Министерство образования РФ

ФГБОУ ВПО Владимирский государственный университет

имени Александра Григорьевича и Николая Григорьевича

Столетовых

Кафедра РТиРС

О.Р.Никитин

Специализация по теме диссертации

Методические указания к лабораторным работам

ВЛАДИМИР 2012

СОДЕРЖАНИЕ

ЛАБОРАТОРНАЯ РАБОТА №1

ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН **4**

ЛАБОРАТОРНАЯ РАБОТА №2

МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ЧИСЕЛ С ЗАДАННЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ **12**

ЛАБОРАТОРНАЯ РАБОТА №3

ЭМПИРИЧЕСКИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ **16**

ЛАБОРАТОРНАЯ РАБОТА №4

ОЦЕНИВАНИЕ КООРДИНАТЫ ЦЕНТРА РАСПРЕДЕЛЕНИЯ ВЫБОРОЧНОЙ СОВОКУПНОСТИ **22**

ЛАБОРАТОРНАЯ РАБОТА №5

ПРОВЕРКА ПРИНАДЛЕЖНОСТИ ВЫБОРОЧНОЙ СОВОКУПНОСТИ К НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ **25**

ПРИЛОЖЕНИЕ

СПРАВОЧНЫЕ СВЕДЕНИЯ ПО ЯЗЫКУ MATLAB **29**

СПИСОК ЛИТЕРАТУРЫ 39

Аннотация

Методические указания к лабораторным работам предназначены для бакалавров направления 210400 «Радиотехника» и специальности 210600 «Радиотехнические системы и компоненты» по дисциплине «Специализация по теме диссертации».

Лабораторный практикум содержит 5 лабораторных работ, в процессе выполнения которых студенты знакомятся с законами распределения случайных чисел, эмпирическими функциями распределения, моделированием случайных чисел с заданным законом распределения, получают знания по оценкам распределения выборочных совокупностей экспериментальных данных. Задания выполняются на ЭВМ в языке MatLab. Пособие снабжено необходимыми краткими положениями теоретического материала, справочными сведениями по языку MatLab и списком необходимой литературы.

В силу своей универсальности данных данные лабораторных работ методических указаний могут быть использованы для других направлений и специальностей, где происходит обучение обработке экспериментальных данных.

ЛАБОРАТОРНАЯ РАБОТА №1

ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Цели работы

1. Изучение законов распределения случайных величин, наиболее часто применяемых при решении задач обработки экспериментальных данных.
2. Изучение инструментов MATLAB для моделирования функций распределений.
3. Исследование с помощью средств MATLAB зависимости функций распределений от их параметров.

Основные теоретические положения

Закон распределения случайной величины – это соотношение, устанавливающее связь между возможными значениями случайной величины и вероятностями их появления.

Интегральная функция распределения (или просто *функция распределения*) *случайной величины X* – это функция *F(x),* определяющая для каждого значения *x* вероятность того, что случайная величина *X* примет значение меньше *x*

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

Свойства функции распределения

1.	Функция распределения – это неотрицательная функция, заключенная между 0 и 1, т.е. $0 \le F(x) \le 1$

2.	Функция распределения – это неубывающая функция, т.е. $F(x_2) \ge F(x_1)$ при $x_2 > x_1$
3.	Вероятность того, что случайная величина примет значение из интервала $[a,b]$, равна приращению функции распределения на этом интервале $P(a \le x \le b) = F(b) - F(a)$
4.	Вероятность того, что случайная величина примет одно определенное значение равна 0.
5.	Если возможные значения случайной величины принадлежат интервалу $[a,b]$, то $F(x) = \begin{cases} 0 \text{ ï ðè } x \le a, \\ 1 \text{ ï ðè } x \ge b. \end{cases}$. В общем случае $F(-\infty) = 0, \quad F(+\infty) = 1$.

Несмотря на то, что интегральная функция распределения является исчерпывающей характеристикой случайной величины, по ней трудно судить о характере распределения случайной величины в небольшой окрестности какой-либо точки числовой оси. Поэтому, наряду с интегральной рассматривают также дифференциальную функцию распределения случайной величины.

Дифференциальная функция распределения (плотность распределения, плотность вероятности) *f(x)* представляет собой производную от функции *F(x).*

$$f(x) = \frac{dF(x)}{dx}.$$

Свойства плотности вероятности

1.	Плотность вероятности – неотрицательная функция, т.е. $f(x) \ge 0$.
2.	Функция распределения равна интегралу от функции плотности вероятности, т.е. $F(x) = \int_{-\infty}^{x} f(x)dx$
3.	Вероятность попадания случайной величины в интервал $[a,b]$ равна площади под кривой *f(x)*, т.е. $P(a \le x \le b) = \int_a^b f(x)dx$.

4.	Интеграл от функции плотности вероятности в бесконечных пределах равен 1, т.е. $\int_{-\infty}^{+\infty} f(x)dx = 1$

Закон распределения Гаусса (нормальный закон распределения)

Наиболее широко применяемое распределение. Нормальное распределение является краеугольным камнем математической статистики в силу ряда причин:

– схема его возникновения соответствует многим реальным физическим процессам, порождающим результаты обрабатываемых наблюдений;

– при возрастании объема выборки предельное распределение для большинства экспериментальных данных может быть аппроксимировано нормальным законом;

– нормальное распределение обладает рядом благоприятных математико-статистических свойств (легко нормируется и аппроксимируется, обладает свойством аддитивности).

Обозначение	$N(\mu, \sigma)$
Параметры	μ – ì àò åì àò è÷åñêî å î æ èäàí èå, σ – äèñïåðñèÿ
Плотность вероятности	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad (-\infty < x < +\infty)$
Функция распределения	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\}dt, \quad (-\infty < x < +\infty)$

Нормальное распределение с параметрами $\mu = 0$, $\sigma = 1$ (т.е. $N(0,1)$) называется стандартным нормальным распределением.

Закон распределения Пирсона

Если $X_1, X_2, \ldots, X_k$ – независимые случайные величины, имеющие распределение $N(0,1)$, то сумма их квадратов подчиняется распределению χ^2 (хи-квадрат) Пирсона с числом степеней свободы k.

Обозначение	$\chi^2(k)$
Параметр	k – число степеней свободы
Плотность вероятности	$f(x) = \left[\sqrt{2} \cdot \Gamma\left(\frac{k}{2}\right)\right]^{-1} \cdot x^{\frac{k-2}{2}} \cdot \exp\left(-\frac{x}{2}\right), \quad x \ge 0$
Функция распределения	$F(x) = \left[\sqrt{2} \cdot \Gamma\left(\frac{k}{2}\right)\right]^{-1} \int\limits_0^x t^{\frac{k-2}{2}} \exp\left(-\frac{t}{2}\right) dt, \quad x > 0$

где $\Gamma(x)$ – гамма-функция.

Закон распределения Стьюдента

Если Y – случайная величина, распределенная по закону $N(0,1)$, а независимая от нее случайная величина X имеет распределение χ^2 с k степенями свободы, то случайная величина $t = Y \cdot \sqrt{\frac{k}{X}}$ подчиняется распределению Стьюдента с k степенями свободы.

Обозначение	$t(k)$
Параметр	k – число степеней свободы
Плотность вероятности	$f(t) = \Gamma\left(\frac{k+1}{2}\right) \cdot \left[\sqrt{\pi k} \cdot \Gamma\left(\frac{k}{2}\right)\right]^{-1} \cdot \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \quad (-\infty < t < +\infty)$
Функция распределения	$F(t) = \frac{1}{\sqrt{\pi k}} \int\limits_{-\infty}^{t} \left(1 + \frac{y^2}{k}\right)^{-\frac{k+1}{2}} dy, \quad (-\infty < t < +\infty)$

где $\Gamma(x)$ – гамма-функция.

Закон распределения Фишера

Если две независимые случайные величины $X = \chi_1^2$ и $Y = \chi_2^2$ распределены по закону Пирсона со степенями свободы, соответственно, k_1 и k_2, то случайная величина $F = \dfrac{X \cdot k_2}{Y \cdot k_1}$ имеет распределение Фишера со степенями свободы k_1 и k_2.

Обозначение	$F(k_1, k_2)$
Параметры	k_1, k_2
Плотность вероятности	$f(x) = \dfrac{\Gamma\left(\dfrac{k_1+k_2}{2}\right)}{\Gamma\left(\dfrac{k_1}{2}\right)\cdot\Gamma\left(\dfrac{k_2}{2}\right)} \cdot \left(\dfrac{k_1}{k_2}\right)^{\frac{k_1}{2}} \cdot \dfrac{x^{\frac{k_1}{2}-1}}{\left(1+\dfrac{k_1}{k_2}x\right)^{\frac{k_1+k_2}{2}}}, \qquad x \ge 0$
Функция распределения	$F(x) = \dfrac{\Gamma\left(\dfrac{k_1+k_2}{2}\right)}{\Gamma\left(\dfrac{k_1}{2}\right)\cdot\Gamma\left(\dfrac{k_2}{2}\right)} \cdot k_1^{\frac{k_1}{2}} \cdot k_2^{\frac{k_2}{2}} \cdot \int\limits_0^x y^{\frac{k_1}{2}-1} \left(k_1+k_2\right)^{\frac{k_1+k_2}{2}} dy, \qquad x \ge 0$

где $\Gamma(x)$ – гамма-функция.

Экспоненциальный закон распределения

Используется для описания внезапных отказов, когда износом изделия можно пренебречь. Наработка на отказ многих невосстанавливаемых изделий и наработка между соседними отказами у восстанавливаемых изделий в случае простейшего потока отказов подчинены экспоненциальному распределению. Наработка на отказ большой многокомпонентной системы может быть описана экспоненциальным распределением при любом распределении наработки на отказ компонентов системы.

Обозначение	$E(b)$

Параметр	b
Плотность вероятности	$f(x) = \frac{1}{b} \cdot \exp\left(-\frac{x}{b}\right), \quad x \geq 0$
Функция распределения	$F(x) = 1 - \exp\left(-\frac{x}{b}\right), \quad x \geq 0$

Двойное экспоненциальное распределение (в записи функций вместо x используется $|x|$ при $-\infty < x < +\infty$) называется распределением Лапласа.

Равномерный закон распределения

Равномерному распределению подчиняются случайные величины, имеющие одинаковую вероятность появления (например, погрешность измерений с округлением).

Обозначение	$U(a,b)$
Параметры	a, b
Плотность вероятности	$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b; \\ 0, & x < a,\, x > b. \end{cases}$
Функция распределения	$F(x) = \begin{cases} 0, & x \leq a; \\ \frac{x-a}{b-a}, & a < x < b; \\ 1, & x \geq b. \end{cases}$

Функции MATLAB для моделирования законов распределения

`gamma(x)`	Расчет значения гамма-функции в точке x

`normpdf(x,μ,σ)`	Расчет значения плотности вероятности нормального распределения с параметрами μ, σ в точке *x*
`chi2pdf(x,k)`	Расчет значения плотности вероятности распределения Пирсона с *k* степенями свободы в точке *x*
`tpdf(x,k)`	Расчет значения плотности вероятности распределения Стьюдента с *k* степенями свободы в точке *x*
`fpdf(x,k1,k2)`	Расчет значения плотности вероятности распределения Фишера с *k1*, *k2* степенями свободы в точке *x*
`exppdf(x,b)`	Расчет значения плотности вероятности экспоненциального распределения с параметром *b* в точке *x*
`unifpdf(x,a,b)`	Расчет значения плотности вероятности равномерного распределения с параметрами *a, b* в точке *x*
`normcdf(x,μ,σ)`	Расчет значения функции распределения для нормального распределения с параметрами μ, σ в точке *x*
`chi2cdf(x,k)`	Расчет значения функции распределения для распределения Пирсона с *k* степенями свободы в точке *x*
`tcdf(x,k)`	Расчет значения функции распределения для распределения Стьюдента с *k* степенями свободы в точке *x*
`fcdf(x,k1,k2)`	Расчет значения функции распределения для распределения Фишера с *k1*, *k2* степенями свободы в точке *x*
`expcdf(x,b)`	Расчет значения функции распределения для экспоненциального распределения с параметром *b* в точке *x*
`unifcdf(x,a,b)`	Расчет значения функции распределения для равномерного распределения с параметрами *a, b* в точке *x*

Порядок выполнения работы

Для каждого закона распределения

1. Написать m-функцию MATLAB для расчета значения плотности вероятности в зависимости от значения случайной величины и параметров распределения. С помощью полученной m-функции

построить график плотности вероятности (составить соответствующий m-скрипт).

2. Построить график плотности вероятности с помощью соответствующей pdf-функции MATLAB. По сравнению графиков сделать выводы о правильности составленной в п.1. m-функции.
3. Составить m-скрипт для построения графика интегральной функции распределения с помощью соответствующей cdf-функции MATLAB.
4. Исследовать поведение интегральной и дифференциальной функций распределения при различных значениях параметров с помощью составленных в пп.1. и 2. скриптов.

Замечание: Необходимую информацию по языку MATLAB можно найти в Приложении.

ЛАБОРАТОРНАЯ РАБОТА №2

МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ЧИСЕЛ С ЗАДАННЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ

Цели работы

1. Изучение алгоритмов моделирования случайных чисел с заданным законом распределения.
2. Приобретение навыков моделирования одномерных случайных чисел в системе MATLAB

Основные теоретические сведения

Случайные числа с различными законами распределения моделируются с помощью преобразований одного или нескольких независимых значений базовой случайной величины α.

Базовая случайная величина α – это случайная величина с распределением $U(0,1)$ (равномерным распределением в интервале [0,1]). В любой системе программирования имеется стандартная программа моделирования базовой случайной величины.

Моделируемое случайное число обозначим ξ.

Алгоритмы моделирования случайных чисел

Закон распределен	**Плотность вероятности**	**Алгоритм моделирования**

ия		
Гаусса $N(\mu,\sigma)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$, где $(-\infty<x<+\infty)$	$\xi=\mu+\sigma\cos(2\pi\alpha_1)\cdot\sqrt{-2\ln(\alpha_2)}$, где α_i - независимые случайные величины с распределением $U(0,1)$
Пирсона $\chi^2(k)$	$f(x)=\frac{x^{\frac{k-2}{2}}}{\sqrt{2}\cdot\Gamma\left(\frac{k}{2}\right)}\cdot\exp\left(-\frac{x}{2}\right)$, где $x\ge 0$	$\xi=\sum_{i=1}^{k}u_i^2$, где u_i - независимые случайные величины с распределением $N(0,1)$
Стьюдента $t(k)$	$f(x)=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\cdot\Gamma\left(\frac{k}{2}\right)}\cdot\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}}$, где $(-\infty<x<+\infty)$	$\xi=u\sqrt{\frac{k}{v}}$, где $u\in N(0,1)$, $v\in\chi^2(k)$
Фишера $F(k_1,k_2)$	$f(x)=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\cdot\Gamma\left(\frac{k_2}{2}\right)}\cdot\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}\cdot\frac{x^{\frac{k_1}{2}-1}}{\left(1+\frac{k_1}{k_2}x\right)^{\frac{k_1+k_2}{2}}}$, где $x\ge 0$	$\xi=\frac{k_2\cdot v_1}{k_1\cdot v_2}$, где $v_1\in\chi^2(k_1)$, $v_2\in\chi^2(k_2)$
Экспонен-циальный $E(b)$	$f(x)=\frac{1}{b}\cdot\exp\left(-\frac{x}{b}\right),\quad x\ge 0$	$\xi=-b\ln\alpha$
Равномерный $U(a,b)$	$f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b;\\ 0, & x<a,\ x>b.\end{cases}$	$\xi=a+(b-a)\alpha$

Функции Matlab для моделирования случайных чисел

Функция	Назначение

`y=rand`	базовая случайная величина, выбранная из равномерного распределения $U(0,1)$
`y=rand(m,k)`	моделирует $(m \times k)$-матрицу со случайными числами, выбранными из равномерного распределения $U(0,1)$
`y=normrnd(m,s)`	случайное число, выбранное из нормального распределения $N(m,s)$
`y=chi2rnd(k)`	случайное число, выбранное из распределения Пирсона $\chi^2(k)$
`y=trnd(k)`	случайное число, выбранное из распределения Стьюдента $t(k)$
`y=frnd(k1,k2)`	случайное число, выбранное из распределения Фишера $F(k_1,k_2)$
`y=exprnd(b)`	случайное число, выбранное из экспоненциального распределения $E(b)$
`y=unifrnd(a,b)`	случайное число, выбранное из равномерного распределения $U(a,b)$

Порядок выполнения работы

Для каждого закона распределения

1. Написать на основе алгоритма моделирования m-функцию, формирующую выборку случайных чисел. Входными параметрами являются параметры распределения и объем выборки.

2. Написать на основе rnd-функции MATLAB аналогичную п.1. m-функцию.
3. Сформировать на основе написанных в п.1. и п.2. функций выборку объемом $n = 1000$.
4. Написать m-скрипт для отображения полученных выборок на действительной числовой оси.
5. Отобразить в одной системе координат график плотности вероятности и выборку на числовой прямой. Для построения графика плотности вероятности использовать программы из лабораторной работы №1.
6. Сделать выводы по результатам сравнения изображения выборки и графика плотности вероятности.

Замечание 1: Изображение выборки на числовой оси представляет собой график функции $y = f(x)$, у которой все координаты $y_i = 0$, а координаты x_i соответствуют элементам выборки. Вектор y можно получить с помощью функции MATLAB

```
y = zeros(1,n)
```

где $n = 1000$ (объем выборки).

Замечание2: Необходимую информацию по языку MATLAB можно найти в Приложении.

ЛАБОРАТОРНАЯ РАБОТА №3

ЭМПИРИЧЕСКИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

Цели работы

1. Изучение методики формирования интервального распределения выборки случайных чисел и построения эмпирических функций распределения.
2. Изучение инструментов MATLAB для формирования интервального распределения выборки случайных чисел и построения эмпирических функций распределения.
3. Исследование зависимости эмпирических функций распределения от объема выборки случайных чисел.

Основные теоретические положения

Формирование интервального распределения выборки

Генеральная совокупность – это набор всех видов значений случайной величины, которые могли бы быть при данном комплексе условий.

На практике при проведении экспериментов из генеральной совокупности (бесконечного множества возможных значений измеряемого параметра) извлекается ограниченное число объектов. Совокупность случайно отобранных объектов из генеральной совокупности называется *выборочной совокупностью (выборкой)*. Объем выборки обозначается n.

Выборка объема n записывается в виде последовательности её значений $x_1, x_2, \ldots, x_n$.

Последовательность элементов выборки, расположенных в порядке возрастания их значений $x_1 \le x_2 \le \ldots \le x_n$, называется *ранжированным рядом*. Номер элемента ранжированного ряда называется *рангом*.

Для представления выборочной совокупности используется *интервальное распределение*. Для этого диапазон значений $[x_{\min}, x_{\max}]$ разбивается на k равных интервалов. Оптимальное количество интервалов определяется по формуле Старджесса

$$k = 1 + 3.31 \cdot \lg n.$$

Полученное значение округляется до целого.

Затем вычисляется ширина интервала h по формуле

$$h = \frac{x_{\max} - x_{\min}}{k}$$

и границы интервалов

$$x_i^{\text{ëåâ}} = x_{\min} + (i-1) \cdot h,$$

$$x_i^{\text{ï ðàâ}} = x_{\min} + i \cdot h,$$

где i – номер интервала.

Затем подсчитывают частоту (количество) m_i попаданий значений элементов выборки в каждый интервал. Считают, что если некоторый элемент выборки равен левой границе, то элемент попал в интервал, а если правой (кроме последнего интервала) – то не попал.

Середины интервалов вычисляются по формуле

$$x_{0i} = \frac{x_i^{\text{ëåâ}} + x_i^{\text{ï ðàâ}}}{2}.$$

Интервальное распределение удобно представлять в виде таблицы

Границы интервалов	**Середины интервалов**	**Частота попадания в интервал**
$x_i^{\text{ëåâ}} \ldots x_i^{\text{ï ðàâ}}$	x_{0i}	m_i

Эмпирические функции распределения

Для анализа выборок используются аналоги интегральной и дифференциальной функций распределения, которые называются *статистическими* (или *эмпирическими*). Можно сказать, что эмпирические функции распределения выборки служат для оценки теоретических (генеральных) функций распределения генеральной совокупности.

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$.

Эмпирическую функцию распределения изображают в виде лестничного графика, длина каждой ступеньки которого равна длине соответствующего интервала, а высота – отношению накопленной частоты для этого интервала к объему выборки, т.е.

$$F^*(x_i) = \frac{m_i^{нак}}{n}$$

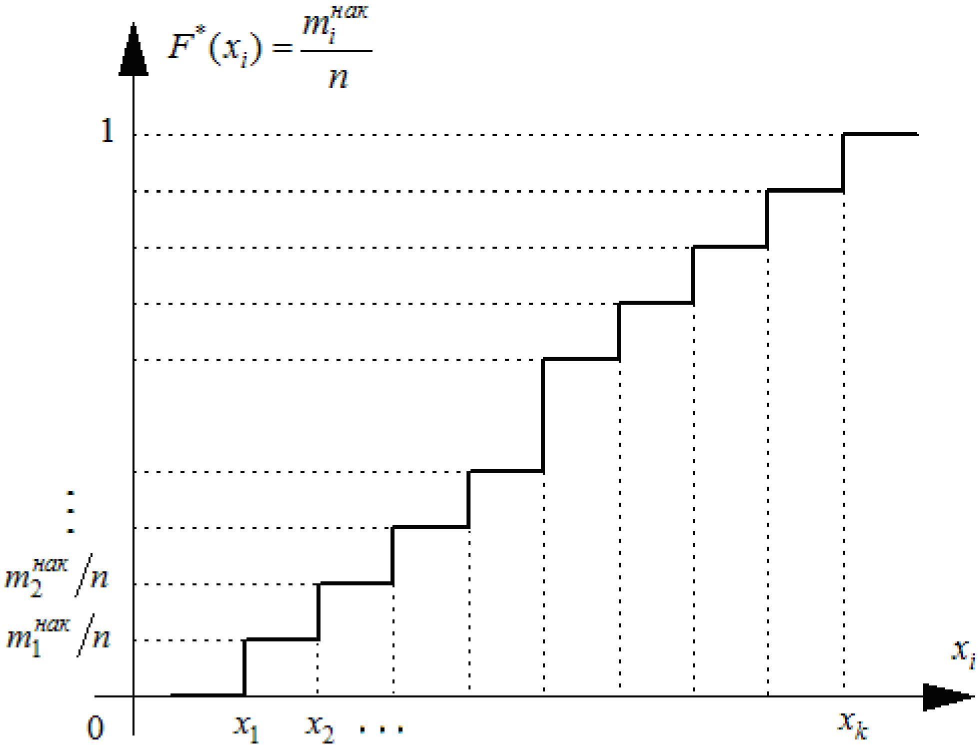
где $m_i^{нак}$ – *накопленная частота,* т.е. число вариант, меньших x_i,

x_i – середина *i*-ого интервала,

n – объем выборки.

Накопленная частота для j-ого интервала определяется последовательным суммированием частот

$$m_j^{нак} = \sum_{i=1}^{j} m_i .$$



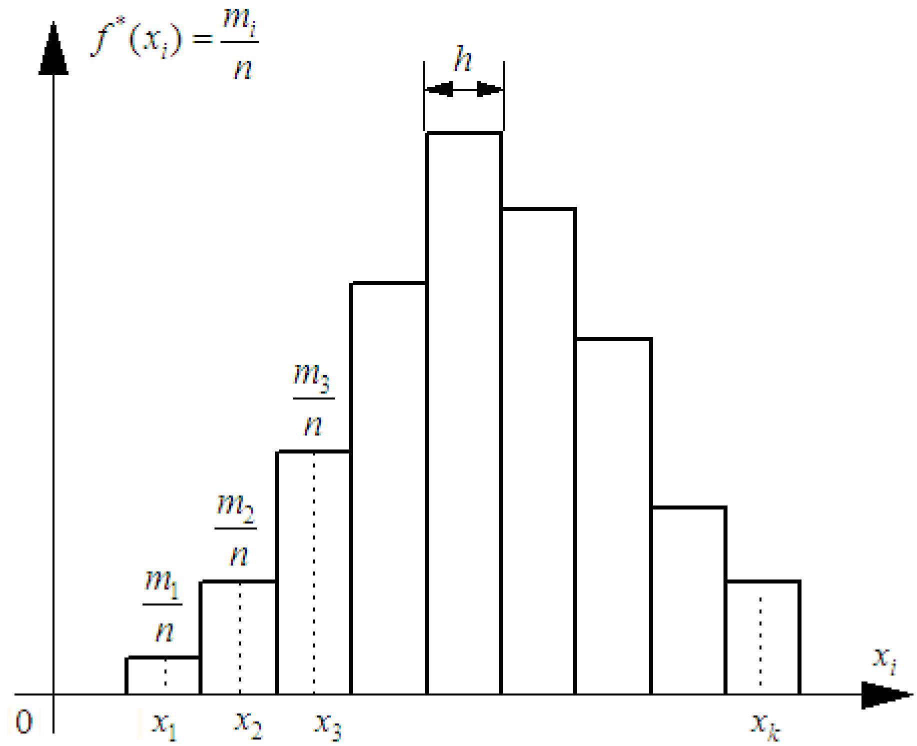
Соответствующая *эмпирическая плотность вероятности* определяется соотношением

$$f^*(x_i)=\frac{m_i}{n}$$

где x_i – середина *i*-ого интервала,

n – объем выборки.

Эмпирическую плотность вероятности изображают в виде гистограммы – фигуры, состоящей из прямоугольников, основаниями которых служат интервалы, а высоты равны отношению частоты для этих интервалов к объему выборки.



Порядок выполнения работы

1. Написать m-функцию, формирующую интервальное распределение выборки. Входной параметр: `x` – вектор элементов выборки. Выходные параметры: `x1` – вектор значений левых границ интервалов, `x2` – вектор значений правых границ интервалов, `x0` – вектор значений центров интервалов, `m` – вектор значений частот.
2. Сформировать выборку `x` объемом $n = 50$ с нормальным, экспоненциальным и равномерным распределениями с помощью m-функций, полученных в лабораторной работе №2.
3. Для полученных выборок сформировать интервальное распределение с помощью m-функции, полученной в п.1. То есть получить векторы значений `x1,x2,x0,m.` Привести получившиеся данные в виде таблицы.
4. С помощью рассчитанных векторов `x1,x2,x0,m` построить графики эмпирических функций распределений.

5. Наложить на полученные графики кривые соответствующих теоретических (генеральных) функций распределения, используя программы, полученные в лабораторной работе №1.
6. Исследовать зависимость вида эмпирических функций распределения от объема выборки n (исследовать сходимость эмпирических функций распределения к генеральным при увеличении объема выборки).

Замечание 1: Формирование ранжированного ряда из выборки `x` осуществляется функцией `sort(x)` (упорядочивание значений выборки по возрастанию).

Замечание 2: График эмпирической плотности вероятности нужно строить с помощью функции `bar`, а эмпирическую функцию распределения выборки – с помощью функции `stairs`.

Замечание 3: Необходимую информацию по языку MATLAB можно найти в Приложении.

ЛАБОРАТОРНАЯ РАБОТА №4

ОЦЕНИВАНИЕ КООРДИНАТЫ ЦЕНТРА РАСПРЕДЕЛЕНИЯ ВЫБОРОЧНОЙ СОВОКУПНОСТИ

Цели работы

1. Знакомство с оценками координаты центра опытного распределения.
2. Изучение алгоритмов расчета оценок центра распределения по выборке экспериментальных данных.
3. Выбор оптимальной оценки для экспериментальных данных с различными видами распределений.

Основные теоретические положения

При многократных измерениях за оценку X^* результата измерения принимается координаты центра опытного распределения $X_{ц.р.}$, т.е.

$$X^* = X_{ö.ð.}\,.$$

Известны несколько оценок координаты центра распределения: среднее арифметическое, медиана, центр срединного размаха, центр размаха.

Среднее арифметическое (выборочное среднее арифметическое) вычисляется по формуле

$$\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i\,,$$

Медианой $\overline{X}_M$ называют значение x_i, которое делит ранжированный ряд на две части, равные по числу элементов

$$\bar{X}_M = \begin{cases} \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right), & \text{åñëè } n - \text{÷åò í î å,} \\ x_{\frac{n+1}{2}}, & \text{åñëè } n - \text{í å÷åò í î å.} \end{cases}$$

Центр размаха $\bar{X}_R$ вычисляется по формуле

$$\bar{X}_R = \frac{x_1 + x_k}{2},$$

где x_1, x_k – крайние значения вариационного ряда.

Центр срединного размаха $\bar{X}_C$ определяется по формуле

$$\bar{X}_C = \frac{1}{2} \cdot \left(x_{0.25} + x_{0.75} \right),$$

где $x_{0.25}, x_{0.75}$ – 25%-ный и 75%-ный квантили – такие элементы ранжированного ряда, левее которых лежат соответственно 25% и 75% всех значений элементов выборки. Для вычисления квантилей $x_{0.25}, x_{0.75}$ могут быть использованы следующие формулы:

	$x_{0.25} = x_k$	$x_{0.75} = x_k$
При *n*, кратном 4	$k = \frac{n}{4} + 1$	$k = \frac{3n}{4}$
При (*n*-1), кратном 4	$k = \frac{n-1}{4} + 1$	$k = n - \frac{n-1}{4}$
При (*n*+1), кратном 4	$k = \frac{n+1}{4} + 1$	$k = n - \frac{n+1}{4}$
Остальные четные *n*	$k = \frac{n+2}{4}$	$k = \frac{3n+2}{4}$

Порядок выполнения работы

1. Написать m-функции для расчета среднего арифметического, медианы, центра размаха, центра срединного размаха. Входной

параметр: `x` – выборка случайных чисел. Выходной параметр: `y` – значение соответствующей оценки координаты опытного распределения.

2. Сформировать с помощью программ, полученных в лабораторной работе №2, выборку `x` с нормальным, экспоненциальным и равномерным распределениями. Параметры распределений выбрать произвольно.
3. Рассчитать для каждой выборки все оценки центра распределения с помощью m-функций, полученных в п.1. Привести результаты расчетов в виде таблицы.
4. По результатам расчетов сделать выбор оптимальной оценки центра распределения для каждого рассматриваемого закона.

Замечание 1: Объемы выборок должны быть большими для получения более точного результата.

Замечание 2: Необходимую информацию по языку MATLAB можно найти в Приложении.

ЛАБОРАТОРНАЯ РАБОТА №5

ПРОВЕРКА ПРИНАДЛЕЖНОСТИ ВЫБОРОЧНОЙ СОВОКУПНОСТИ К НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ

Цели работы

1. Изучение методики проверки нормальности опытного распределения с помощью критерия Пирсона.
2. Реализация изученных алгоритмов в MATLAB.

Основные теоретические положения

Критерии для проверки статистических гипотез о принадлежности выборки к конкретным законам распределения называются *критериями согласия*.

Критерий χ^2 Пирсона эффективен при большом объеме выборки (условно $n > 50$).

Идея критерия состоит в контроле отклонений гистограммы исследуемой выборки от гистограммы с таким же числом интервалов, построенной на основе нормального распределения.

Для использования критерия необходимо, чтобы в каждый интервал статистического распределения попадало не меньше 5 значений. В случае, если это не так, следует объединить несколько расположенных рядом интервалов в один.

Алгоритм проверки гипотезы по критерию Пирсона:

1. Выдвигается нулевая гипотеза: “генеральная совокупность распределена нормально”. Выбирается значение доверительной вероятности p.

2. Рассчитывается для каждого интервала параметр

$$z_i = \frac{\left| x_{i0} - \bar{X} \right|}{S^*},$$

где x_{i0} – середина соответствующего интервала,

$S^* = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}$ – несмещенная оценка стандартного отклонения выборки,

$\bar{X}$ – выборочное среднее арифметическое,

n – объем выборки,

3. По значению z_i определяется значение дифференциальной функции стандартного нормального распределения $f(z_i)$ (вероятность попадания значений в интервал в случае, если распределение нормальное)

$$f(z_i) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z_i^2}{2}\right\}$$

4. Вычисляются теоретические частоты попадания значений выборки в i-й интервал

$$m_i' = \frac{n \cdot h}{S^*} \cdot f(z_i),$$

где h – длина интервала,

S^* – несмещенная оценка стандартного отклонения выборки.

5. В качестве наблюдаемого значения критерия проверки нулевой гипотезы (“генеральная совокупность распределена нормально”) принимается случайная величина, определяемая по формуле

$$\chi^2_{\text{í àáë}} = \sum_{i=1}^{k}\frac{\left(m_i - m_i'\right)^2}{m_i'},$$

где m_i– эмпирические частоты (частота попадания значений в i-й интервал),

m_i' – теоретические частоты, вычисленные в предположении нормально-распределенной генеральной совокупности,

k – количество интервалов.

6. Вычисляется число степеней свободы

$$r = k - 3,$$

где k – количество интервалов,

7. При заданном значении доверительной вероятности p определяется квантиль распределения Пирсона $\chi_p^2(r)$, значение которого является критическим значением критерия. Квантиль распределения Пирсона, соответствующий заданной доверительной вероятности p можно вычислить с помощью аппроксимации Корниша-Фишера

$$\chi_p^2(r) = r + A\sqrt{r} + B + \frac{C}{\sqrt{r}} + \frac{D}{r} + \frac{E}{r\sqrt{r}},$$

где $A = d\sqrt{2}$, $B = \frac{2}{3}\left(d^2 - 1\right)$, $C = d \cdot \frac{d^2 - 7}{9\sqrt{2}}$, $D = \frac{6d^4 + 14d^2 - 32}{405}$,

$$E = d \cdot \frac{9d^4 + 256d^2 - 433}{4860\sqrt{2}}.$$

$$d = 2.0637 \cdot \left(\ln\frac{1}{1-p} - 0.16\right)^{0.4274} - 1.5774, \quad \left(\text{ï ðè } 0.5 \le p \le 0.999\right),$$ – p-квантиль стандартного нормального распределения ,

8. Нулевая гипотеза принимается с заданной вероятностью, если

$$\chi_{\text{íàáë}}^2 \le \chi_p^2(r).$$

Порядок выполнения работы

1. Написать m-функцию для расчета *p*-квантиля распределения Пирсона. Входные параметры: `r` – число степеней свободы, `p` – доверительная вероятность. Выходной параметр: `y` – значение *p*-квантиля распределения Пирсона.
2. Написать m-функцию для проверки нормальности опытного распределения с помощью критерия Пирсона. Входные параметры: `x` – выборка случайных чисел, `p` – доверительная вероятность. Выходной параметр: `y` – принимает значение 1, если нулевая гипотеза принимается, 0 – в противном случае. Для формирования интервального распределения использовать программы, полученные в лабораторной работе №3. Для расчета критического значения критерия использовать m-функцию, полученную в п.1.
3. Сформировать выборку `x` объемом $n = 1000$ с нормальным и равномерным распределениями, используя программы, полученные в лабораторной работе №2.
4. Проверить принадлежность полученных выборок к нормальному распределению с помощью m-функции, полученой в п.2. Убедиться, что функция возвращает значение 1 в случае нормального закона распределения и 0 – в случае равномерного.

ПРИЛОЖЕНИЕ

СПРАВОЧНЫЕ СВЕДЕНИЯ ПО ЯЗЫКУ MATLAB

Описание языка MATLAB

Язык программирования MATLAB является интерпретируемым. Этап компиляции полной программы отсутствует и исполняемые программы не создаются. Программы существуют лишь в виде m-файлов (файлов с расширением m) и для их выполнения необходимо находиться в среде MATLAB. Однако существуют технологии автоматизированного перевода программ на языке MATLAB в коды языков программирования C и C++.

Существует 2 вида программ: m-функции и m-скрипты (сценарии).

Сценарий представляет собой последовательность команд без входных и выходных параметров. Он имеет следующую структуру:

```
% Основной комментарий
% Дополнительный комментарий
Тело файла
```

Комментарии нужны для того, чтобы ознакомиться с назначением файла через справочную систему. Основной комментарий выводится при исполнении команды

```
>> help имя_каталога
```

Полный комментарий выводится при исполнении команды

```
>> help имя_файла
```

Для выполнения сценария в среде Matlab необходимо ввести его имя

```
>> имя_файла
```

и нажать клавишу ENTER.

М-функция является типичным объектом языка программирования системы Matlab. Структура m-функции с одним выходным параметром выглядит следующим образом:

```
function var=func_name(par1,…,parN)
% Основной комментарий
% Дополнительный комментарий
Тело файла
var=выражение;
```

Здесь переменная `var` – выходной параметр, `par1,…,parN` – входные параметры, `func_name` – имя функции. Функция возвращает свое значение `var` и может использоваться в математических выражениях в виде

```
func_name(список параметров);
```

Файл m-функции должен иметь то же имя `func_name`, что и сама функция (т.е. `func_name.m`).

Все переменные, имеющиеся в теле файла-функции, являются *локальными*, то есть действуют только в пределах тела функции, в отличие от файла-сценария, все переменные которого являются *глобальными*.

Правила вывода комментариев те же, что и у файлов-сценариев.

Последняя конструкция `var=выражение` вводится, если требуется, чтобы функция возвращала результат вычислений. М-функция завершается по достижению конца файла. Для досрочного выхода из функции и возврата значения функции используется программный оператор `return`.

Если выходных параметров больше одного, то структура файла имеет вид:

```
function [var1,…,varN]=func_name(par1,…,parM)
% Основной комментарий
% Дополнительный комментарий
Тело файла с любыми выражениями
var1=выражение;
…
varN=выражение;
```

Данная функция вызывается в виде:

```
[var1,var2,…]=func_name(par1,…,parM);
```

Если такая функция используется в виде

```
func_name(par1,…,parM);
```

то возвращается значение только первого выходного параметра – переменной `var1`.

Циклы и операторы выбора

<table>
<tr>
<td><pre>for var= b:s:e
  Инструкция1;
  Инструкция2;
  …
end</pre></td>
<td>Используется для организации вычислений с заданным числом повторений цикла
b – начальное значение переменной цикла var,
s – приращение (шаг) этой переменной,
e–конечное значение управляющей переменной, при достижении</td>
</tr>
</table>

	которого цикл завершается.

`while Условие`
`  Инструкция1;`
`  Инструкция2;`
`  …`
`end`
	Цикл выполняется до тех пор, пока выполняется `Условие`

`if Условие1`
`   Инструкции1`
`[elseif Условие2`
`   Инструкции2`
`else`
`   Инструкции3]`
`end`
	Оператор выбора: позволяет сделать выбор хода вычислений в зависимости от условий. Если выполняется `Условие1`, то вызываются `Инструкции1`. В противном случае проверяется `Условие2`. Если выполняется `Условие2`, то вызываются `Инструкции2`. Если `Условие2` не выполняется, то вызываются `Инструкции3`. Квадратными скобками выделены необязательные элементы оператора (при использовании оператора скобки ставить не нужно)

`switch`
`Выражение0`
`case Выражение1`
`     Инструкции1`
`case Выражение2`
`     Инструкции2`
`…`
`otherwise`
`     Инструкции`
`end`
	Оператор множественного выбора. Выполняются `ИнструкцииN`, если `ВыражениеN` соответствует `Выражению0`. Если ни одно из выражений после операторов `case` не соответствует `Выражению0`, то выполняются `Инструкции` после `otherwise`. `Выражение1`, `Выражение2`,… не должны быть равны между собой.
	Выражение после `case` может быть перечислением
	`case {выражение1_1, выражение1_2, выражение1_3}`
	В этом случае инструкции после case выполняются, если `Выражение0` соответствует одному из перечисленных выражений.

`Условие` в операторах записывается в виде:

`Выражение1 Оператор_отношения Выражение2`

В качестве оператора отношения могут быть использованы:

Оператор	Значение
`>`	больше
`<`	меньше
`>=`	не меньше

`<=`	не больше
`==`	равно
`~=`	не равно
`&`	логическое И
`|`	логическое ИЛИ

Представление данных в MATLAB

Все данные в MATLAB представляются в виде массивов.

Массив – это упорядоченная, пронумерованная совокупность однородных данных. Массивы различаются по числу размерностей (измерений): одномерные, двумерные, многомерные.

Размер массива – это количество элементов вдоль каждого из измерений. Номер элемента в массиве называется *индексом*. В отличие от других языков программирования, в MATLAB нумерация элементов начинается не с нуля, а с единицы, поэтому индексы всегда ≥ 1.

Массивы являются способом хранения математических объектов – векторов и матриц. Скалярные величины хранятся в одномерных массивах с размерностью 1. Векторы могут записываться в виде строки или в виде столбца. В этом случае они называются соответственно *вектором-строкой* или *вектором-столбцом*.

Доступ к элементу массива (вектора) осуществляется с помощью его индексов (номера строки и номера столбца), которые записываются в круглых скобках после имени массива `a(2,5)`. Если вместо одного из индексов поставить двоеточие `:`, то это равносильно обращению ко всему столбцу или всей строке `a(:,5)`.

Удаление элемента матрицы осуществляется присвоением ему пустого массива, который обозначается квадратными скобками `[]`. Например `a(2,5)=[]`.

К векторам и матрицам можно применять арифметические операции

Оператор MATLAB	Назначение
`+`	сложение

-	вычитание
*	умножение
/	деление
^	степень

Следует помнить, что все арифметические операторы в MATLAB представляют матричные операции. Так например, выражение `C=A*B` вычислит произведение матриц A и B. Для того, чтобы выполнить поэлементные действия, нужно перед соответствующим арифметическим оператором ставить точку. В этом случае, например, выражение `C=A.*B` вычислит матрицу, состоящую из произведения соответствующих элементов матриц A и B, т.е. $c_{ij} = a_{ij} \cdot b_{ij}$.

Основные функции для работы с массивами.

Функция MATLAB	Назначение/результат выполнения
d=det(A)	d - определитель матрицы A
n=length(A)	n - длина вектора A
[n,m]=size(A)	n – количество строк матрицы A m – количество столбцов матрицы A
A=zeros(n,m)	A – матрица размером m×n, состоящая из нулей
A=eye(n,m)	A – единичная матрица размером m×n, (все элементы на главной диагонали раны 1, остальные элементы раны 0)
A=ones(n,m)	A – матрица размером m×n, состоящая из единиц
B=A′	B – матрица, полученная транспонированием матрицы A

Основные функции обработки статистических данных

Функция	**Результат**
min(X)	Наименьшее значение элемента выборки *X*
max(X)	Наибольшее значение элемента выборки *X*
mean(X)	Выборочное среднее арифметическое

`trimmean(X,e)`	Среднее арифметическое е-процентной выборки
`mad(X)`	Среднее абсолютное отклонение от среднего значения выборки *X*
`median(X)`	Медиана выборки *X*
`sum(X)`	Сумма элементов выборки *X*
`cumsum(X)`	Накопленная сумма элементов выборки *X*. (Например, если *X* – вектор значений частот статистического распределения, то результатом выполнения функции будет вектор значений накопленной частоты)
`std(X)`	Несмещенная оценка стандартного отклонения (СКО) выборки *X*
`var(X)`	Несмещенная оценка дисперсии выборки *X*
`sort(X)`	Ранжированный ряд, составленный из выборки *X*
`prod(X)`	Произведение элементов выборки *X*
`moment(X,k)`	Оценка центрального момента порядка k выборки *X*
`range(X)`	Размах выборки *X*
`kurtosis(X)-3`	Оценка коэффициента эксцесса выборки *X*
`skewness(X)`	Оценка коэффициента асимметрии выборки *X*
`prctile(X,p)`	p-квантиль (процентиль) выборки *X*

Форматы вывода результатов вычислений и функции округления

По умолчанию результаты вычислений представляются в формате short (4 знака после запятой). Для отображения большего количества знаков (например, при вычислении квантилей распределений) необходимо представить результат в формате long (15 знаков после запятой) с помощью инструкции `format long`. Эта инструкция вводится до вывода результата. Для возвращения к формату short следует выполнить инструкцию `format short`.

Для округления используются следующие функции

Функция MATLAB	Назначение/результат выполнения
round(x)	Округление числа x до ближайшего целого
ceil(x)	Округление числа x до ближайшего целого в сторону $+\infty$
floor(x)	Округление числа x до ближайшего целого в сторону $-\infty$
fix(x)	Округление числа x до ближайшего целого в сторону 0

Построение и оформление графиков функций

График функции, заданной таблицей экспериментальных данных строится с помощью функции

```
plot(x,y,s)
```

где x – вектор значений аргумента,

y – вектор значений функции,

s – строка, определяющая цвет и тип линии, а также маркер точек графика

Строка **s**	Значение
	Цвет линии
‘y’	Желтый
‘m’	Розовый
‘c’	Голубой
‘r’	Красный
‘w’	Белый
‘g’	Зеленый
‘b’	Синий
‘k’	Черный
	Тип линии
‘-’	Сплошная линия
‘:’	Пунктирная линия
‘-.’	Штрихпунктирная линия
‘--’	Штриховая линия

Строка **s**	Значение
	Тип маркера
‘.’	Точка
‘o’	Круг
‘x’	Крест
‘+’	
‘*’	
‘s’	Квадрат
‘d’	Ромб
‘v’	Треугольник вершиной
‘^’	вниз
‘<’	Треугольник вершиной
‘>’	вверх
‘p’	Треугольник вершиной
‘h’	влево
	Треугольник вершиной
	вправо
	Пятиконечная звезда

	Шестиконечная звезда

Строка s формируется из комбинации значений типа и цвета линии и типа маркера. Так, например, 'k-.p' указывает, что график будет построен черной штрихпунктирной линией, а точки будут отмечены пятиконечной звездой.

Если указать тип маркера точек, но не указывать тип линии, то на график будут нанесены только точки указанным маркером без соединяющей их линии.

Функция plot(y) строит график зависимости значений элементов вектора y от их индекса.

Функция plot(x,y) строит график зависимости *y=f(x)* сплошной синей линией.

Для построения графика в отдельном окне нужно предварительно вызвать функцию figure (пустое графическое окно). С помощью этой функции можно установить цвет графического окна. Например, figure('Color','w') устанавливает белый цвет и открывает новое графическое окно.

Для нанесения координатной сетки на график используется инструкция grid on, которую нужно указать сразу после функции plot.

Для построения нескольких графиков в одной системе координат возможны два варианта. Пусть нужно отобразить зависимости *y1=f(x1)* и *y2=f(x2)* .

- вариант 1:

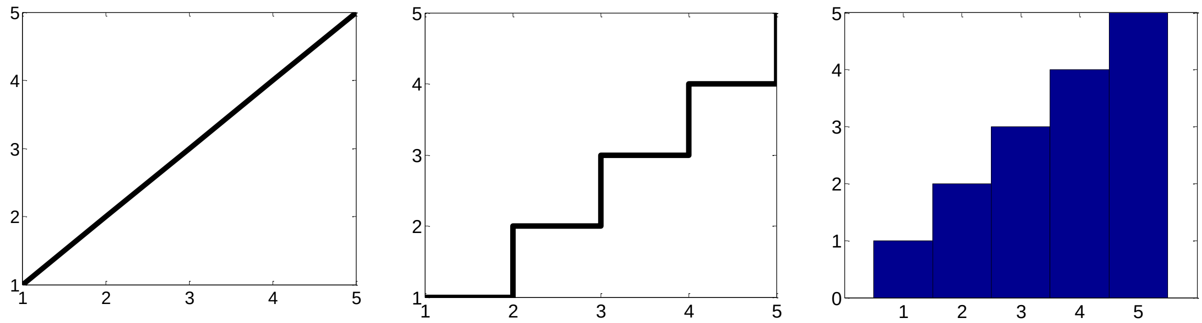
```
plot(x1,y1,s1,x2,y2,s2)
```

- вариант 2:

```
plot(x1,y1,s1)
hold on
plot(x2,y2,s2)
```

Для построения лестничного графика используется функция `stairs(x,y,s)`, формат вызова которой аналогичен формату функции `plot`.

Для построения столбчатой диаграммы используется функция `bar(x,y,w,s)`, формат вызова которой аналогичен формату функции `plot`, за исключением параметра `w`, значение которого определяет ширину стобца. При `w =1` столбцы примыкают друг к другу.



`plot(x,y)` `stairs(x,y)` `bar(x,y,1)`

Функции для оформления графиков

Функция	Назначение
`grid on`	Отображение координатной сетки
`title('Текст')`	Заголовок графика
`xlabel('Текст')`	Подпись оси абсцисс
`ylabel('Текст')`	Подпись оси ординат
`legend('Текст1',…,'ТекстN')`	Легенда

Легенда представляет собой поясняющие надписи к линиям графика. Количество надписей должно соответствовать количеству линий на графике. Если указать после текстов надписей числовой параметр `p`

`legend('Текст1',…,'ТекстN',p)`

то можно управлять положением легенды на графике

Значение p	Положение легенды
-1	В правом верхнем углу графического окна (вне графика)
0	Положение, при котром легенда наименьшим образом перекрывает графики

1	В верхнем правом углу графика (по умолчанию)
2	В верхнем левом углу графика
3	В нижнем левом углу графика
4	В нижнем правом углу графика

Можно использовать формат TeX в надписях.

СПИСОК ЛИТЕРАТУРЫ

1. Ануфриев И.Е. Самоучитель MatLab 5.3/6.x. – СПб.: БХВ-Петербург, 2004.
2. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения: Учеб. пособие для вузов. – М.: Высш. шк., 2000.
3. Гмурман В.Е. Теория вероятностей и математическая статистика: Учеб. пособие для вузов. – М.: Высш. шк., 2003.
4. Дьяконов В., Круглов В. Математические пакеты расширения MATLAB. Специальный справочник. – СПб.: Питер, 2001.
5. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006.
6. Половко А.М., Бутусов П.Н. MATLAB для студента. – СПб.: БХВ-Петербург, 2005.