x) = \sum_i p_i f(x,i), \qquad (10)$$

где $p_i = p(Y=i)$.

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

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

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

Отсюда получаем алгоритм моделирования:

а) для распределения p0, p1, p2, … выбираем Y;

б) если Y=n, то $X=R^{1/(n+1)}$

Комбинация метода суперпозиций и метода отбора

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

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

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

$\omega i(x)$ – некоторые известные плотности вероятностей,

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

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, …;

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

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

Пусть

$$\begin{cases} p_1=\sqrt{\dfrac{2}{\pi}}; \\ \omega_1(x)=\begin{cases}1, \text{при } 0\le x\le 1;\\ 0, \text{при } x<0,\ x>1;\end{cases} \\ g_1(x)=e^{-\frac{x^2}{2}}; \end{cases} \qquad \begin{cases} p_2=\dfrac{1}{\sqrt{2\pi}}; \\ \omega_2(x)=\begin{cases}2e^{-2(x-1)}, \text{при } x>1;\\ 0, \text{при } x\le 1;\end{cases} \\ g_2(x)=e^{-\frac{(x-2)^2}{2}}; \end{cases}$$

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), $-\infty<x<\infty$.

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.

Биноминальное распределение

Биноминальным 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

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

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 биноминальному закону распределения, в виде суммы n независимых случайных величин X_i, каждая из которых имеет распределение

X_i	1	0
p_i	p	q=1-p

(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, …, r_n проверяется выполнение неравенства $r_i<p$(i=1, 2, …, n). Если неравенство выполняется, принимается $X_i=1$, в противном случае $X_i=0$.

в) вычисляется сумма $X_1 + X_2 + … + X_n$ 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, …, s_k, которые являются реализацией биноминально распределенной случайной величины.

Распределение Пуассона

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

Случайная величина X имеет распределение Пуассона, если

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

где λ - параметр распределения.

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

$$\sum_{i=1}^{n} x_i \le \lambda < \sum_{i=1}^{n+1} x_i, \tag{16}$$

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

$$\prod_{i=1}^{n+1} r_i \le e^{-\lambda} < \prod_{i=1}^{n} r_i, \tag{17}$$

где $\prod_{i=1}^{n} r_i$ – произведение всех r_i, таких, что значение i – целое и выполняется соотношение i=1…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, …, r_n независимых случайных величин, равномерно распределенных на [0; 1);

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

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

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 $N\to\infty$, $p\to 0$, $N_p=\lambda$ 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 y_j фактического наступления события A в серии с номером j. Числа y_j будут приближенно следовать закону Пуассона, причем тем точнее, чем больше N. Практически N должно выбираться таким образом, чтобы p=0.1 + 0.2.

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

$$y_i=\mu+\sigma x_i,\ i=1, 2, \ldots, \qquad (18)$$

при любом μ и σ>0 можно затем получить последовательность случайных чисел y_1, y_2, …, отвечающих распределению N (μ; σ^2) с математическим ожиданием M(Y)=μ 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, …, n) приближается к нормальному 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, …, 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

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

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 (μ=0, σ=1) из двух стандартных случайных чисел (R_1 и R_2 на интервале от 0 до 1) с помощью следующих уравнений:

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

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.

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

Моделирование (бета-распределения)

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

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

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

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

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

где R_1, R_2, …, 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

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

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

[m] – обозначает целую часть m.

С учетом последнего распределения получаем формулу для бета-распределения:

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

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).

Моделирование гамма-распределения

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

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

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:

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

Первичная обработка статистических данных

Выборка. Эмпирическая функция распределения. Гистограмма

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

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

Числа $x_1,,\ldots,x_n$, $x_i \in \mathbf{X}$, $i = \overline{1,n}$, 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 7 – Параметры выборки

Параметр	Обозначение	Определение
Выборочные данные	x_i, где i = 1, ... n	Наблюденные значения случайной величины
Объем 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, …, X_n), координаты которого являются независимыми, распределенными 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.

Табл. 8 – 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 ($y_1, \ldots, y_r$) ($r<n$) – различные значения среди $x_1, \ldots, x_n$; 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 $\left(y_i, p_i^*\right)$, $i = \overline{1,r}$ называют иногда *эмпирическим законом распределения*, а саму таблицу 7 – *таблицей частот*. Выборочные значения $x_1, \ldots, x_n$, упорядоченные по возрастанию, носят название *вариационного ряда*:

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)},$$

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

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

Эмпирической функцией распределения, соответствующей выборке $x_1, \ldots, x_n$, называется функция

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

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

- если все значения $x_1, \ldots, x_n$ различны, то $F_n^*(x)=i/n$ при $x \in [x_{(i)}, x_{(i+1)})$, $x_{(0)}=-\infty$, $x_{(n+1)}=\infty$;
- если $y_1, \ldots, y_r$ –различные значения среди $x_1, \ldots, x_n$, 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 $x_1, \ldots, x_n$ – 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 $x_1, \ldots, x_n$ строят вариационный ряд $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$.
2. Промежуток $[x_{(1)}, x_{(n)}]$ разбивают точками $u_0 = x_{(1)}, u_1, \ldots, u_L = x_{(n)}$: $u_0<u_1<\ldots<u_L$ на L непересекающихся интервалов $J_k = [u_{k-1}, u_k)$ (на практике $L<<n$).
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 8).

Табл.9 – 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 $\sum_{k=1}^{L} \nu_k = n, \quad \sum_{k=1}^{L} p_k^* = 1$. 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 Δu=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≥1000 L=11…15;

n≥400 L=10;

n≥200 L=9;

100<n<200 критерий применяют в исключительных случаях с числом интервалов L=7…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:

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

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 $x_1, \ldots, x_n$:

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

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:

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

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

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

Проверка статистических гипотез

Статистической гипотезой называют любое утверждение о виде или свойствах наблюдаемых в эксперименте случайных величин. Правило, позволяющее 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 величии, то она называется *простой*, в противном случае – *сложной*.

Пусть относительно наблюдаемой случайной величины X сформулирована некоторая гипотеза H_0; $x_1, \ldots, x_n$ – выборка объема n, являющаяся реализацией случайного вектора (X_1, ..., 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 $T = T(X_1, \ldots, X_n)$ (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 $K = \{t: t = T(x_1, \ldots, x_n)\}$ возможных значений статистики T, так чтобы $P\{T\{X_1, \ldots, X_n) \in K_\alpha / H_0\} \le \alpha$. Критерий согласия имеет следующий вид:

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

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

Проверка гипотезы о виде распределения

Пусть $x_1, ..., x_n$ – 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 критерии χ^2 **Пирсона**. Чтобы воспользоваться критерием χ^2 Пирсона, выборочные данные $x_1, ...,x_n$ 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; $\nu_1,\ldots,\nu_L$ – частоты попадания выборочных значений в интервалы $J_1,...,J_L$ соответственно ($\nu_1 + ... + \nu_L = n$). Обозначим p_k теоретическую (соответствующую H_0) вероятность попадания случайной величины X в интервал $J_k : p_k = P\{u_k \le X < u_{k+1}\} = F(u_{k+1}) - F(u_k), k = \overline{1, L}$, 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 $\chi_n^2 = \sum_{k=1}^{L} \frac{(\nu_k - np_k)^2}{np_k}$,

где L – количество интервалов гистограммы, построенной по таблице 3;
ν_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 $F_n^*(x)$ (ν_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 ν_k/n сближаются с вероятностями p_k при $n \to \infty$, $k = \overline{1, L}$, 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 $K_\alpha = \{t \ge t_\alpha\}$, где $t = \chi_n^2(x_1, ...,x_n)$ – 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 $n \to \infty$ предельное распределение хи-квадрат с (L-1) степенью свободы $\chi^2(L\text{-}1)$.

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

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 $\chi^2_n (x_1, \ldots, x_n) = t$.

6. По таблице распределения для вычисленного значения χ^2_n 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 $v_k=m_k$, $k = \overline{1,L}$, где m_k – частота выборочного значения x_k.

Задание

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

б) Моделирование случайной величины $X = \{x_i\} = x_1, x_2, \ldots, x_n$ выполнить с помощью последовательности чисел $R = \{r_i\} = r_1, r_2, \ldots, r_n$, 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 $X = \{x_i\} = x_1, x_2, \ldots, x_n$ 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.

Табл.10 – Варианты заданий

Вариант	Закон распределения	Способ построения
1	Экспоненциальный, λ=2	Метод обратных функций, формула (2)
2	Нормальный, N(0;1)	Метод отбора
3	Экспоненциальный, λ=3	Метод отбора
4	Распределение Пуассона, формула (15); λ=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); λ=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 10

12	Бета-распределение, формула (22); m=5, p=2.3	Метод Йонка
13	Дискретное распределение; k=10; 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; σ=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; σ=3	Процедура 2
19	Закон Релея: $f(y)=\frac{y}{\sigma^2}\exp\left(-\frac{y^2}{2\sigma^2}\right)$, $y\geq 0; \sigma^2=0.5$	Процедура 1
20	Экспоненциальный; λ=0.4	Метод обратных функций, формула (2)
21	Нормальный; N(0; 1)	Процедура Марсальи и Брея, формула (21)
22	Дискретное распределение; k=12; 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); λ=4	Формула (26)
24	Нормальный, N(2, 1)	Аппроксимация распределения; формула (19)

Лабораторная работа № 4
Линейная и квадратичная аппроксимация статистических данных
Теоретический материал
Линейная аппроксимация статистических данных

Постановка задачи. Пусть требуется определить функциональную зависимость между двумя величинами X и Y. Они могут быть параметрами некоторого либо физического процесса, либо экономического показателя, либо других явлений природного характера. Для определения этой зависимости проводят наблюдения или эксперименты, результаты которых записывают в виде таблицы значений

рассматриваемых величин. Используя эти данные, требуется определить математическую зависимость между этими величинами.

Математическая модель задачи. Искомая зависимость записывается в виде линейной функции $y = a \cdot x + b,$ где a и $b -$ неизвестные пока параметры, значения которых должны быть определены.

Метод решения задачи. Данная задача решается известным методом наименьших квадратов, сущность которого заключается в следующем. График аппроксимирующей функции должен проходить очень близко от статистических точек. Это условие может быть осуществлено, если следующая функция имеет наименьшее значение:

$$U(a,b) = \sum_{k=1}^{n} [y_k - (a \cdot x_k + b)]^2 \Rightarrow \min .$$

Здесь $n -$ количество статистических точек, $(x_k, y_k) -$ координаты статистических точек. Необходимым и достаточным условием минимума данной функции может быть записано:

$$\frac{\partial U}{\partial a} = 0, \quad \frac{\partial U}{\partial b} = 0.$$

Из этих условий будут определены следующие формулы для определения значений неизвестных параметров a и b:

$$a = \frac{n \cdot S_4 - S_1 \cdot S_2}{n \cdot S_3 - S_1^2}, \quad b = \frac{S_3 \cdot S_2 - S_1 \cdot S_4}{n \cdot S_3 - S_1^2},$$

где $S_1 = \sum_{k=1}^{n} x_k, \quad S_2 = \sum_{k=1}^{n} y_k, \quad S_3 = \sum_{k=1}^{n} x_k^2, \quad S_4 = \sum_{k=1}^{n} x_k y_k.$

Алгоритм решения задачи:

- ввод массивов $x_k, y_k, k = 1,2,\cdots,n,$ в память компьютера;
- определение значений следующих сумм $S_1, S_2, S_3, S_4;$
- определить значения искомых параметров a и b.

Пример. Использовать статистические данные, приведенные в таблице1.

Табл. 11- Статистические данные значений величин X и Y

x_k	50	60	70	80	90	100	110	120	130	140
y_k	65	75	95	100	115	130	155	170	180	190
x_k	150	160	170	180	190	200	210	220	230	240
y_k	200	210	235	250	265	290	310	350	370	400

Результатом решения задачи является линейная аппроксимирующая функция $y = a \cdot x + b$ и ее график в виде прямой линии.

Квадратичная аппроксимация статистических данных

Постановка задачи. Пусть требуется определить функциональную зависимость между двумя величинами X и Y в виде квадратичной функции $y = a \cdot x^2 + b \cdot x + c$.
Для определения этой зависимости используются данные эксперимента или статистики. Этот процесс называется квадратичной аппроксимацией.
Математическая модель задачи. Искомая зависимость записывается в виде функции $y = a \cdot x^2 + b \cdot x + c,$ где $a, b, c -$ неизвестные пока параметры, значения которых должны быть определены.
Метод решения задачи. Данная задача решается методом наименьших квадратов. Используется условие минимума следующей функции:

$$U(a,b,c) = \sum_{k=1}^{n} [y_k - (a \cdot x_k^{\,2} + b \cdot x + c)]^2 \Rightarrow \min.$$

Здесь $n -$ количество статистических точек, $(x_k, y_k) -$ координаты статистических точек. Необходимым и достаточным условием минимума данной функции имеет вид:

$$\frac{\partial U}{\partial a} = 0, \quad \frac{\partial U}{\partial b} = 0, \quad \frac{\partial U}{\partial c} = 0.$$

Из этих условий будут получена следующая система алгебраических уравнений относительно неизвестных параметров a, b и c:

$$a \cdot \sum_{k=1}^{n} x_k^{\,4} + b \cdot \sum_{k=1}^{n} x_k^{\,3} + c \cdot \sum x_k^{\,2} = \sum_{k=1}^{n} y_k x_k^{\,2},$$

$$a \cdot \sum_{k=1}^{n} x_k^{\,3} + b \cdot \sum_{k=1}^{n} x_k^{\,2} + c \cdot \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k x_k,$$

$$a \cdot \sum_{k=1}^{n} x_k^{\,2} + b \cdot \sum_{k=1}^{n} x_k + c \cdot n = \sum_{k=1}^{n} y_k.$$

Алгоритм решения. Полученную систему алгебраических уравнений можно решить с помощью MS Excel или с помощью программы на алгоритмическом языке. При этом используется один из существующих методов решения системы алгебраических уравнений. Алгоритм решения этой системы состоит из следующих этапов:

- ввод массивов $x_k, y_k, k = 1,2,\cdots,n,$ в память компьютера;
- определение значений следующих сумм $S_1, S_2, S_3, S_4, S_5, S_6, S_7;$
- решить систему алгебраических уравнений;
- определить значения искомых параметров $a, b, c.$

Здесь суммы определяются по формулам:

$$S_1 = \sum_{k=1}^{n} x_k, \; S_2 = \sum_{k=1}^{n} x_k^{\,2}, \; S_3 = \sum_{k=1}^{n} x_k^{\,3}, \; S_4 = \sum_{k=1}^{n} x_k^{\,4}, \; S_5 = \sum_{k=1}^{n} y_k,$$

$$S_1 = \sum_{k=1}^{n} x_k y_k, \qquad S_1 = \sum_{k=1}^{n} x_k^{\,2} y_k.$$

Пример. В качестве статистических данных использовать данные, приведенные в таблице 1.
Результатом решения задачи является квадратичная функция и ее график.

Задание

1) Изучить теоретические материалы о линейной и квадратичной аппроксимации данных эксперимента или статистики и ознакомиться с методическими указаниями по выполнению данной работы.
2) Сформулировать постановку задачи.
3) Изучить метод решения задачи и разработать алгоритм ее решения.
4) Решить задачу с помощью MS Excel или программы на алгоритмическом языке.
5) Построить график аппроксимирующей функции.

Список литературы

1. *Беляева М.А.* Моделирование систем : методич. указания по выполнению лабораторных работ / М.А. Беляева. — М. : МГУП, 2011.

2. *Беляева М.А.* Моделирование систем : электронное учеб. пособие / М.А. Беляева. — М. : МГУП, 2010.

3. *Беляева М.А.* Многокритериальная оптимизация процессов тепловой обработки мясных полуфабрикатов при ИК-энергоподводе : автореферат на соиск. уч. ст. д.т.н. / М.А. Беляева. — М. : МГУПБ, 2009. — 50 с.

4. Закгейм А.Ю. Введение в моделирование химико-технологических процессов. Математическое описание процессов. М.:Химия, 2008 г.

5. Воронин, А. И. Основы научных исследований : учеб. пособие (курс лекций) / А. И. Воронин ; Мин-во образования и науки Рос. Федерации, ГОУ ВПО Сев.-Кав. гос. техн. ун-т . - Ставрополь : Издательство СевКавГТУ, 2008. - 133 с

6. Королев А.Л. Компьютерное моделирование – М.: БИНОМ. Лаборатория знаний, 2010.-230 с.

7. *Васильев К.К.* М.Н. Служивый : учеб. пособие «Математическое моделирование систем связи» / К.К. Васильев. — УлГТУ, 2008 — 168 с.

СОДЕРЖАНИЕ

Методические указания

Редактор
Подписано в печать
Формат 60х84 1/16 Бумага газетная.
Печать офсетная 2.0 п.л., уч.-изд.л.
Тираж 100 экз.
Заказ №
Редакционно-издательский отдел ДГТУ

ИПЦ ДГТУ
367015 Махачкала, пр.Шамиля, 70