

Министерство образования Российской Федерации
Томский политехнический университет

Ю.И. ГАЛАНОВ

Статистическое моделирование

Методические указания по выполнению лабораторных работ и
индивидуального задания по математической статистике

Издательство ТПУ
Томск 2004

Введение

При изучении эффективности различных статистических процедур необходимо иметь эталонные наборы выборок различного объема с известными законами распределения. Обработывая эталонные данные по какому-либо алгоритму, мы имеем возможность сравнить предсказания теории с известными параметрами эталона, установить границы надежности статистических выводов, выработать, наконец, особую «статистическую интуицию», иммунитет против необоснованных заключений на основании ненадежных статистических данных.

Для генерации выборки с заданным законом распределения достаточно уметь создавать какое-нибудь одно «стандартное» распределение, а затем подбирать такую функцию от него, которая имела бы требуемый закон распределения.

В качестве исходного материала для моделирования распределений применяют случайную величину, равномерно распределенную на отрезке $[0,1]$. Значения такой случайной величины называют *случайными числами*.

Существует много способов получения случайных чисел на ЭВМ, включая встроенные генераторы, а также специальные алгоритмы [1,2].

Например, *линейная конгруэнтная последовательность* создается следующим образом. Задаем 4 «магических» числа: X_0 – начальное значение, a – множитель, c – приращение, m – модуль.

Искомая последовательность случайных чисел получается из соотношения:

$$X_{n+1} = (aX_n + c) \bmod m, n \geq 0$$

Такие последовательности через некоторый период повторяются, поэтому их называют псевдослучайными. Подбирая параметры последовательности можно значительно увеличить её период. Псевдослучайные последовательности широко применяются в компьютерных расчетах.

Во всех алгоритмических языках программирования и специальных вычислительных средах предусмотрены функции, возвращающие значения случайных чисел. В среде MathCAD есть функция $rnd(x)$, возвращающая значение случайной величины, равномерно распределенной на отрезке $[0, x]$ [6]. Мы рассмотрим некоторые простейшие алгоритмы преобразования случайных последовательностей в выборки из генеральных совокупностей, обладающих необходимыми законами распределения.

Глава 1.

Моделирование дискретных случайных величин

1.1. Случайные испытания с двумя исходами

Опыт с двумя исходами $\{A, \bar{A}\}$, независимо от природы исследуемых явлений, полностью определяется заданием вероятности p события A . В этом смысле события, происходящие с равной вероятностью, являются эквивалентными и моделируют одно другое. Так, например, опыт по выниманию шаров из урны является идеальной моделью испытаний с двумя исходами. Для этого достаточно подобрать необходимый состав шаров и схему выбора.

Идею моделирования случайных событий на компьютере легко уяснить с помощью понятия геометрической вероятности. Пусть мы случайным образом бросаем точку на отрезок единичной длины и нас интересует попадание точки на отрезок длиной p . Тогда вероятность попадания точки на этот отрезок и будет численно равна его длине p , то есть события $\{A\}$ и $\{rnd(1) < p\}$ равносильны (рис.1).

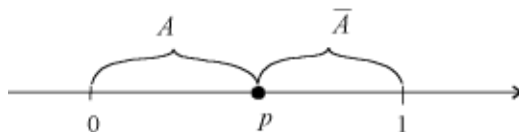


Рис. 1.1. Равносильность событий A и $\{rnd(1) < p\}$

Для подсчета числа появлений события A в серии из n испытаний введем индикатор события $J(p)$ — случайную величину, равную 1, если событие произошло и равную 0 в противном случае.¹

¹ При программировании в среде MathCad индикатор легко реализовать разными методами: использовать логические операции, возвращающие единицу, если результат операции «истина» и ноль — если «ложно», а также встроенные функции Хевисайда и Кронеккера.

$$J(p) := \text{if}(\text{rnd}(1) < p, 1, 0), \quad (1.1.)$$

Число m появлений события в серии из n испытаний можно получить просуммировав индикаторы событий для каждого испытания.

$$n := 10 \quad i := 0.. n - 1 \quad m := \sum_i J(p) \quad (1.2.)$$

1.2. Дискретная случайная величина с заданным законом распределения

Пусть закон распределения дискретной случайной величины задан таблицей.

X	x_1	x_2	\dots	x_i	\dots	x_n	\dots
P	P_1	P_2	\dots	P_i	\dots	P_n	\dots

Таблица 1.1. Ряд распределения дискретной случайной величины

Моделью такой случайной величины является опыт с k исходами. Разобьем отрезок $[0, 1]$ на k отрезков точками u_0, u_1, \dots, u_k так, чтобы выполнялись условия:

$$u_i - u_{i-1} = p_i, (i = 1 \dots k), \quad \sum_{i=1}^k p_i = 1 \quad (1.3.)$$

При моделировании единичного испытания будем полагать, что случайная величина приняла значение x_i , если случайное число попало в i -тый интервал. Отметим, что последовательность точек u_0, u_1, \dots, u_k есть не что иное, как последовательность значений функции распределения.

1.2. Дискретная СВ с заданным законом распределения

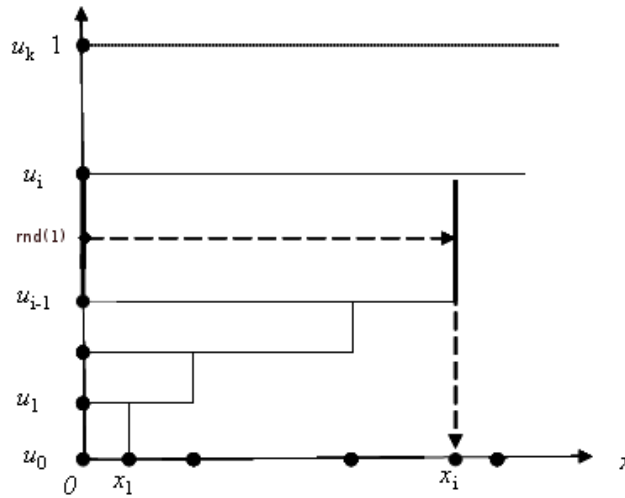


Рис. 1.2. Моделирование исхода случайного эксперимента

Примеры

Моделирование опыта с четырьмя исходами.

Задаем значения случайной величины и их вероятности:

$$p_1 := 0.1 \quad p_2 := 0.3 \quad p_3 := 0.4 \quad p_4 := 0.2$$

$$X_1 := 1 \quad X_2 := 3 \quad X_3 := 4.5 \quad X_4 := 12$$

Задаем вектор индикаторов событий:

$$K := 4 \quad i := 1..K \quad j := 2..K-1 \quad \sum_i p_i = 1$$

$$U_1 := p_1 \quad U_j := U_{j-1} + p_j \quad U = \begin{bmatrix} 0.1 \\ 0.4 \\ 0.8 \end{bmatrix}$$

$$M(x) := \text{if} \left[x < U_1, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \text{if} \left[x < U_2, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \text{if} \left[x < U_3, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right] \right] \right]$$

Создаем последовательность N индикаторов:

$$N := 50 \quad k := 1.. N \quad L^{(k)} := M(\text{rnd}(1)) \quad J := L^T \cdot X$$

Вектор J содержит выборку из заданного полиномиального распределения. Вектор частот B служит оценкой вектора вероятностей P :

$$B := \frac{\sum_k L^{(k)}}{N} \quad B^T = [0.04 \quad 0.3 \quad 0.46 \quad 0.2] \quad \sum B = 1$$

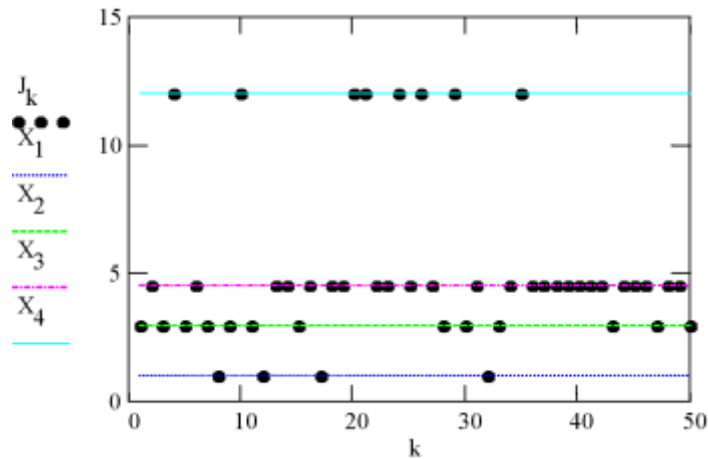


Рис. 1.3. Выборка из полиномиального распределения

Моделирование случайного блуждания материальной частицы по плоской сетке

Рассмотрим материальную частицу, способную в дискретные моменты времени случайным образом перемещаться в одном из четырех направлений по узлам квадратной плоской сетки. Данный пример является модификацией предыдущего.

Введем функцию случайного смещения

1.2. Дискретная СВ с заданным законом распределения

$$D(p) := \text{if} \left[p < \frac{1}{4}, \left[\begin{array}{c} 1 \\ 0 \end{array} \right], \text{if} \left[p < \frac{1}{2}, \left[\begin{array}{c} -1 \\ 0 \end{array} \right], \text{if} \left[p < \frac{3}{4}, \left[\begin{array}{c} 0 \\ -1 \end{array} \right], \left[\begin{array}{c} 0 \\ 1 \end{array} \right] \right] \right] \right]$$

Зададим вектор начального смещения:

$$R^{(0)} := \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Зададим число смещений и текущее положение частицы:

$$N := 1000; \quad i := 1..N; \quad R^{(i)} := R^{(i-1)} + D(\text{rnd}(1))$$

Укажем начальное и конечное положение частицы:

$$K^{(0)} := R^{(0)} \quad K^{(1)} := R^{(N)}$$

Результаты моделирования приведены на рисунках.(1.4. а, в.).

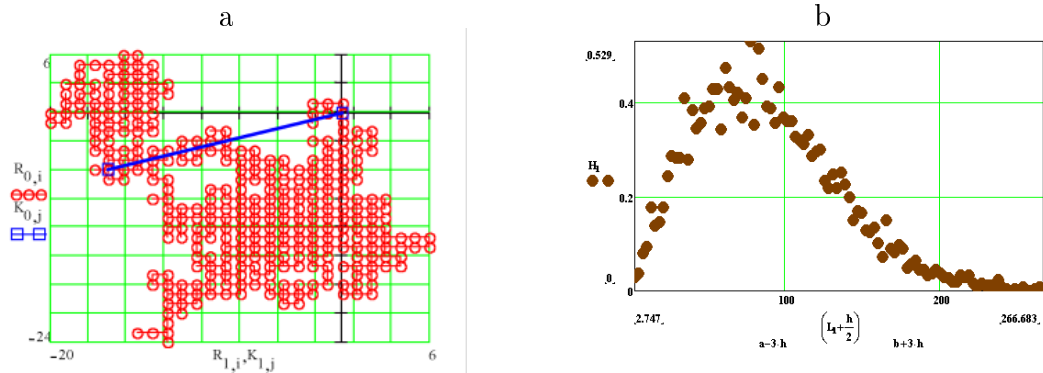


Рис. 1.4. а – Смещение одной точки за 1000 шагов. в – Плотность распределения длин смещения 5000 точек за 10 000 шагов.

При большом числе исходов опыта для построения индикаторов событий получаются громоздкие формулы. В этом случае можно использовать другой алгоритм идентификации исхода.

1.3. Программа моделирования опытов с большим числом исходов

Моделируемое распределение может быть задано либо рядом, либо функцией распределения.

В начале смоделируем исходные данные:

$$n := 20 \quad i := 0.. n - 1 \quad x_i := 2 \cdot i + rnd(1) \quad j := 0.. n - 1$$

Значения вероятностей вводим в виде компонент вектора p_i (в тексте не показано). По заданному закону распределения строим теоретическую функцию распределения и строим ее график (рис. 1.5):

$$efr_i := \sum_j p_j \cdot (x_i \geq x_j)$$

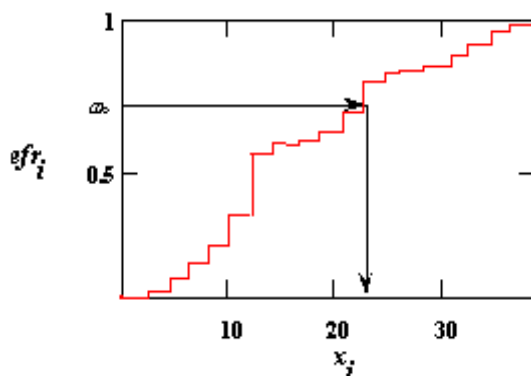


Рис. 1.5. Функция распределения моделируемой дискретной случайной величины и схема ее моделирования.

Задаем объем выборки, номер элемента и набор случайных чисел a_k :

$$K := 300 \quad k := 0.. K - 1 \quad a_k := rnd(1)$$

Создадим массив индикаторов событий путем подсчета максимального номера значения случайной величины, удовлетворяющего

1.4. Биномиальное распределение

условию $a_k > efr_i$ и подсчитаем частоту событий, просуммировав индикаторы событий:

$$D_{\sum_i (a_k > efr_i), k} := 1 \quad S := \frac{1}{K} \cdot \sum_k D^{(k)} \quad s := 0.. \text{last}(S)$$

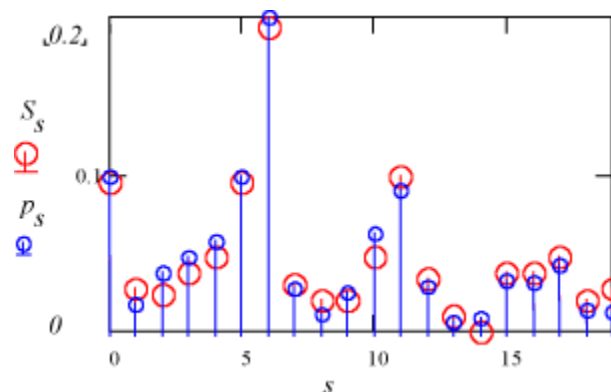


Рис. 1.6. Модельные и экспериментальные распределения.

Рассмотренный алгоритм моделирования дискретных распределения является универсальным, его можно применять и для моделирования непрерывных распределений, если их функцию распределения предварительно заменить кусочно-постоянной функцией.

1.4. Биномиальное распределение

Биномиальное распределение можно смоделировать с помощью рассмотренного выше универсального алгоритма, однако мы в данном разделе сделаем это исходя из «первых принципов» без предварительного задания функции распределения.

Биномиальное распределение описывает число m появлений события в серии из n испытаний, в каждом из которых событие появляется с вероятностью p .

Программа подсчета числа событий в серии:

$$n := 10 \quad i := 0.. n - 1 \quad m := \sum_i J(p)$$

Здесь $J(p)$ — определенный выше индикатор наступления события в отдельном испытании.

Программа моделирования выборки. Задаем объем выборки и проиндексируем её элементы:

$$K := 100 \quad j := 0.. K - 1$$

Количество испытаний в серии

$$n := 20 \quad i := 0.. n - 1$$

Создаем выборку В:

$$B_j := \sum_i J(p)$$

1.5. Распределение Пуассона

При моделировании будем использовать свойство пуассоновского процесса, состоящего в том, что время ожидания появления события имеет показательное распределение.

$$F_\tau(t) = 1 - e^{-\lambda \cdot t} \quad (1.4.)$$

Следовательно, последовательность наступления событий в пуассоновском процессе можно задать через *последовательность времен ожидания* этих событий. При этом надо проверять, что бы *суммарное время ожидания событий в цепочке не превышала единицы*.

Последовательность времен ожидания можно получить методом обратных функций (см. раздел 2.1)

$$\tau_i = -\frac{1}{\lambda} \cdot \ln(1 - u_i), \quad (1.5.)$$

где $u_i = rnd(1)$ — *случайные числа*, т. е. значения случайной величины, равномерно распределенной на $[0, 1]$.

Последовательность (1.5.) следует продолжать пока не нарушается условие:

$$\sum_{i=1}^j \tau_i = \sum_{i=1}^j \left(-\frac{1}{\lambda} \cdot \ln(1 - u_i)\right) \leq 1. \quad (1.6.)$$

1.5. Распределение Пуассона

Максимально возможное значение слагаемых в сумме (1.6.) и будет равно числу появления событий в данной серии, т.е. эти числа и есть значения случайной величины, имеющей распределений Пуассона.

Замечание. Программисты всегда старались и, надеюсь, стараются сделать свои программы как можно короче. Проследуем этому хорошему примеру. Во-первых, мы можем исключить одну операцию в вычислительной формуле, заменив выражение $1 - u_i$ просто на u_i , поскольку они имеют один и тот же закон распределения. Во-вторых, избавимся от операций логарифмирования в каждом слагаемом, для чего пропотенцируем выражение (1.6.).

В результате этих преобразований мы определим случайную величину

$$\xi = \max \left\{ j : \prod_{i=1}^j u_i \geq e^{-\lambda} \right\}, \lambda > 0, \quad (1.7.)$$

которая описывается распределением Пуассона. Элемент выборки можно получить последовательно увеличивая число членов (j) в произведении до тех пор, пока не нарушится условие

$$\prod_{i=1}^j u_i \geq e^{-\lambda} \quad , \quad (1.8.)$$

максимальное значение (j), удовлетворяющее этому условию и есть очередное значение случайной величины.

Программа создания выборки.

$$D(\lambda, N) := \left| \begin{array}{l} d \leftarrow \exp(-\lambda) \\ j \leftarrow 0 \\ \text{for } k \in 0 \dots N \\ \quad \left| \begin{array}{l} x \leftarrow \text{rnd}(1) \\ \text{while } x > d \\ \quad \left| \begin{array}{l} x \leftarrow x \cdot \text{rnd}(1) \\ j \leftarrow j + 1 \end{array} \right. \\ \quad P_k \leftarrow j \\ \quad j \leftarrow 0 \end{array} \right. \\ P \end{array} \right.$$

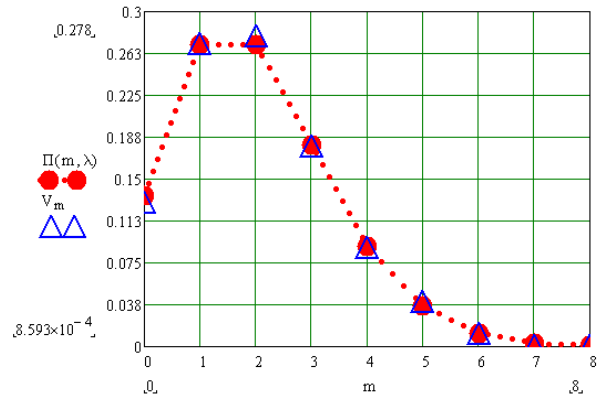


Рис. 1.7. Модельное распределение Пуассона и его выборочная оценка

Глава 2.

Моделирование непрерывных распределений

2.1. Применение обратной функции

Пусть $F_\xi(x)$ — непрерывная и монотонно возрастающая функция распределения случайной величины ξ . Тогда существует обратная ей функция $F^{-1}(y)$, определенная на отрезке $[0, 1]$. Пусть U — случайная величина, равномерно распределенная на $[0, 1]$. Тогда, на основании равносильности двух событий и можем записать

$$P(U_i \leq F(x)) = P(F^{-1}(U_i) \leq x) = P(x_i \leq x) = F(x), \quad (2.1.)$$

где $x_i = F^{-1}(U_i)$.

Следовательно, случайная величина X имеет распределение $F_\xi(x)$. Таким образом, если известен явный вид обратной функции $F^{-1}(y)$, то получаем следующий естественный алгоритм моделирования случайной величины с заданным законом распределения.

1. Генерируем очередное значение случайной величины U_i .
2. По формуле $x_i = F^{-1}(U_i)$ находим очередное значение случайной величины X .

Если обратную функцию для $F(x)$ в явном виде найти не удастся, то можно использовать ее аппроксимацию многочленами, рациональными дробями или цепными дробями [2].

Примеры распределений

- *Распределение Релея*

Плотность распределения

$$f(y) = \frac{y}{\sigma^2} \cdot \exp\left(-\frac{y^2}{2 \cdot \sigma^2}\right), \quad y \geq 0. \quad (2.2.)$$

Функция распределения

$$F(y) = 1 - e^{-\frac{y^2}{2\sigma^2}}$$

Математическое ожидание

$$m_y = \sqrt{\frac{\pi}{2}} \cdot \sigma$$

Дисперсия

$$\sigma_y^2 = \left(2 - \frac{\pi}{2}\right) \cdot \sigma^2$$

Обратная функция

$$y = \sigma \cdot \sqrt{-2 \ln(1 - u)}$$

- *Показательное распределение*

Плотность распределения

$$f(y) = \lambda \cdot \exp(-\lambda \cdot y), \quad y \geq 0.$$

Функция распределения

$$F(y) = 1 - e^{-\lambda y}$$

Математическое ожидание и дисперсия

$$m_y = \sigma_y = \frac{1}{\lambda}$$

Обратная функция

$$y = -\frac{1}{\lambda} \cdot \ln(1 - u)$$

2.1. Применение обратной функции

- *Распределение Коши*

Плотность распределения

$$f(y) = \frac{1}{\pi \cdot b \cdot \left[1 + \frac{(y-a)^2}{b^2}\right]}$$

Функция распределения

$$F(y) = \frac{1}{\pi} \cdot \operatorname{arctg}\left(\frac{y-a}{b}\right) + \frac{1}{2}$$

Математическое ожидание и дисперсия не определены. Определены мода и медиана:

$$\operatorname{mod}_y = \operatorname{med}_y = a$$

Обратная функция

$$y = b \cdot \operatorname{tg}\left[\pi \cdot \left(u - \frac{1}{2}\right)\right] + a$$

- *Закон арксинуса*

Плотность распределения

$$f(y) = \frac{1}{\pi \cdot b \cdot \sqrt{1 - \frac{(y-a)^2}{b^2}}} \quad (2.3.)$$

Функция распределения

$$F(y) = \frac{1}{\pi} \cdot \arcsin\left(\frac{y-a}{b}\right) + \frac{1}{2}$$

Математическое ожидание и дисперсия

$$m_y = a, \quad \sigma_y^2 = \frac{b^2}{2}$$

Обратная функция

$$y = b \cdot \sin\left[\pi \cdot \left(u - \frac{1}{2}\right)\right] + a$$

Пример: моделирование выборки из распределения арксинуса

Задаем параметры распределения и объем выборки:

$$a := 1 \quad b := 2 \quad N := 5000 \quad i := 0.. N - 1$$

Создаем выборку и вариационный ряд:

$$Y_i := a + b \cdot \text{in}(\pi \cdot (\text{rnd}(1) - 0.5)) \quad Z := \text{sort}(Y)$$

Поместим на один рисунок графики теоретической и экспериментальной плотности распределения, построенной по данной выборке:

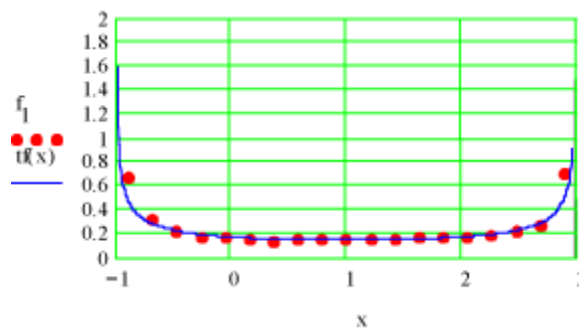


Рис. 2.1. Экспериментальная (точки) и теоретическая (сплошная линия) плотности распределения закона арксинуса

2.2. Нормальное распределение

Из многих способов моделирования нормально распределенных случайных величин мы рассмотрим только те, которые проще всего реализовать в среде MathCAD.

Заметим, что в версии MathCAD 6.0 и выше для наиболее употребляемых распределений созданы функции, генерирующие значения соответствующих случайных величин.

Обратную функцию для нормального распределения можно получить в виде какой-либо аппроксимации [2]. После чего применить

2.2. Нормальное распределение

описанный выше алгоритм. Однако проще всего воспользоваться некоторыми специальными свойствами нормального распределения.

2.2.1. Аналитический способ

Известно, что распределение произведения двух независимых случайных величин, одна из которых имеет релеевское распределение (2.2.), а другая распределена по закону арксинуса (2.3.) с параметрами $(0, 0.5)$, является нормальным [3].

Это позволяет формировать нормальную случайную величину с помощью следующих преобразований:

$$y = \sin(2 \cdot \pi \cdot U_1) \cdot \sqrt{-2 \cdot \ln(U_2)} \quad (2.4.)$$

или

$$y = \cos(2 \cdot \pi \cdot U_1) \cdot \sqrt{-2 \cdot \ln(U_2)} \quad (2.5.)$$

где U_1 и U_2 — независимые реализации случайных чисел. Параметры получаемой нормальной случайной величины будут $(0, 1)$.

Программа моделирования нормального распределения

$$n := 100 \quad k := 0.. N - 1 \quad s := 4.5 \quad m := 10$$

$$B_k := m + s \cdot \sin(2 \cdot \pi \cdot \text{rnd}(1)) \cdot \sqrt{-2 \cdot \ln(\text{rnd}(1))}$$

2.2.2. Способ, основанный на центральной предельной теореме

Согласно центральной предельной теореме, распределение суммы независимых одинаково распределенных случайных величин при неограниченном увеличении числа слагаемых стремится к нормальному распределению.

Пусть U_1, U_2, \dots, U_n — независимые случайные величины, равномерно распределенные на $[0, 1]$. Числовые характеристики U_i равны:

$$M[U_i] = 0.5; \quad D[U_i] = 1/12. \quad (2.6.)$$

Из данных случайных величин легко сформировать случайную величину, имеющую распределение, близкое к стандартному:

$$y_n = \frac{\sum_i U_i - n/2}{\sqrt{n/12}} \quad (2.7.)$$

Уже при $n = 12$ формула (2.7.) дает случайную величину с распределением, близким к стандартному. В этом случае (2.7.) имеет особенно простой вид:

$$y_{12} = \sum_{i=1}^{12} U_i - 6 \quad (2.8.)$$

Распределение с произвольными параметрами (m, s^2) легко получить из стандартного с помощью преобразования:

$$y = m + s \cdot y_{12} \quad (2.9.)$$

Программа моделирования стандартного распределения

$$n := 10 \quad i := 0.. n - 1 \quad N(m, s) := m + s \cdot \left(\sum_i (\text{rnd}(1) - 6) \right)$$

$$K := 500 \quad k := 1.. K \quad B_k := N(0, 1)$$

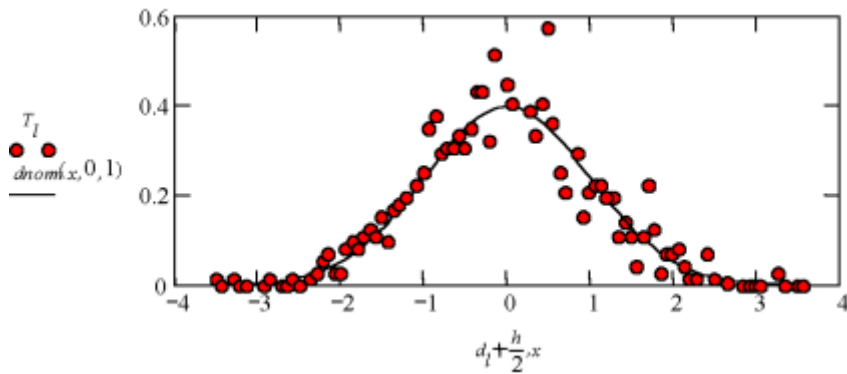


Рис. 2.2. Оценка плотности по выборке из нормального распределения

2.3. Распределения, связанные с нормальным

Из независимых выборок из стандартного распределения, легко получить выборки из хи-квадрат распределения, распределения Стьюдента и Фишера-Снедекора.

хи-квадрат распределение

Распределение хи-квадрат имеет случайная величина

$$\chi_n^2 = \sum_{i=1}^n \xi_i^2, \quad (2.10.)$$

где ξ_i — независимые случайные величины, имеющие стандартное распределение.

Программа моделирования

Выборку объема K можно получить, суммируя независимые стандартно распределенные числа, получаемые по описанным выше алгоритмам

$$K := 1000 \quad j := 0.. K - 1 \quad n := 10 \quad i := 0.. n - 1$$

$$t_i := \sum_j \left[[\sin(2 \cdot \pi \cdot rnd(1))]^2 \cdot \ln(-2 \cdot \ln(rnd(1))) \right]$$

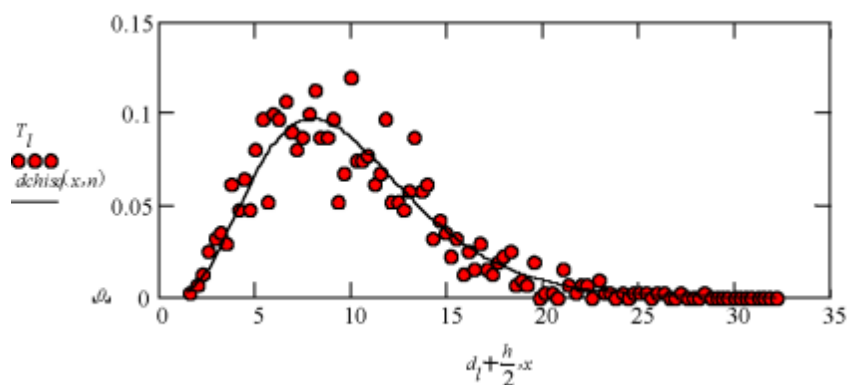


Рис. 2.3. Моделирование распределения хи-квадрат с 10-ю степенями свободы

Распределение Стьюдента

Распределение Стьюдента имеет случайная величина, равная отношению независимых стандартно распределенной величины ξ и корня квадратного из случайной величины χ^2 , имеющей хи-квадрат распределение, деленной на число ее степеней свободы n :

$$T_n = \frac{\xi}{\sqrt{\frac{\chi^2}{n}}} \quad (2.11.)$$

Программа моделирования

$$K := 100 \quad j := 0..K-1 \quad n := 10 \quad j := 0..n-1$$

$$t_i := \frac{\sin(2 \cdot \pi \cdot rnd(1)) \cdot \sqrt{\ln(-2 \cdot \ln(rnd(1)))}}{\sqrt{\frac{1}{n} \cdot \sum_j [[\sin(2 \cdot \pi \cdot rnd(1))]^2 \cdot \ln(-2 \cdot \ln(rnd(1)))]}}$$

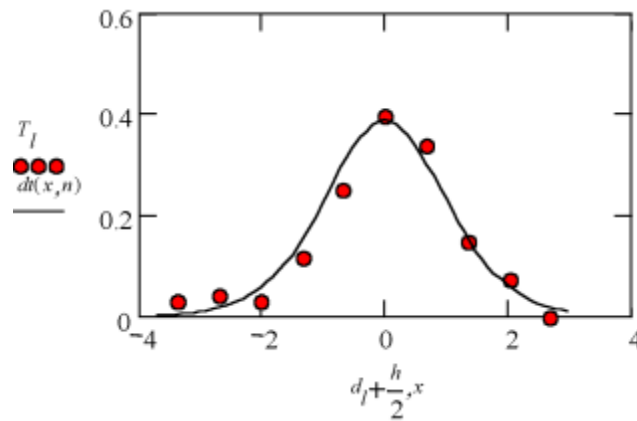


Рис. 2.4. Моделирование распределения Стьюдента с n степенями свободы

Распределение Фишера-Снедекора

Данное распределение описывает отношение двух независимых случайных величин, имеющих хи-квадрат распределение:

$$F = \frac{\chi_1}{\chi_2} : \frac{N_1}{N_2} \quad , \quad (2.12.)$$

2.4. Изотропный вектор

где N_1, N_2 — степени свободы распределений.

Программа моделирования

$$K := 1000 \quad i := 0.. K - 1$$

$$N1 := 10 \quad u := 0.. N1 - 1 \quad N2 := 20 \quad \nu := 0.. N2 - 1$$

$$F_i := \frac{\sum_u \left[[\sin(2 \cdot \pi \cdot rnd(1))]^2 \cdot 2 \cdot \ln(rnd(1)) \right]}{\sum_\nu \left[[\sin(2 \cdot \pi \cdot rnd(1))]^2 \cdot 2 \cdot \ln(rnd(1)) \right]} \cdot \frac{N2}{N1}$$

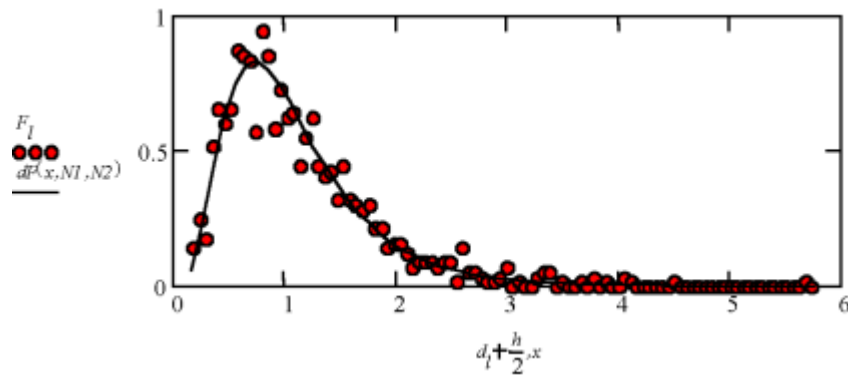


Рис. 2.5. Моделирование распределения Фишера-Снедекора с N_1 и N_2 степенями свободы

2.4. Изотропный вектор

Для практических целей важным является моделирование изотропного направления в трехмерном пространстве. Рассмотрим радиус-вектор, свободный конец которого может находиться в произвольной точке на сфере единичного радиуса.

Изотропность вектора предполагает, что вероятность попадания свободного конца вектора в произвольную область на сфере пропорциональна площади (мере) этой области и не зависит от ее формы.

В сферической системе координат ориентация вектора задается двумя углами: θ — углом между вектором и осью OZ и φ — углом между проекцией вектора на плоскость XOY и осью OX (рис.2.6).

Выберем область D на сфере в виде сферического пояса, параллельного плоскости XOY . Такая область обладает замечательным свойством: ее площадь не зависит от положения на сфере (на экваторе или на полюсе), а определяется толщиной слоя dZ :

$$S_{\omega} = 2 \cdot \pi \cdot R \cdot dZ,$$

где R — радиус сферы. Два события: {Конец вектора попал в область D } и {Проекция R_z приняло данное значение} — эквивалентны. Вероятность этих событий постоянна при изменении проекции на отрезке $[-R, R]$. Следовательно, можно утверждать, что проекция R_z единичного изотропного вектора равномерно распределена на отрезке $[-1, 1]$.

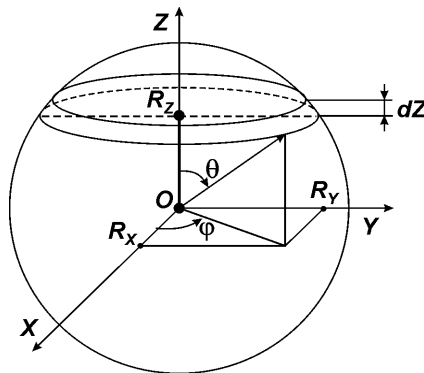


Рис. 2.6. Задание ориентации радиус-вектора в сферических координатах

В силу симметрии области D угол φ будет равномерно распределен на отрезке $[0, \pi]$.

Связь между декартовыми и сферическими координатами дается следующими соотношениями:

$$\begin{cases} R_x = |R| \cdot \sin(\theta) \cdot \cos(\varphi) \\ R_y = |R| \cdot \sin(\theta) \cdot \sin(\varphi) \\ R_z = |R| \cdot \cos(\theta) \end{cases} \quad (2.13.)$$

Следовательно, для моделирования изотропного вектора необходимо вначале задать две равномерно распределенные случайные величины: R_z и φ , а затем по формулам (2.13.) рассчитать остальные проекции вектора. При этом надо учесть, что $\sin(\theta) = \sqrt{1 - R_z^2}$.

2.5. Методы Монте-Карло

Методами Монте-Карло называют численные методы расчета, использующие статистическое моделирование распределений случайных величин.

Суть методов Монте-Карло состоит в том, что величина, подлежащая вычислению, представляется как математическое ожидание некоторой функции от случайной величины с известным распределением.

Таким образом, расчетная задача сводится к оценке математического ожидания случайной величины с известным распределением, которая легко реализуется рассмотренными выше методами статистического моделирования.

Рассмотрим алгоритм вычисления однократного определенного интеграла

$$J = \int_a^b h(x) dx.$$

Пусть $f_\xi(x)$ — плотность распределения некоторой случайной величины, заданной на $[a, b]$. Проведем тождественное преобразование в подынтегральном выражении.

$$J = \int_a^b h(x) dx = \int_a^b \frac{h(x)}{f_\xi(x)} \cdot f_\xi(x) dx = \int_a^b \psi(x) \cdot f_\xi(x) dx = M[\psi(x)] \quad (2.14.)$$

Таким образом, мы представили интеграл как математическое

ожидание функции

$$\psi(x) = \frac{h(x)}{f_{\xi}(x)} \quad (2.15.)$$

от случайной величины с плотностью распределения $f_{\xi}(x)$, при этом

$$f(x) \neq 0, \quad \forall x \in [a, b].$$

Порядок оценки математического ожидания функции $\psi(x)$ (2.15.):

1. Создаем выборку с заданным распределением $\{x_i\}$, $i = 1 \dots n$.
2. Создаем выборку $\{\psi(x_i)\}$.
3. Получаем точечную оценку интеграла:

$$J \approx J_n = \frac{1}{n} \cdot \sum_{i=1}^n \psi(x_i).$$

4. Оцениваем среднеквадратичную погрешность и строим доверительный интервал.

Плотность $f(x)$ выбирают такой, чтобы погрешность оценки была минимальной. В данном случае удобно использовать случайную величину, равномерно распределенную на $[a, b]$. Тогда имеем

$$f(x) = \frac{1}{b-a}; \quad \psi(x) = (b-a) \cdot h(x); \quad J_n = \frac{b-a}{n} \cdot \sum_{i=1}^n h(x_i)$$

Пример. Вычисление пятикратного интеграла методом Монте-Карло

Задание. Методом Монте-Карло вычислить интеграл

$$\int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 \sin(x^2 + y^2 + z^2 + v^2 + t^2) dx dy dz dv dt$$

2.5. Методы Монте-Карло

по единичному пятимерному кубу.

Программа вычислений:

$$f(x, y, z, v, t) := \sin(x^2 + y^2 + z^2 + v^2 + t^2); \quad m := 0 \dots 5000$$

$$I_m := f(\text{rnd}(1), \text{rnd}(1), \text{rnd}(1), \text{rnd}(1), \text{rnd}(1))$$

$$MF := \text{mean}(I) \quad MF = 0.79501$$

$$DF := \text{var}(I) \quad DF = 0.06556$$

$$DI := \begin{bmatrix} MF - \sqrt{\frac{DF}{N}} \cdot 1.96 \\ MF + \sqrt{\frac{DF}{N}} \cdot 1.96 \end{bmatrix} \quad DI = \begin{bmatrix} 0.77914 \\ 0.81088 \end{bmatrix}$$

Вывод. Приближенное значение интеграла, равно 0.80. С доверительной вероятностью 0.95 точное значение лежит в интервале (0.77 ; 0.81).

Глава 3.

Индивидуальное задание и образцы лабораторных работ

3.1. Индивидуальное задание к теме «Моделирование случайных величин»

Исходные данные. Даны случайные величины с известными распределениями:

Случайная величина	Закон распределения
ξ	Нормальное распределение с параметрами a, σ
η	Показательное распределение с параметром λ
ζ	Биномиальное распределение с параметрами n, p
ψ	Распределение Пуассона с параметром λ
ρ	Распределение Релея с параметром σ
κ	Распределение Коши с параметрами a, b
ϕ	Распределение арксинуса с параметрами a, b
ω	Изотропный вектор с координатами $(\omega_1, \omega_2, \omega_3)$

Задание.

1. Смоделировать необходимые выборки согласно индивидуальным заданиям.
2. Оценить среднее значение функции и сравнить его с функцией от средних значений.
3. Найти 95-ти процентный доверительный интервал для функции $F(\xi, \eta, \zeta, \psi, \omega)$.
4. Построить 95-ти процентный доверительный интервал для среднего значения функции. Подогнать его так, чтобы он укладывался в 5-ти процентную относительную ошибку.

5. Сопоставить выборочные законы распределения случайных величин $\xi, \eta, \zeta, \psi, \omega$ и функции $F(\xi, \eta, \zeta, \psi, \omega)$.

Пример выполнения индивидуального задания

Дано:

случайная величина V — распределена нормально с параметрами $N(2, 0.8)$;

случайная величина B имеет биномиальное распределение с параметрами $n = 5, p = 0.35$;

X, Y, Z — координаты изотропного вектора.

Принять относительную погрешность вычисления значения функции равной 2.5%.

Создаем выборку из нормального распределения:

$$k := 0 \dots N - 1$$

$$V_k := 2 + 0.8 \cdot \left(\sin(2 \cdot \pi \cdot rnd(1)) \cdot \sqrt{-2 \cdot \ln(rnd(1))} \right)$$

Создаем выборку из биномиального распределения

$$n := 5 \quad i := 0 \quad p := 0.35 \quad B_k := \sum_i \Phi(p - rnd(1))$$

Моделируем изотропный вектор

$$Z_k := 2 \cdot rnd(1) - 1 \quad Fi_k := rnd(2 \cdot \pi)$$

$$X_k := \sqrt{1 - Z_k^2} \cdot \cos(Fi_k) \quad Y_k := \sqrt{1 - Z_k^2} \cdot \sin(Fi_k)$$

Создаем массив значений функции:

$$Fun := \overline{F(V, B, X, Y, Z)}$$

Оцениваем значения функции и ее дисперсии:

$$MF := mean(Fun) \quad MF := 3.3757$$

$$DF := var(Fun) \quad DF := 7.8320$$

Значение функции от средних:

$$F(\text{mean}(V), \text{mean}(B), \text{mean}(X), \text{mean}(Y), \text{mean}(Z)) := 121.4548$$

При построении 95-ти процентного доверительного интервала используем статистику $STAT$, полагая при этом, что ее распределение близко к стандартному :

$$STAT := \frac{Fun - MF}{\sqrt{\frac{DF}{N}}} \quad |STAT| < 1.96$$

Рассчитаем границы доверительного интервала и относительную погрешность значения функции:

$$DI := \begin{bmatrix} MF - \sqrt{\frac{DF}{N}} \cdot 1.96 \\ MF + \sqrt{\frac{DF}{N}} \cdot 1.96 \end{bmatrix} \quad DI := \begin{bmatrix} 3.20228 \\ 3.54919 \end{bmatrix}$$

$$N \equiv 1000 \quad \delta := \frac{(DI_1 - DI_0)}{2 \cdot MF} \quad \delta = 0.04921$$

Пока данная величина не станет меньше, чем заданная погрешность, будем увеличивать число элементов выборки N .

N	50	100	200	250	500	2000	4000
δ	0.218	0.171	0.102	0.099	0.068	0.032	0.024

Получены результаты, из которых видно, что требуемая точность вычислений достигается при $N = 4000$.

Построение гистограммы функции:

$$SF := \text{sort}(Fun) \quad R := \frac{N}{20} \quad r := 0.. R - 1 \quad h := \frac{SF_{N-1} - SF_0}{R}$$

$$d_r := SF_0 + r \cdot h \quad H := \frac{\text{hist}(d, SF)}{N \cdot h} \quad H_{\text{last}(H)+1} := 0$$

Выведем результаты на график:

3.1. ИДЗ «Моделирование случайных величин»

$$x := -1, -0.8.. 10 \quad m := 0.. 5$$

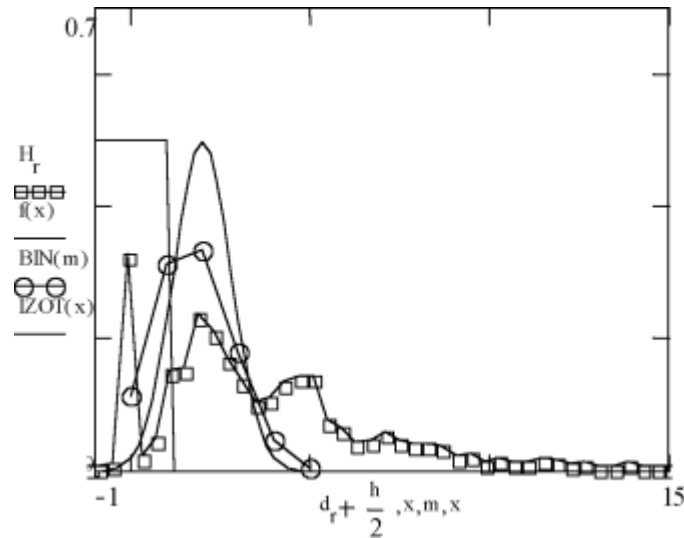


Рис. 3.1. Распределения исходных случайных величин и функции от них

Модельные распределения задаются следующими выражениями:

$$BIN := \frac{5!}{m! \cdot (5 - m)!} \cdot 0.35^m \cdot 0.65^{5-m},$$

$$f(x) \equiv \frac{1}{\sqrt{2 \cdot \pi \cdot 0.8}} \exp\left(-\frac{(x - 2)^2}{2 \cdot 0.64}\right),$$

$$IZOT(x) \equiv if(x < -1, 0, if(x < 1, 0.5, 0)).$$

Как видно из рисунка, плотность распределения функции имеет несколько максимумов в отличие от распределений аргументов. Это приводит к тому, что значения функции от *средних значений аргументов* значительно отличаются от оценки *среднего значения функции*.

3.2. Лабораторная работа по теме «Моделирование биномиального распределения»

Задание.

1. Смоделировать выборку из биномиального распределения с заданными параметрами.
2. Построить вариационный ряд и выборочную функцию распределения.
3. Построить гистограмму и полигон частот.
4. Построить и сравнить выборочные частоты с теоретическими.
5. Сравнить выборочные числовые характеристики с модельными.

Схема Бернулли

Событие A зададим вероятностью p , с которой оно может произойти в единичном эксперименте (или не произойти с вероятностью $(1-p)$). Для генерирования случайных чисел равномерно распределенных на отрезке $[0, 1]$ используем функцию, генерирующую случайные числа — $rnd(1)$. Зададим число p , принадлежащее отрезку $[0,1]$. Будем считать, что появление случайного числа $rnd(1) \leq p$ соответствует появлению события A . Введем индикатор события $J(p$ с помощью функции Хевисайда:

$$J(p) := \Phi(p - rnd(1))$$

Эта функция равна единице, если событие A произошло. И равна нулю в противном случае.

Моделирование выборки из биномиального распределения $Bi(n, p)$.

Смоделируем опыт, состоящий из n независимых испытаний. Зададим n и подсчитаем число испытаний $Bi(n, p)$, в которых произошло событие в данной серии испытаний:

3.2. ЛБ «Моделирование биномиального распределения»

$$i := 0.. n - 1 \quad Bi(n, p) := \sum_i I(p)$$

Параметр n задается ниже в удобном для расчетов месте с помощью оператора глобального определения.

Создадим выборку B из N элементов, подчиняющихся биномиальному распределению:

$$N := 201 \quad k := 0.. N - 1 \quad B_k := Bi(n, p)$$

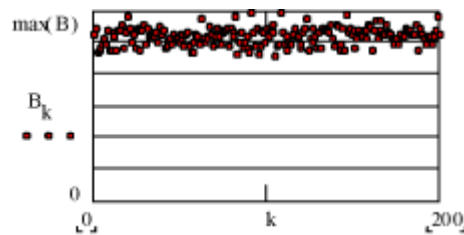


Рис. 3.2. Представление первичного экспериментального материала в виде *статистического ряда*.

Упорядочим выборку:

$$C := \text{sort}(B)$$

Выборочная функция распределения. Выборочную функцию распределения можно задать двумя способами: задав ее в явном виде

$$F_N(x) := \frac{1}{N} \cdot \sum_i \text{if}(x < B_i, 1, 0)$$

или используя вариационный ряд $F_N(k) := \frac{k}{N}$, где k — номер члена вариационного ряда.

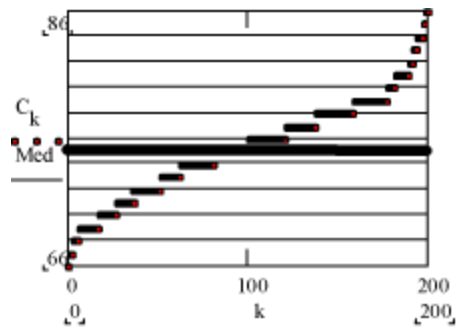


Рис. 3.3. Представления экспериментального материала в виде *вариационного ряда*. Сплошной линией показана медиана

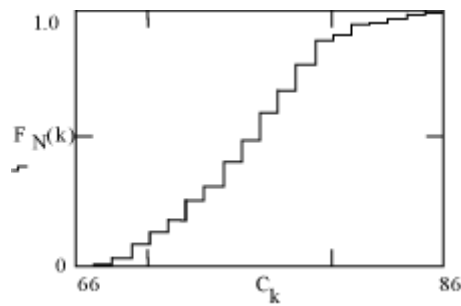


Рис. 3.4. Эмпирическая функция распределения

Гистограмма и полигон частот

Зададим полуинтервал D , полностью покрывающий $[a, b]$, состоящий из M одинаковых непересекающихся полуинтервалов $[x_l, x_{l+1})$.

Зададим шаг дискретизации

$$l := 0.. M - 1 \quad h := \frac{\max(B) - \min(B)}{M - 1} \cdot 1.001$$

Создадим массив \mathbf{x} , содержащий координаты точек разбиения области D :

$$x_l := \min(B) + h \cdot l \quad \min(B) = 66 \quad \max(B) = 86$$

При построении гистограммы используем системную функцию $hist(vx, vy)$:

3.2. ЛБ «Моделирование биномиального распределения»

$$H := \frac{hist(x, B)}{N \cdot h}.$$

Поскольку число элементов у массива $hist(vx, vy)$ на единицу меньше, чем у массива x , то добавим ему еще один нулевой элемент:

$$H_{last(H)+1} := 0$$

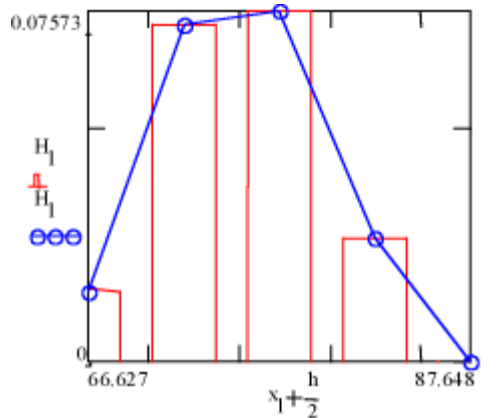


Рис. 3.5. Гистограмма и полигон частот

Сравним выборочную и теоретическую плотности распределения.

$$P(x, p, n) := \frac{n! \cdot p^x \cdot (1 - p)^{n-x}}{x! \cdot (n - x)!}$$

$$M \equiv 5 \quad t := 0.. \max(B) \quad n \equiv 100 \quad p \equiv 0.75$$

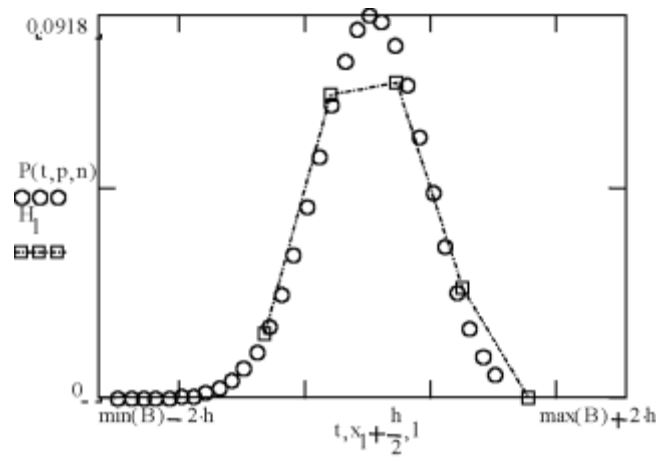


Рис. 3.6. Закон биномиального распределения и его экспериментальная оценка

Числовые характеристики			
Теоретические		Выборочные	
Мат. ожидание	$m = 75$	Выборочное среднее	$\text{mean}(B) = 74.79$
Дисперсия	$n \cdot p \cdot (1 - p) = 18.75$	Выборочная дисперсия	$\text{var}(B) = 17.173$ $\frac{n}{n-1} \cdot \text{var}(B) = 17.347$

Рис. 3.7. Сравнение теоретических и выборочных числовых характеристик

Для дискретных распределений вернее будет рассчитывать не гистограмму, а относительные частоты появления того или иного значения случайной величины. Для этого сначала определим область на-

3.3. Моделирование подбрасывания монеты

блюдаемых значений случайной величины:

$$m := C_0.. C_{last(C)}.$$

Введем индикатор события {СЛУЧАЙНАЯ ВЕЛИЧИНЫ ПРИНЯЛА ЗАДАННОЕ ЗНАЧЕНИЕ}:

$$Ident(x, y) := if(x \neq y, 0, 1).]$$

С помощью данного индикатора подсчитаем частоты событий:

$$\eta_m := \sum_k \frac{Ident(m, C_k)}{N - 1}$$

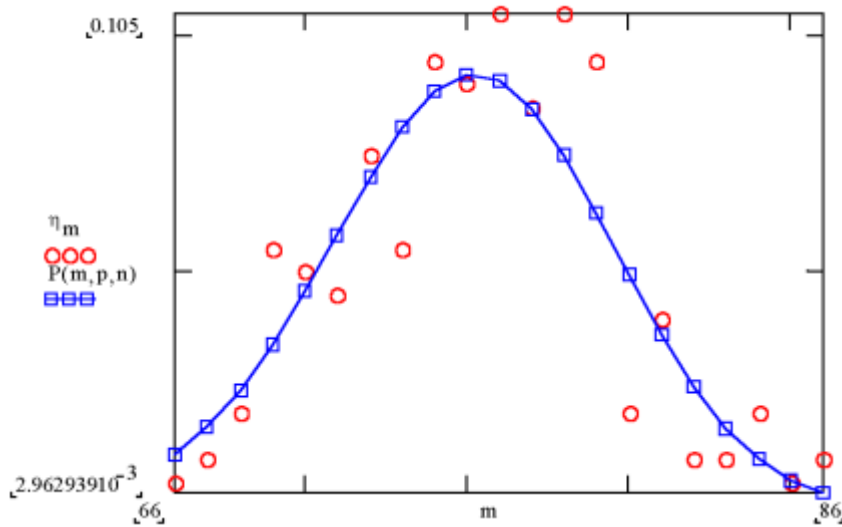


Рис. 3.8. Частотное представление результатов моделирования

3.3. Моделирование подбрасывания монеты

Задание. Смоделировать опыт подбрасывания монеты. Проследить поведение частоты появления «орла» с увеличением числа опытов.

Задаем индикатор события и для подсчета числа событий в серии используем рекуррентную формулу:

$$ORIGTN := 1 \quad N \equiv 5000 \quad i := 1.. N$$

$$S_1 := if(rnd(1) > 0.5, 0, 1) \quad S_{i+1} := S_i + if(rnd(1) > 0.5, 0, 1)$$

Выведем результаты на график

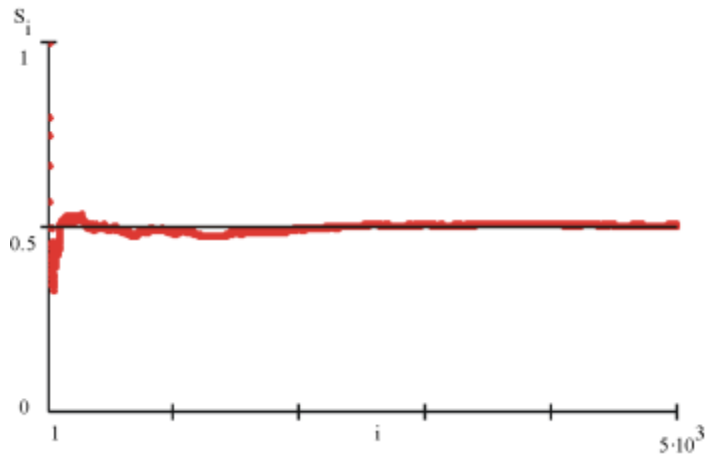


Рис. 3.9. Поведение частоты появления «орла» при подбрасывании монеты в зависимости от числа испытаний

Данный опыт иллюстрирует теорему Бернулли: частота появления события в серии испытаний сходится по вероятности к вероятности события в отдельном испытании.

Варианты заданий

№	$a; b$	σ	λ	n	p	$F(\xi, \eta, \zeta, \psi, \omega)$
1	1	0.5	2	—	—	$(\xi^2 + \eta^2)/(1 + \eta)$
2	0.1	1.1	—	—	—	$\sin(2 \cdot \omega_1) \cdot \exp(\omega_2)$
3	$\pi/4$	$\pi/16$	2	—	—	$\arctan(\xi) \cdot (1 + \psi^2)^{-1}$
4	0.2	0.1	—	—	—	$\ln(1 + \xi) \cdot [\omega_1^2 + \omega_2^2]^{-1}$
5	$\pi/2$	0.5	1	—	—	$(\cos(\xi) + \eta) \cdot [\omega_1^2 + \omega_2^2]^{-1}$
6	1	0.1	—	10	0.2	$\xi^2 \cdot \exp(\zeta^2)$
7	2	0.5	—	—	—	$(\xi + \exp(\omega_1)) \cdot (1 + \exp(2 \cdot \omega_2^2))^{-2}$
8	0	1	1.2	10	0.1	$\xi^2 + \zeta^2 + \eta^2$
9	0.5	1	0.8	10	0.2	$(\xi^2 + \zeta^2 + \eta^2) \cdot (2 + \tan(\zeta^2))^2$
10	0.4	0.2	3	10	0.5	$(\cos(\xi) + \eta) \cdot (2 + \tan(\zeta^2))^2$
11	0.1	1	3	15	0.45	$\exp[(\cos(\xi) + \eta) \cdot (1 + h)^{-1}]$
12	—	3	4	—	—	$\cos(\psi) \cdot \exp(-\omega_1^2)$
13	0.1	2	—	20	0.3	$(\omega_1 + \xi) \cdot \arctan(1 + h)^{-1}$
14	1; 2	3	0.8	—	—	$\sin(h^3) \cdot (\omega_1 + \xi) \cdot \arctan(1 + h)^{-1}$
15	2; 3	—	1	—	—	$\sin(\xi + \cos(\sqrt{1 + \eta}))^3$
16	1	0.8	2.5	—	—	$\sqrt{\xi^2 + \eta^2 + \omega_1^2}$
17	0.5	1.2	0.7	9	0.4	$\sqrt{\xi^2 + \sqrt{\eta^2 + \zeta^2}}$
18	0.1	0.7	3.5	—	—	$\sin(\xi + \sqrt{\eta^2 + \omega_1^2})^3$
19	-1	1	-0.5	—	—	$\sin((\cos(\xi^2) + \eta^2))$
20	$\pi/4$	$\pi/32$	—	—	—	$\arctan(\xi^2) \cdot (1 + \arctan(\omega_1^2))^{-1}$

Таблица 3.1. Варианты заданий

Список литературы

1. Ермаков С.М., Михайлов Г.А. Статистическое моделирование.— М.: Наука, 1982, 296 с.
2. Шторм П., Нойман П. Таблицы по математической статистике. — М.: Финансы и статистика, 1982.
3. Ивченко Г.И., Медведев Ю.И. Математическая статистика. — М.: Высшая школа, 1992. 304 с.
4. Гмурман В.Е. Теория вероятностей и математическая статистика. М. — 1998.
5. Плис А.И., Сливина Н.А. Лабораторный практикум по высшей математике. — М.: Высш. шк., 1994.—416 с.
6. MATHCAD 6.0 PLUS. Финансовые инженерные и научные расчеты в среде WINDOWS 95. /Перевод с англ.— М.: Информ.-издат.дом “ФилинЪ”. 1996.—712с.

Содержание

Введение	3
1. Моделирование дискретных случайных величин	5
1.1. Случайные испытания с двумя исходами	5
1.2. Дискретная СВ с заданным законом распределения . .	6
1.3. Опыт с большим числом исходов	10
1.4. Биномиальное распределение	11
1.5. Распределение Пуассона	12
2. Моделирование непрерывных распределений	15
2.1. Применение обратной функции	15
2.2. Нормальное распределение	18
2.2.1. Аналитический способ	19
2.2.2. Следствие центральной предельной теоремы . .	19
2.3. Распределения, связанные с нормальным	21
2.4. Изотропный вектор	23
2.5. Методы Монте-Карло	25
3. Идз и образцы лабораторных работ	28
3.1. ИДЗ «Моделирование случайных величин»	28
3.2. ЛБ «Моделирование биномиального распределения» .	32
3.3. Моделирование подбрасывания монеты	37
Варианты заданий	39

СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Метод. указания по выполнению лабораторных работ и индивидуального задания по математической статистике

Юрий Иванович Галанов