**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИИ**
**Дагестанский государственный университет**

**НАЗАРАЛИЕВ М.А., БЕЙБАЛАЕВ В.Д.**

# ЛАБОРАТОРНЫЙ ПРАКТИКУМ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ В СРЕДЕ MATLAB
## (Примеры и методические указания к выполнению лабораторных работ)

**Махачкала 2016**

Печатается по решению редакционно-издательского совета Дагестанского государственного университета.

Назаралиев М.А., Бейбалаев В.Д.
Лабораторный практикум по математической статистике в среде MATLAB: Учебное пособие.- Махачкала: Издательство ДГУ, 2016.- 53 с.

Главной целью пособия является научить студентов пользоваться методами математической статистики с использованием современных информационных технологий. Наиболее подходящей для этой цели является одна из самых развитых и эффективных математических систем - MatLAB, которая занимает особое место среди множества таких систем (MathCAD, Maple, Mathematica и др.). Пособие включает в себя также и лабораторный практикум по математической статистике, апробированный на занятиях со студентами ДГУ, обучающимися по естественно-научным направлениям.

Предназначено для направления подготовки бакалавров на естественнонаучных факультетах высших учебных заведений.

Рецензент: д. ф.-м. н., профессор Ахмедов С.А..

ИПЦ ДГУ 2016

# Содержание

## Введение

Одним из важных направлений развития современных вычислительных технологий в настоящее время является появление мощных математических пакетов таких, как MAPLE, MathCAD, Mathematica, MATLAB. При помощи таких пакетов одной командой выполняются решения математических задач, например, обращение матриц, решение проблемы собственных чисел, вычисление пределов, разложение функции в ряд, решение дифференциальных уравнений и многих других.

Среди множества математически ориентированных пакетов система научно- инженерных расчётов MATLAB в одинаковой мере ориентирована на применение как символьных, так и численных методов. Включённое в MATLAB ядро Maple хорошо справляется с символьными вычислениями, а сам MATLAB и его инструментарии прекрасно работают с числами. Особенностью MATLAB является его способность встраиваться в другие приложения Windows и динамически обмениваться с ними данными.

## §1. Случайные величины и их распределения. Основные дискретные и непрерывные распределения случайных величин

Величина, которая в результате опыта или испытания принимает одно и только одно возможное числовое значение из заранее известной совокупности значений называют *случайной величиной.*

*Случайная величина*, которая в результате испытания принимает либо конечное множество, либо счетное множество возможных значений называют *дискретной случайной величиной*.

Если множество возможных значений несчетно, то случайная величина *называется непрерывной.*

*Законом распределения* дискретной случайной величины называется правило, которое устанавливает связь между возможными значениями случайной величины и соответствующими им вероятностями.

При табличном способе задания закона распределения первая строка таблицы содержит возможные значения случайной величины (обычно в порядке возрастания), а вторая – соответствующие вероятности ( $\Sigma p_i = 1$ ):

| $x_i$ | $x_1$ | $x_2$ | … | $x_n$ |
|---|---|---|---|---|
| $p_i$ | $p_1$ | $p_2$ | … | $p_n$ |

Одним из универсальных способов, пригодным как для дискретных, так и для непрерывных случайных величин, является ее функция распределения.

*Функцией распределения случайной величины* X называется функция $F(x)$, которая для любого числа $x \in R$ равна вероятности события $\{X < x\}$.

Функция распределения обладает следующими свойствами:

1. Функция распределения $F(x)$ ограничена, т.е. $0 \le F(x) \le 1$.

2. Функция распределения $F(x)$ на числовой оси R есть неубывающая функция, т.е. если $x_2 > x_1$, то $F(x_2) \ge F(x_1)$.

3. Функция распределения $F(x)$ на минус бесконечности обращается в ноль и в плюс бесконечности равна единице, т.е. $F(-\infty) = 0, F(+\infty) = 1$.

4. Вероятность попадания случайной величины X в промежуток $[a,b)$ равна приращению ее функции распределения на этом промежутке, т.е.

$$P\{a \le X < b\} = F(b) - F(a).$$

5. Функция распределения $F(x)$ есть непрерывная слева функция, т.е. $\lim\limits_{x \to x_0 - 0} F(x) = F(x_0)$.

Одним из важнейших характеристик непрерывной случайной величины (помимо функции распределения) является ее плотность распределения вероятностей.

Первая производная от функции распределения случайной величины называют ее *плотностью распределения вероятностей*.

Обозначается плотность распределения вероятностей непрерывной случайной величины через $f(x)$ или $p(x)$. Таким образом, согласно определению

$$f(x) = F^{/}(x).$$

Свойства плотности распределения:

1. Плотность распределения $f(x)$ случайной величины X есть неотрицательная функция, т.е. $f(x) \ge 0$.
2. Вероятность того, что непрерывная случайная величина принимает возможные значения из промежутка $[a,b]$ равна определенному интегралу от ее плотности в пределах от $a$ до $b$, т.е. $P(a \le X \le b) = \int\limits_a^b f(x)dx$.
3. Функция распределения непрерывной случайной величины можно представить через плотность распределения в виде: $F(x) = \int\limits_{-\infty}^{x} f(t)dt$.
4. Несобственный интеграл от плотности вероятности в бесконечных пределах равен единице, т.е.

$$\int\limits_{-\infty}^{+\infty} f(x)dx = 1.$$

## 1.1. Основные распределения дискретных случайных величин

**Биномиальный закон распределения.** Дискретная случайная величина распределена по биномиальному закону распределения, если она принимает возможные значения $0,1,\ldots,n$ с вероятностями $p_k = P(X = k) = C_n^k p^k q^{n-k}$, где $0 < p < 1$, $q = 1 - p$, $k = 0,1,\ldots,n.$

Ряд распределения имеет вид:

| | | | | |
|---|---|---|---|---|
| $X$ | 0 | 1 | $\ldots$ | $n$ |
| $P$ | $q^n$ | $C_n^1 pq^{n-1}$ | $\ldots$ | $p^n$ |

$\sum\limits_{k=0}^{n} p_k = (p+q)^n = 1$, $MX = np$, $DX = npq$.

**Геометрическое распределение.** Дискретная случайная величина распределена по геометрическому закону распределения, если она принимает возможные значения $1,2\ldots,k,\ldots$ с вероятностями $p_k = P(X = k) = q^{k-1}p$, где $0 < p < 1$, $q = 1 - p$, $k = 1,2\ldots,k,\ldots$

Ряд распределения имеет вид:

| | | | | |
|---|---|---|---|---|
| $X$ | 1 | 2 | 3 | $\ldots$ |
| $P$ | $q$ | $qp$ | $q^2 p$ | $\ldots$ |

$\sum\limits_{k=1}^{\infty} p_k = p \sum\limits_{k=1}^{\infty} q^{k-1} = p\dfrac{1}{1-q} = 1$, $MX = \dfrac{1}{p}$, $DX = \dfrac{q}{p^2}$.

**Распределение Пуассона.** Дискретная случайная величина распределена по закону Пуассона, если она принимает возможные значения $0,1,2\ldots,k,\ldots$ с вероятностями $p_k = P(X = k) = \dfrac{a^k e^{-a}}{k!}$, где $a = np, 0 < p < 1$, $q = 1 - p$, $k = 0,1,2\ldots,k,\ldots$

Ряд распределения имеет вид:

| | | | | | |
|---|---|---|---|---|---|
| $X$ | 0 | 1 | $\ldots$ | $k$ | $\ldots$ |
| $P$ | $e^{-a}$ | $ae^{-a}$ | $\ldots$ | $\dfrac{a^k e^{-a}}{k!}$ | $\ldots$ |

$\sum\limits_{k=0}^{\infty} p_k = e^{-a} \sum\limits_{k=0}^{\infty} \dfrac{a^k}{k!} = e^{-a} e^{a} = 1$, $MX = a$, $DX = a$.

## 1.2. Основные распределения непрерывных случайных величин

**Равномерный закон распределение.** Непрерывная случайная величина $X$ распределена по равномерному закону на отрезке $[a,b]$, если плотность распределения ее имеет вид:

$$f(x)=\begin{cases}\dfrac{1}{b-a}, \ при\ x\in[a,b],\\ 0, \quad при \quad x\notin[a,b].\end{cases}$$

Функция распределения случайной величины $X \sim R[a,b]$ имеет вид:

$$F(x)=\begin{cases}0, \ при\ x\le a,\\ \dfrac{x-a}{b-a}, \ при \ \ a<x\le b,\\ 1, \quad при \quad x>b.\end{cases}$$

$MX=\dfrac{a+b}{2}, \quad DX=\dfrac{(b-a)^2}{12}$.

**Показательный закон распределения.** Непрерывная случайная величина $X$ распределена по показательному закону распределения, если плотность распределения ее имеет вид:

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, \ при\ x\ge 0,\\ 0, \quad при \quad x<0.\end{cases}$$

Функция распределения случайной величины $X \sim P(\lambda)$ имеет вид:

$$F(x)=\begin{cases}0, \ при\ x\le 0,\\ 1-e^{-\lambda x}, \ при\ x>0.\end{cases}$$

$MX=\dfrac{1}{\lambda}, \quad DX=\dfrac{1}{\lambda^2}$.

**Нормальный закон распределения.** Непрерывная случайная величина $X$ распределена по нормальному закону с параметрами $a,\sigma$, если плотность распределения ее имеет вид:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x\in R.$$

Функция распределения случайной величины $X \sim N(a,\sigma)$ имеет вид:

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}dt.$$

$MX=a, \quad DX=\sigma^2$.

**Распределение Релея.** Непрерывная случайная величина $X$ распределена по закону Релея, если плотность распределения ее имеет вид:

$$f(x)=\begin{cases}0 & , \quad x<0\\ \dfrac{x}{\sigma^2}e^{-\frac{x^2}{2\cdot\sigma^2}} & , \quad x\geq 0\end{cases} \quad ; \; M[X]=\sigma\cdot\sqrt{\frac{\pi}{2}} \; .$$

## §2. Статистическое распределение выборки. Эмпирическая функция распределения полигон и гистограмма

Совокупность всех подлежащих к изучению объектов или возможных результатов всех мыслимых наблюдений, производимых в неизменных условиях над одним объектом, называется *генеральной совокупностью*.

Совокупность всех отобранных случайным образом объектов из генеральной совокупности называют *выборкой*.

Ряд составленный из элементов выборки, расположенных по неубыванию, называется *вариационным рядом*.

*Частотами* называют числа $n_i$, показывающие, сколько раз варианта $x_i$ встречается в выборке, а их отношение к объему выборки называют относительными частотами.

*Статистическим рядом распределения* выборки называется перечень возможных значений и соответствующих частот или относительных частот.

| | | | | |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | $\dots$ | $x_k$ |
| $n_i$ | $n_1$ | $n_2$ | $\dots$ | $n_k$ |

Когда число элементов выборки велико или признак непрерывный, составляют интервальный статистический ряд. В первую строчку статистического ряда вписывают частичные интервалы $[x_0,x_1),[x_1,x_2),\dots,[x_{k-1},x_k)$, которые обычно имеют одинаковую длину, а во вторую строчку вписывают число наблюдений $n_i$, $i=1,2,\dots,k$, попавших в каждый интервал.

| | | | | |
|---|---|---|---|---|
| Интер. | $[x_0,x_1)$ | $[x_1,x_2)$ | $\dots$ | $[x_{k-1},x_k)$ |
| $n_i$ | $n_1$ | $n_2$ | $\dots$ | $n_k$ |

Обычно длину интервала определяют с помощью формулы Стерджеса

$$h=\frac{x_{\max}-x_{\min}}{1+\log_2 n},$$

где $x_{max} - x_{min}$ - разность между максимальным и минимальным значениями признака, $m = 1 + \log_2 n$ - число интервалов.

*Эмпирической функцией распределения* называется функция $F_n^*(x)$, определяющая для каждого значения $x$ относительную частоту события $X < x$, т.е.

$$F_n^*(x) = \frac{n_x}{n},$$

где $n$ - объем выборки, $n_x$ - число наблюдений меньших $x$.

*Полигоном частот* называют ломанную, звеньями которой являются отрезки, соединяющие последовательно точки $(x_1, n_2), (x_2, n_2), ..., (x_k, n_k)$.

*Гистограммой частот* (*относительных частот*) называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины $h$, а высоты равны отношениям $\frac{n_i}{h}$ - плотности частот ($\frac{n_i}{n \cdot h}$ - плотности относительных частот). При этом площадь гистограммы частот равна объему выборки, а площадь гистограммы относительных частот единице.

**Числовые характеристики статистического распределения.**

*Выборочным средним* называется среднее арифметическое значений выборки, т.е.

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \text{ или же } \overline{x}_в = \sum_{i=1}^{k} x_i \nu_i,\ \nu_i = \frac{n_i}{n}.$$

*Выборочной дисперсией* называется среднее арифметическое квадратов отклонений выборочных значений от выборочного среднего, т.е.

$$D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 n_i \text{ или же } D_в = \sum_{i=1}^{k}(x_i - \overline{x}_в)^2 \nu_i .$$

Величина, которая определяется формулой

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 n_i ,$$

называется *исправленной выборочной дисперсией*.

*Выборочным средним квадратическим отклонением* называется квадратный корень от выборочной дисперсии

$$\sigma_в = \sqrt{D_в},$$

а исправленным выборочным средним квадратическим отклонением квадратный корень из исправленной выборочной дисперсии

$$S = \sqrt{S^2}\ .$$

Варианту с наибольшей частотой называют *модой* $M_o^*$ вариационного ряда, а значение признака, приходящееся на середину ряда, называется *медианой* $M_e^*$.

Решение уравнения $F_X(x_p) = p$, где $p$ некоторое число, удовлетворяющее условию $0 < p < 1$, называется *квантилью уровня* $p$. При этом $x_{0,25}$-нижняя квантиль, $x_{0,5}$- медиана, $x_{0,75}$-верхняя квантиль.

## §3. Статистические оценки параметров распределения

*Статистической оценкой* $\tilde{\Theta}_n$ неизвестного параметра $\Theta$ теоретического распределения называется статистика, которая принимается в качестве приближенного значения параметра и зависящая от данных выборки.

Качество оценки зависит от свойств несмещенности, состоятельности и эффективности оценок.

Если $M\tilde{\Theta} = \Theta$, то оценка называется *несмещенной*. В противном случае оценка называется *смещенной*.

Если $\tilde{\Theta}_n \xrightarrow[n \to \infty]{P} > \Theta$ ($\Theta_n$ стремиться по вероятности к оцениваему параметру $\Theta$), то оценка называется *состоятельной*. Несостоятельные оценки не рассматривают.

Несмещенная оценка $\tilde{\Theta}_n$ с наименьшей дисперсией среди всех несмещенных оценок называется *эффективной.*

Оценка, определяемая двумя числами- концами интервала, называется *интервальной*.

Интервал $(\overline{\Theta}_1, \overline{\Theta}_2)$, покрывающий с вероятностью $\gamma$ истинное значение параметра $\Theta$, называется *доверительным интервалом*, а вероятность $\gamma$- *надежностью оценки или доверительной вероятностью.*

**Доверительный интервал для математического ожидания при известной дисперсии.** Доверительный интервал для математического ожидания $MX$ при известной дисперсии $\sigma^2$ имеет вид:

$$\left(\overline{X} - t \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + t \cdot \frac{\sigma}{\sqrt{n}}\right),$$

где $t$ определяется из равенства $\Phi_0(t) = \frac{\gamma}{2}$ (или $\Phi(t) = \frac{1+\gamma}{2}$ ). При заданном значении $\gamma$ по таблице функции Лапласа $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$ находим аргумент $t$.

**Доверительный интервал для математического ожидания при неизвестной дисперсии.** Доверительный интервал для математического ожидания при неизвестной дисперсии имеет вид:

$$\left(\overline{X} - t_\gamma \cdot \frac{S}{\sqrt{n}}, \overline{X} + t_\gamma \cdot \frac{S}{\sqrt{n}}\right),$$

где $t_\gamma$ определяется из таблицы квантилей распределения Стьюдента в зависимости от доверительной вероятности $\gamma$ и числа степеней свободы $n-1$ ($t_\gamma$ - квантиль уровня $1-\gamma$), $S$ - исправленное выборочное среднее квадратическое отклонение.

**Доверительный интервал для среднего квадратического отклонения нормального распределения.** Доверительный интервал для среднего квадратического отклонения $\sigma$ нормального распределения имеет вид:

$$\left(\frac{\sqrt{n} \cdot S_0}{\chi_2}, \frac{\sqrt{n} \cdot S_0}{\chi_1}\right),$$

где $n$- объем выборки, $S_0^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - a)^2$, а $\chi_1^2 = \chi^2_{\frac{1+\gamma}{2}, n}$, $\chi_2^2 = \chi^2_{\frac{1-\gamma}{2}, n}$ являются квантилями $\chi^2$- распределения с $n$ степенями свободы, определяемые по таблице квантилей $\chi^2_{\alpha, n}$.

**Доверительный интервал для вероятности успеха $p$.** Доверительный интервал, который при больших значениях $n$ с надежностью $\gamma$ покрывает оцениваемый параметр $p$ имеет вид $(p_1, p_2)$, где

$$p_1 = \nu - t\sqrt{\frac{\nu(1-\nu)}{n}}, \quad p_2 = \nu - t\sqrt{\frac{\nu(1-\nu)}{n}},$$

$\nu = \dfrac{n_A}{n}$ - относительная частота события, $t$ определяется из равенства $\Phi_0(t) = \dfrac{\gamma}{2}$.

## §4. Знакомство с функциями пакета Statistics Toolbox системы MATLAB

Приведём список некоторых функций пакета Statistics Toolbox. Для получения полного описания функции, например, stats следует в командном окне MATLAB ввести команду help stats.

### 4.1. Функции распределения

Следующие функции пакета вычисляют значения функций распределений $F(x)$ для различных законов вероятностей. Все имена функций пакета имеют суффикс *cdf (сокращённое от cumulative distribution function – функции распределения).

- binocdf – биномиальное распределение;
- poisscdf – распределение Пуассона;
- normcdf – нормальное распределение;
- unifcdf – равномерное распределение;
- chi2cdf – $\chi^2$- распределение Пирсона;
- tcdf – $t$ –распределение Стьюдента.

Для вычисления вероятности события $P(\alpha < X < \beta)$, например, нормально распределенной случайной величины $X$ с параметрами $a = 0, \sigma = 1$ в интервале (0;2) нужно в командной строке набрать следующие команды

```
>> a = 0
>> sigma = 1
>> alfa = 0
>> betta = 2
>> normcdf (betta, a, sigma) – normcdf (alfa, a, sigma)
```

## 4.2. Плотности распределения.

С помощью этих функций пакета вычисляют значения плотности распределения $f(x)$ для различных видов непрерывных распределений. Все имена имеют суффикс *pdf (сокращение от probability density function – плотность распределения).

- binopdf – биномиальное распределение;
- poisspdf – распределение Пуассона;
- normpdf – нормальное распределение;
- unifpdf – непрерывное равномерное распределение;
- chi2pdf – $\chi^2$ - распределение Пирсона;
- tpdf – $t$ – распределение Стьюдента.

Например, для вычисления максимального по модулю отклонения распределения Пуассона от биномиального распределения можно использовать следующий m-файл.

```
function f=lab_1_2(n,p)
f=0;
lambda=n*p;
for i=0:1:n
z=abs(poisspdf(i,lambda)-binopdf(i,n,p));
if f<z f=z;
end
end
```

## 4.3. Квантили распределения.

С помощью этих функций пакета вычисляют критические значения, или квантили, или обратную функцию распределения для различных вероятностных законов. Все имена функций имеют суффикс *inv (сокращение от inverse cumulative function – обратная функция распределения)

- binoinv – биномиальное распределение;
- poissinv – распределение Пуассона;
- norminv – нормальное распределение;
- unifinv – непрерывное равномерное распределение;
- chi2inv – $\chi^2$ - распределение Пирсона;
- tinv – $t$ – распределение Стьюдента.

### 4.4. Генераторы случайных чисел.

С помощью этих функций пакета генерируют случайные числа. Имена этих функций имеют суффикс *rnd (от английского RANDOM- случайный).

- binornd – биномиальное распределение;
- poissrnd – распределение Пуассона;
- normrnd – нормальное распределение;
- unifrnd – равномерное распределение;
- chi2rnd – $\chi^2$ - распределение Пирсона;
- trnd – $t$ – распределение Стьюдента.

## §5. Построение графика функции распределения случайной величины

Для того, чтобы построить график функции распределения $F(x)$ нормально распределённой случайной величины $X$ с параметрами $a, \sigma$ в интервале $[xb, xc]$, например, для $xb = -1,\ xc = 1$ и параметров $a = 0, \sigma = 1$ можно ввести следующую последовательность команд в командном окне MATLAB.

$>> xb = -1$

$>> xc = 1$

$>> a = 0$

$>> \mathrm{sigma} = 1$

$>> x = xb : 0.1 : xc$

Здесь вводятся переменные $xb = -1,\ xc = 1,\ \ a = 0, \sigma = 1$. Командой $x = xb : 0.1 : xc$ вводится массив переменной $x$ в пределах от $[xb, xc]$ с шагом $0.1$.

***Затем командой***

$>> y = \mathrm{normcdf}\ (x, a, \mathrm{sigma}\ )$

вычисляется массив $y = F(x)$. Далее вводится команда построения графика.

$>> \ \mathrm{plot}(x, y)$

Построение графика функции распределения $F(x)$ нормально распределённой случайной величины $X$ с параметрами $a = 0, \sigma = 1$ можно выполнить также с помощью следующего m-файла (m- функции)

```
function f=lab_1_1(xb,xc,a,sigma)
x=xb:0.1:xc;
y=normcdf(x,a,sigma);
plot(x,y);
```

*Замечание*: m-файлом называется текстовый файл с расширением .m, содержащий набор инструкций на языке системы MATLAB.

Графики плотностей распределения непрерывных случайных величин строятся аналогично. Например, график плотности распределения нормально распределенной случайной величины $X$ с параметрами $a, \sigma$ в интервале $[xb, xc]$ строится следующим образом

$>> xb = -1$

$>> xc = 1$

$>> a = 0$

$>> \text{sigma} = 1$

$>> x = xb : 0.1 : xc$

$>> y = \text{normpdf}(x, a, \text{sigma})$

$>> \text{plot}(x, y)$

## Лабораторная работа № 1

## Основные распределения вероятностей, используемые в математической статистике. Выборка и ее характеристики

*Цель работы:* Знакомство с функциями пакета Statistics Toolbox системы MATLAB. Закрепление понятий функции распределения, плотности распределения.

### Задания на лабораторную работу

1. В командном окне пакета MATLAB введите команду help normcdf. Разберитесь с синтаксисом этой команды. Вычислите вероятность события $P(X < 2)$, где $X$ нормально распределённая случайная величина с параметрами

$a = 0, \sigma = 1$. А также вычислите вероятность события $P(-3 \le X < 3)$, где $X$ нормально распределённая случайная величина с параметрами $a = 0, \sigma = 1$.

2. Постройте график функции распределения $F(x)$ случайной величины $X \sim N(a, \sigma)$ в интервале $[xb, xc]$ (параметры и интервал указываются преподавателем отдельно для каждого студента свои).
3. Вычислите вероятность $P(x_1 \le X < x_2)$, для случаев биномиального распределения, распределения Пуассона, равномерного распределения, распределения Пирсона и распределения Стьюдента. Параметры этих распределений выберите самостоятельно. Постройте графики функций распределений перечисленных распределений.
4. Постройте графики плотностей распределений биномиального, пуассоновского, нормального, непрерывного, распределения Пирсона и распределения Стьюдента.
5. Численно убедитесь, что плотность $\chi^2 -$ распределения Пирсона с $n$ степенями свободы задается следующей формулой:

$$k_n(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}} \Gamma\left(\dfrac{n}{2}\right)} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}, & \text{при } x > 0 \\ 0, & \text{при } x \le 0 \end{cases},$$

где $\Gamma\left(\dfrac{n}{2}\right) = \int\limits_0^\infty x^{\frac{n}{2}-1} e^{-x} dx$ - гамма-функция. Известно, что $M[\chi^2] = n, \ D[\chi^2] = 2n$.

***Указание 1.*** Для этого с помощью средств пакета Mat LAB постройте график функции $K_n(x)$. Сравните полученный график с графиком, полученным с помощью функции chi2pdf.

***Указание2.*** Для вычисления гамма- функции в MATLAB имеется команда gamma.

Например:

```
>>  g=gamma(6);
>>  g =120
```

6. Численно убедитесь, что плотность распределения Стьюдента с n степенями свободы имеет вид

$$s_n(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi\cdot n}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{\left(\frac{n+1}{2}\right)}.$$

   Число степеней свободы $n$ для каждого студента свое и равно его номеру в списке.
7. Выясните, при каких значениях произведения $n\cdot p-$ параметров биномиального распределения, для приближённого вычисления вероятности $b_n(k,p)$ следует пользоваться приближением Пуассона, а при каких – приближением Муавра-Лапласа. В качестве оценки точности приближения выберите максимальное по модулю отклонение одного распределения от другого.

8. Найдите медиану, верхнюю и нижнюю квантили для случаев биномиального распределения, пуассоновского распределения, непрерывного равномерного распределения, распределения Пирсона и распределения Стьюдента. Параметры этих распределений выберите самостоятельно. Как изменятся квантили нормального распределения с параметром $a=0$, если параметр $\sigma$ изменить в $k$ раз?

## Контрольные вопросы

1. Какие математические пакеты прикладных программ известны вам? В чём преимущества использования математических пакетов в решении прикладных задач? В чём их недостатки?
2. Перечислите некоторые функции пакета Statistics Toolbox системы MATLAB. В чем заключается смысл суффиксов в названиях функций?
3. Что собой представляет командное окно MATLAB? Что такое m-файл MATLAB?
4. Дайте определения функции и плотности распределения случайной величины.

5. Перечислите параметры биномиального распределения, распределения Пуассона, нормального распределения, равномерного распределения. Какие значения возвращают функции пакета binopdf и poisspdf , когда переменная не является натуральным числом.
6. Запишите плотность нормального распределения случайной величины с параметрами $(a,\sigma)$.
7. Запишите плотность равномерного распределения случайной величины на отрезке $[a,b]$.
8. Запишите плотность показательного распределения случайной величины.
9. Что называется медианой и квантилями распределения случайной величины? Чему равна медиана нормально распределённой случайной величины с параметрами $(a,\sigma)$? Чему равна медиана равномерно распределённой случайной величины с параметрами $(a,b)$?
10. Сформулируйте центральную предельную теорему.

## §6. Графическое представление статистических данных в MATLAB

Пусть задан файл типа .txt, в котором записаны n случайных чисел $x(i),\ i \in 1:n$, распределённых, например, по нормальному закону. В дальнейшем эту выборку $x(i),\ i \in 1:n$ будем называть большой выборкой. Сгенерировать такой файл можно с помощью m-файла MATLAB (см. ПРИЛОЖЕНИЕ 1). Графическое представление статистических данных в MATLAB состоит из следующих шагов:

1. Сделаем выбор $m$ случайных чисел $y(j),\ j \in 1:m$ с помощью датчика случайных чисел из большой выборки. Для этого число $m$ возьмем из условия $m < 30 < n$. Выборка $y(j),\ j \in 1:m$ будем называть малой. Датчики случайных чисел в MATLAB можно реализовать функциями, приведенными в параграфе 1.4, например, функция **y=unidrnd(n,m,1)** формирует одномерный массив объёма m натуральных дискретных равномерных случайных чисел в пределах от 1 до n.

2. Выборки $x(i),\ i \in 1:n$ и $y(j),\ j \in 1:m$ ранжируем, то есть представим по неубыванию. В MATLAB эту процедуру можно выполнить с помощью функции:

   **x=sort(x(:)); y=sort(y(:));**

3. Вычислим выборочное среднее $\overline{X}$ и исправленную выборочную дисперсию $s^2$ большой выборки по следующим формулам

$$\overline{X} = \frac{\sum_{i=1}^{n} x(i)}{n}, \quad s^2 = \frac{\sum_{i=1}^{n} \left(x(i) - \overline{X}\right)^2}{n-1}.$$

   В MATLAB для вычисления выборочной средней $\overline{X}$ используется функция **Mx=mean(x);** а для вычисления исправленной выборочной дисперсии $s^2$ - функция **Dx=var(X),** которая вычисляет несмещённую (исправленную) точечную оценку дисперсии. Аналогично, вычисляются выборочное среднее и исправленная выборочная дисперсия малой выборки.

4. Строим полигоны большой и малой выборки с числом вершин $k = \left\lfloor \sqrt{n} \right\rfloor$. Для нахождения середин интервалов разбиения и числа попаданий в каждый интервал в MATLAB используется функция **[njx,xm]=hist(x,k);**. Результатом функции **[njx,xm]=hist(x,k);** является массив **xm** середин интервалов и массив **njx** числа попаданий в интервал. Полигон малой выборки строится на том же разбиении, что и полигон большой выборки. Для построения полигона используется функция **plot(xm,njx,ym,njy);**, здесь **njy** – массив числа попаданий малой выборки в интервал разбиения.

5. Строим гистограммы большой и малой выборок с $k = \left\lfloor \sqrt{n} \right\rfloor$ числом участков разбиения. Построение гистограммы для выборки, например, $x(i),\ i \in 1:n$ в MATLAB выполняется с помощью следующих функций **[njx,xm]=hist(x,k)**; и **bar(xm,nj);** Функция **bar(xm,nj);** выполняет графический вывод построенной гистограммы. Гистограмма малой выборки строится так, чтобы число интервалов и середины интервалов этих гистограмм (большой и малой выборок) совпадали.

```
>> x=normrnd(a,si,n,1)
>> y=unidrnd(n,m,1)
```

```
>> x=sort(x(:)); y=sort(y(:));
>> Mx=mean(x);
>> Dx=var(X);
>> [njx,xm]=hist(x,k);
>>[njy,ym]=hist(y,k);
>> plot(xm,njx,ym,njy);
>> bar(xm,nj);
```

## §7. Построение доверительных интервалов для математического ожидания и среднего квадратического отклонения нормально распределённой генеральной совокупности в MATLAB

Интервальные оценки математического ожидания и дисперсии большой и малой выборок, полученных в предыдущем параграфе, при значениях доверительных вероятностей $0.9, 0.95, 0.99$ находят из следующего равенства

$$P\left(\overline{X}-t(\gamma)\cdot\sqrt{\frac{s^2}{n}}\le M[X]\le \overline{X}+t(\gamma)\cdot\sqrt{\frac{s^2}{n}}\right)=\gamma .$$

Здесь $\gamma$ - доверительная вероятность, $t(\gamma)$ - значение квантиля $\frac{1+\gamma}{2}$ распределения Стьюденты с $n-1$ степенями свободы, которое в MATLAB вычисляется функцией **tinv**$\left(\frac{1+\gamma}{2},\mathrm{n}-1\right)$**.**
Интервальная оценка дисперсии находится из следующего равенства

$$P\left(\frac{(n-1)s^2}{\chi_1^2}\le D[X]\le\frac{(n-1)s^2}{\chi_2^2}\right)=\gamma ,$$

где $\gamma$ - доверительная вероятность, а величины $\chi_1^2$ и $\chi_2^2$ выбираются так, чтобы

$$\int_0^{\chi_1^2}\chi_{n-1}^2(t)dt=\frac{1-\gamma}{2}\quad \text{и}\quad \int_0^{\chi_2^2}\chi_{n-1}^2(t)dt=\frac{1+\gamma}{2}.$$

Квантили $\chi^2$ –распределения Пирсона с $n-1$ степенями свободы в MATLAB вычисляются функцией **chi2inv($\alpha, n-1$),** где $\alpha$ значение переменной, в которой вычисляется квантиль**.** Так, что $\chi_1^2 = \text{chi2inv}\left(\frac{1-\gamma}{2}, n-1\right)$, $\chi_2^2 = \text{chi2inv}\left(\frac{1+\gamma}{2}, n-1\right)$.

## Лабораторная работа № 2
## Предварительная обработка статистических данных

*Цель работы*: Закрепление понятий генерального и выборочного распределений. Отработка способов графического представления статистических данных, связанных с понятиями: вариационный ряд, полигон, гистограмма. Численное построение доверительных интервалов для выборок различного объёма нормально распределённой генеральной совокупности.

### Задания на лабораторную работу

1. Выберите параметры $(a, \sigma)$ нормального распределения. С помощью команды normrnd получите последовательность $n$ нормально распределенных случайных чисел. Вычислите их среднее и сравните с выбранным параметром $a$. Как будут отличаться среднее и параметр $a$, если увеличить параметр $\sigma$?
2. Сгенерируйте последовательность $X_1, X_2, ..., X_n$ равномерно распределённых на интервале [0,1] независимых случайных величин. Вычислите $Z = \sum_{i=1}^{n} X_i$. Согласно центральной предельной теореме распределение случайной величины $Z$ приближается с ростом $n$ к нормальному распределению $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$, где $a = \frac{n}{2}$, $\sigma = \sqrt{\frac{n}{12}}$. Проверьте это числено. Для чего получите $N$ значений $Z_1, Z_2, ..., Z_N$ и постройте гистограммы при различных $n$. Интервал $(-\infty, +\infty)$ разбейте на $(k+2)$ интервалов: $(-\infty, -3)$; $(3, +\infty)$ и интервал $(-3; 3)$ на $k$ подинтервалов.

3. Студенту предлагается файл типа .txt , в котором записаны n случайных чисел $x(i),\ i \in 1:n$, распределённых по нормальному закону. Нужно сгенерировать такой файл с помощью m-файла MATLAB (см. ПРИЛОЖЕНИЕ 1).
4. С помощью датчика случайных чисел сделать выбор $m$ случайных чисел $y(j),\ j \in 1:m$ из большой выборки. Число $m$ взять из условия $m < 30 < n$.
5. Выборки $x(i),\ i \in 1:n$ и $y(j),\ j \in 1:m$ ранжировать, то есть упорядочить.
   Вычислить выборочное среднее $\overline{X}$ и выборочную исправленную дисперсию $s^2$.
6. Построить полигоны большой и малой выборки с числом вершин $k = \lfloor \sqrt{n} \rfloor$.
7. Построить гистограммы большой и малой выборок с $k = \lfloor \sqrt{n} \rfloor$ числом участков разбиения.
8. Вычислить интервальные оценки математического ожидания и дисперсии большой и малой выборок при значениях доверительных вероятностей $0.9,\ 0.95,\ 0.99$.

***Замечание.*** Для выполнения этой лабораторной работы можно использовать m-файл ПРИЛОЖЕНИЯ 2.

## Контрольные вопросы

1. Что такое выборка и генеральная совокупность?
2. Как сделать случайную повторную выборку?
3. Что значить ранжировать выборку?
4. Как вычисляются выборочное среднее, выборочную дисперсию и выборочное среднее квадратическое отклонение.
5. Что такое интервальный статистический ряд? Как находят число интервалов для интервального статистического ряда?
6. Что такое частота и относительная частота? Как определяют основания и высоты столбиков гистограммы, которая строится с помощью m-файла ПРИЛОЖЕНИЯ 2.
7. Дайте определение интервальной оценки, доверительного интервала и доверительной вероятности. Что такое уровень значимости?
8. Что такое квантиль распределения?

9. Как определяют доверительный интервал для математического ожидания нормально распределённой случайной величины при известной дисперсии и при неизвестной дисперсии? Схематично изобразите график плотности распределения Стьюдента при различном числе степеней свободы?
10. Как определяют доверительный интервал для дисперсии нормально распределённой случайной величины? Схематично изобразите график плотности распределения Пирсона при различном числе степеней свободы?
11. Как соотносятся доверительные интервалы большой и малой выборок при одинаковых значениях доверительных вероятностей?

## §8. Проверка гипотезы о распределении генеральной совокупности с помощью критерия $\chi^2$

Пусть задан файл типа .txt , в котором записаны n случайных чисел $x(k),\ k \in 1:n$, имеющих некоторый закон распределения. Сгенерировать файл такой можно с помощью ПРИЛОЖЕНИЯ 1.

Построим гистограмму с $k = \lfloor \sqrt{n} \rfloor$ числом участков разбиения. Вычислим выборочное среднее $\overline{X}$ и выборочную исправленную дисперсию $s^2$ с помощью следующих встроенных функций: **Mx=mean(x); Dx=var(X).**

На основании метода моментов параметры, входящие в выражения для генеральных плотностей распределения $f(x)$, подбираются так, чтобы моменты генеральных распределений совпадали с выборочными моментами. Рассмотрим выборки основных генеральных распределений.

На основании метода моментов полагаем, что точечные оценки параметров для

- показательного распределения $\alpha = \dfrac{1}{\overline{X}}$
- нормального распределения $m = \overline{X}$ , $\sigma = \sqrt{s^2}$
- равномерного распределения

$$a = \overline{X} - \sqrt{3s^2}\ ,\ \ b = \overline{X} + \sqrt{3s^2}$$

- распределения Релея $\sigma = \sqrt{\frac{2}{\pi}} \cdot \overline{X}$.

Сравним гистограмму с графиками функций плотностей генеральных распределений с выбранными параметрами. Выдвинем гипотезу о виде распределения генеральной совокупности.

Далее вычислим выборочное значение критерия хи-квадрат

$$\chi_B^2 = \sum_{k=1}^{k} \frac{\left(n_i - n \cdot p_i^{теор}\right)^2}{n \cdot p_i^{теор}}$$

(см. ПРИЛОЖЕНИЕ 4). Найдем квантиль $\chi_r^2(1-\alpha)$, где $\alpha-$уровень значимости, $r-$число степеней свободы. Квантиль $\chi_r^2(1-\alpha)$ находится с помощью функции MATLAB **chi2inv(**$1-\alpha, r$**).** Например, chi2inv(.95,4) вычисляет квантиль $\chi_4^2(0.95)$. Сравним значение критерия $\chi_B^2$ и квантиль $\chi_r^2(1-\alpha)$. Если $\chi_B^2 < \chi_r^2(1-\alpha)$, то гипотеза о виде распределения принимается с доверительной вероятностью $1-\alpha$.

## Лабораторная работа № 3

## Статистическая проверка статистических гипотез

*Цель работы*: Закрепление понятия статистической гипотезы. Отработка практических навыков проверки гипотезы о виде распределения генеральной совокупности.

### Задания на лабораторную работу

1. Студенту предлагается файл типа .txt , в котором записаны n случайных чисел $x(k),\ k \in 1:n$, имеющих некоторый закон распределения. Сгенерировать такой файл можно с помощью ПРИЛОЖЕНИЯ 1 .
2. Построить гистограмму с $k = \left[\sqrt{n}\right]$ числом участков разбиения.

   ***Замечание.*** При использовании m-файлов ПРИЛОЖЕНИЙ п. 2-4 являются ознакомительными.

3. Вычислить выборочное среднее $\overline{X}$ и выборочную исправленную дисперсию $s^2$ (в m-файлах MATLAB ПРИЛОЖЕНИЙ для вычисления $\overline{X}$ и $s^2$ используются встроенные функции).
4. На основании метода моментов параметры, входящие в выражения для генеральных плотностей распределения $f(x)$, подбирать так, чтобы моменты генеральных распределений совпадали с выборочными моментами.
5. Гистограмму сравнить с графиками функций плотностей генеральных распределений с параметрами, выбранными в п.4. Выдвинуть гипотезу о виде распределения генеральной совокупности. (см. ПРИЛОЖЕНИЕ 3)
6. Вычислить выборочное значение критерия хи-квадрат (см. ПРИЛОЖЕНИЕ 4).
7. Найти квантиль $\chi_r^2(1-\alpha)$, где $\alpha-$уровень значимости, $r-$число степеней свободы. Квантиль $\chi_r^2(1-\alpha)$ находится с помощью функции MATLAB **chi2inv(**$1-\alpha, r$**).**
8. Сравнить значение критерия $\chi_B^2$ и квантиль $\chi_r^2(1-\alpha)$ и сделать вывод относительно принятия выдвинутой гипотезы.

## Контрольные вопросы

1. Запишите вид функций плотности экспоненциального распределения, распределения Релея, нормального распределения, равномерного распределения.
2. Запишите формулы для вычисления математического ожидания экспоненциального распределения, распределения Релея, нормального распределения, равномерного распределения.
3. Запишите формулы для вычисления дисперсии нормального распределения и равномерного распределения.
4. В чём заключается метод моментов выбора параметров распределений?
5. Как на основании метода моментов выбираются параметры экспоненциального распределения, распределения Релея, нормального распределения, равномерного распределения?
6. Как вычисляется выборочное значение критерия $\chi^2$.

7. Что такое число степеней свободы при вычислении теоретического значения критерия $\chi^2$.

## §9. Вычисление параметров регрессии по методу наименьших квадратов

Пусть заданы три файла типа .txt , в которых имеются возмущённые значения $n$ пар чисел, $(x(i), y(i))$ $i \in 1:n$, теоретически, которые имеют функциональную зависимость. В первых двух файлах предполагаются, что эти зависимости задаются полиномиальными функциями: линейной, которая имеет вид, т.е. $y = a_1 \cdot x + a_2$ и квадратичной, которая имеет вид, т.е. $y = a_1 \cdot x^2 + a_2 \cdot x + a_3$. В третьем файле предполагается, что эта зависимость имеет экспоненциальный вид, т.е. $y = a_1 \cdot e^{a_2}$. Такой файл можно сгенерировать с помощью пакета MATLAB (см. ПРИЛОЖЕНИЕ 5) .

С помощью встроенных функций MATLAB нужно вычислить параметры линий регрессии. Потом выводим график линий регрессии. Вычисление и графический вывод полиномиальных аппроксимирующих кривых можно выполнить с помощью m-файлов ПРИЛОЖЕНИЯ 6.

Вычисление и графический вывод экспоненциальной аппроксимирующей кривой можно выполнить с помощью m-файла ПРИЛОЖЕНИЯ 7.

Нужно построить интервальную таблицу совместного распределения случайных величин $X, Y$. Например,

| X/Y | 1.02 | 1.82 | 2.61 | 3.41 | 4.20 | 5.00 |
|---|---|---|---|---|---|---|
| -1.12 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 12.08 | 0.03 | 0.18 | 0.12 | 0.00 | 0.00 | 0.00 |
| 25.29 | 0.00 | 0.00 | 0.10 | 0.16 | 0.00 | 0.00 |
| 38.49 | 0.00 | 0.00 | 0.00 | 0.05 | 0.14 | 0.00 |
| 54.69 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.08 |
| 64.90 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |

Построить эту таблицу можно с помощью m-файла (ПРИЛОЖЕНИЯ 8). При этом, в верхнюю строку таблицы вносят значения середин интервального ряда распределения случайной величины $X$, а в левый вертикальный столбец значения середин интервального ряда распределения случайной величины $Y$.

Для каждой из этих таблиц вычисляются условные распределения $P(Y|X)$, условные средние $M[Y|X]$. Параметры кривых регрессии $Y$ на $X$ вычисляются по методу наименьших квадратов. Полученные параметры кривых регрессии сравниваются с параметрами кривых регрессии, полученных в п.2

## Лабораторная работа № 4

## Корреляция и регрессия

*Цель работы:* Закрепление понятий совместного распределения, условного распределения, условного среднего, корреляции и регрессии.

### Задания на лабораторную работу

Даны три файла типа .txt , в которых записаны возмущённые значения $n$ пар чисел, $(x(i), y(i))$ $i \in 1:n$, теоретически имеющих функциональную зависимость. В первом и втором файлах предполагаются полиномиальные функции – линейная вида $y = a_1 \cdot x + a_2$ и квадратичная вида $y = a_1 \cdot x^2 + a_2 \cdot x + a_3$. В третьем файле предполагается экспоненциальная зависимость вида $y = a_1 \cdot e^{a_2}$. Сгенерировать такой файл можно с помощью MATLAB (см. ПРИЛОЖЕНИЕ 5) .

1. Необходимо с помощью встроенных функций MATLAB вычислить параметры линий регрессии. Выполнить графический вывод линий регрессии.

   **Указание.** Вычисление и графический вывод полиномиальных аппроксимирующих кривых можно выполнить с помощью m-файлов ПРИЛОЖЕНИЯ 6. Вычисление и графический вывод экспоненциальной аппроксимирующей кривой можно выполнить с помощью m-файла ПРИЛОЖЕНИЯ 7.
2. Построить интервальную таблицу совместного распределения случайных величин $X, Y$.

   **Указание.** Использовать m-файл ПРИЛОЖЕНИЯ 8. В верхней строке таблицы отложены значения середин интервального ряда распределения случайной величины $X$, а в левом вертикальном столбце значения середин интервального ряда распределения случайной величины $Y$.

2. Для каждой из таблиц вычислить условные распределения $P(Y|X)$, условные средние $M[Y|X]$. Вычислить по методу наименьших квадратов параметры кривых регрессии $Y$ по $X$. Полученные параметры кривых регрессии сравнить с параметрами кривых регрессии, полученных в п.2

## Контрольные вопросы

1. Какая может быть зависимость между парой случайных величин?
2. Что такое условное распределение? Как вычисляется условные вероятности в дискретном случае? Как вычисляются условные функции распределения (условные плотности) в непрерывном случае?
3. Как вычисляются условные математические ожидания в дискретном и непрерывном случаях?
4. Что такое кривая регрессии? В чём заключается экстремальное свойство условного среднего?

## Список литературы


1. Гмурман В.Е. «Теория вероятностей и математическая статистика», М., Высшая школа, 2000 г., с. 479.

2. Иглин С.П. «Математические расчёты на базе MATLAB», Санкт-Петербург, БХВ-Петербург, 2005 г., с.636.

3. Калинина В.Н., Панкин В.Ф. «Математическая статистика», М., Высшая школа, 1998 г., с. 336.

4. Бейбалаев В.Д. MATLAB. Лабораторный практикум, ИПЦ ДГУ, 2014.- 67 с.

*ПРИЛОЖЕНИЕ 1*

Генерация выборок с заданными параметрами распределений

```
% формируется одномерный массив размерности n
% выборочного распределения
%---------------------------------------
function f=gener_exp(alpha,n)
x=exprnd(1/alpha,n,1) % экспоненциальное
% x=normrnd(a,si,n,1) % нормальное
% x=raylrnd(sigma,n,1)% релеевское
% x=unifrnd(a,b,n,1) % равномерное
          %распределения
filestud=['d:\Lab_TV_MS\gr351\351_1.txt']
delete(filestud) % удаление предыдущих данных
% из файла
fid=fopen(filestud,'w') % открытие файла
fprintf(fid,'%7.4f %7.4f %7.4f %7.4f ... %7.4f\r\n',x) % запись
данных
fclose(fid) % закрытие файла
```

Графическое представление данных и вычисление доверительных интервалов

```
function f=lab_2

%путь к файлу данных студента
filestud=['d:\Lab_TV_MS\gr351\351_1.txt'];

%-------------------------------------------
% чтение и ранжирование n нормально
% распределённых случайных чисел (большая
% выборка)

x=load(filestud); %чтение данных из файла
x=sort(x(:));     % сортировка в порядке
              % возрастания большой выборки

n=length(x); % число заданных значений
% случайной величины


%-----------------------------------------
% Формирование малой выборки объёма m<30

m=30;

kk=unidrnd(n,m,1); % массив объёма m
          % натуральных случайных чисел
          % равномерно распределённых в
          % пределах от 1 до n

for i=1:m
   y(i)=x(kk(i)); % выбор m случайных чисел из
              % большой выборки
end

y=sort(y(:)); % сортировка в порядке
% возрастания  малой выборки
```

```
%-------------------------------------
% Вычисление выборочных математического
%       ожидания и дисперсии

MX=MEAN(X); % ВЫБОРОЧНОЕ МАТЕМАТИЧЕСКОЕ

        % ОЖИДАНИЕ БОЛЬШОЙ ВЫБОРКИ

DX=VAR(X); % ИСПРАВЛЕННАЯ ДИСПЕРСИЯ БОЛЬШОЙ
       % выборки

MY=MEAN(Y); % ВЫБОРОЧНОЕ МАТЕМАТИЧЕСКОЕ

        % ОЖИДАНИЕ МАЛОЙ ВЫБОРКИ

Dy=var(y); % исправленная дисперсия малой
       % выборки

%------------------------------------------
% вывод результатов

fprintf('объём большой выборки %5i\r\n',n)
fprintf('выборочное математическое ...
  ожидание большой выборки M=%7.4f\r\n',Mx);
fprintf('выборочная дисперсия большой ...
  выборки D=%7.4f\r\n',Dx);
fprintf('объём малой выборки %5i\r\n',m)
fprintf('выборочное математическое ...
  ожидание малой выборки M=%7.4f\r\n',My);
fprintf('выборочная дисперсия малой ...
   выборки D=%7.4f\r\n',Dy);

%      ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ДАННЫХ

k=round(sqrt(n)); % число интервалов разбиения

%--------------------------------------
% Построение гистограммы большой выборки
```

```
xmax = max(x); % максимальный элемент массива
xmin = min(x); % минимальный элемент массива

d=(xmax-xmin)/20; % добавки влево и вправо

xl=xmin-d; xr=xmax+d; % границы интервала
                 % вывода графиков

[nj,xm]=hist(x,k); %  вычисляются nj – число
             %  попаданий в интервал
             %  и xm - середины интервалов

figure; % обьявление новой фигуры

xlim([xl,xr]) ; % границы графика по оси OX

% заголовок графика
set(get(gcf,'CurrentAxes'),'FontName', ...
   'Times New Roman Cyr','FontSize',10);
title('\bf Гистограмма большой выборки')


xlabel('\it x_i') % метка по оси OX
ylabel('\it n_i') % метка по оси OY

bar(xm,nj);  % построение гистограммы

% ----------------------------------------
% Построение гистограммы малой выборки

figure; % обьявление новой фигуры

xlim([xl,xr]) ; % границы графика по оси OX

% заголовок графика
set(get(gcf,'CurrentAxes'),'FontName', ...
  'Times New Roman Cyr','FontSize',10);
title('\bf Гистограмма малой выборки')

xlabel('\it y_j') % метка по оси OX
ylabel('\it m_j') % метка по оси OY
```

```
del=xm(2)-xm(1); % длина интервала

xy(1) = xmin; % левая граница интервалов
           % разбиения
for i=1:k
   xy(i+1)=xy(i)+del; % границы интервалов
                  % разбиения
end
for i=1:k
   ny(i)=0;
   for j=1:m
    if (y(j)<=xy(i+1))&(y(j)>xy(i)) %
         ny(i)=ny(i)+1;  % число точек малой
         % выборки, попавших в интервал
      end
   end
end
ny(1)=ny(1)+1; % добавление одной точки в
            % первый интервал

bar(xm,ny);  % вывод гистограммы

figure; % обьявление новой фигуры

plot(xm,nj,xm,ny); % построение полигонов

%    ВЫЧИСЛЕНИЕ ИНТЕРВАЛЬНЫХ ОЦЕНОК

gamma(1)=.9;     % доверительные
gamma(2)=.95;    %
gamma(3)=.99 ;   % вероятности

%--------------------------------------------
% Интервальные оценки математического ожидания

for i=1:3
   gg=(1+gamma(i))/2; % переменная, в которой
                % вычисляется квантиль
                % распределения Стьюдента

   tgl=tinv(gg,n-1)   % вычисления квантили
```

```
                % для большой выборки

    LML(i)=Mx-tgl*sqrt(Dx/n);    % левая
            % граница доверительного
          % интервала большой выборки

    RML(i)=Mx+tgl*sqrt(Dx/n); % правая
         % граница доверительного
       % интервала большой выборки

    tgs=tinv(gg,m-1)  % вычисление квантили
                 % для малой выборки

    LMS(i)=My-tgs*sqrt(Dy/m); % левая
          % граница доверительного
         % интервала малой выборки
    RMS(i)=My+tgs*sqrt(Dy/m); % правая
           % граница доверительного
          % интервала малой выборки
end

% _______ вывод результатов _______________

fprintf('Интервальные оценки ...
  математического ожидания\r\n');
fprintf('выборочное математическое ...
  ожидание большой выборки M=%7.4f\r\n',Mx);
for i=1:3
  fprintf('вероятность %3.2f интервал ...
   (%7.4f %7.4f) интервал ...
   (%7.4f %7.4f)\r\n', ...
   gamma(i),LML(i),RML(i),LMS(i),RMS(i));
end


% -----------------------------------------
%     Интервальные оценки дисперсии

for i=1:3
```

```
ggl=(1+gamma(i))/2; % переменная, в которой
          % вычисляется квантиль левой
          % границы интервала

 ggr=(1-gamma(i))/2; % переменная, в которой
          % вычисляется квантиль правой
          % границы интервала


 tgl=chi2inv(ggl,n-1) % вычисления квантили
           % распределения Стьюдента
           % для левой границы
           % интервала большой выборки

 tgr=chi2inv(ggr,n-1) % вычисления квантили
           % распределения Стьюдента
           % для правой границы
           % интервала большой выборки

 LDL(i)=(n-1)*Dx/tgl;   % левая граница
         % доверительного интервала
         % большой выборки
 RDL(i)=(n-1)*Dx/tgr;  % правая граница
         % доверительного интервала
         % большой выборки

 tgl=chi2inv(ggl,m-1) % вычисления квантили
            % распределения Стьюдента
            % для левой границы
            % интервала малой выборки

 tgr=chi2inv(ggr,m-1) % вычисления квантили
            % распределения Стьюдента
            % для правой границы
            % интервала большой выборки

 LDS(i)=(m-1)*Dy/tgl;   % левая граница
         % доверительного интервала
         % большой выборки
```

```
   RDS(i)=(m-1)*Dy/tgr;  % правая граница
             % доверительного интервала
             % большой выборки

end

% _______ вывод результатов _______________

fprintf('Интервальные оценки дисперсии\r\n');
fprintf('выборочное математическое ...
   ожидание большой выборки M=%7.4f\r\n',Dx);
for i=1:3
   fprintf('вероятность %3.2f интервал ...
   (%7.4f %7.4f) интервал ...
   (%7.4f %7.4f)\r\n', ...
    gamma(i),LDL(i),RDL(i),LDS(i),RDS(i));
end
```

Построение графиков плотностей выборочного и теоретического распределений

```
function f=lab_3

%  путь к файлу данных студента
filestud=['d:\Lab_TV_MS\gr351\351_1.txt'];


X=load(filestud); %чтение данных из файла

%---------------------------------------
% сортировка массива данных и вычисление
% выборочного математического ожидания и
% дисперсии

X=sort(X(:)); % сортировка в порядке
           % возрастания

M=mean(X); % выборочное математическое ожидание

D=var(X); % исправленная дисперсия

n=length(X); % число заданных значений
          % случайной величины

%_______вывод результатов______________

fprintf('число выборочных значений ...
  случайной величины %5i\r\n',n)
fprintf('выборочное математическое ...
  ожидание M=%7.4f\r\n',M);
fprintf('выборочная дисперсия D=%7.4f\r\n',D);


%  ФОРМИРУЕМ ГРАФИК ПЛОТНОСТИ ВЫБОРОЧНОГО
%  РАСПРЕДЕЛЕНИЯ
```

```
% фиксируем число интервалов графика плотности
% выборочного распределения
k=round(sqrt(n));

% Замечание: число интервалов - фиксируется как
% целая часть корня квадратного объёма выборки,
% но может быть выбрано другим

%  вычисляются nj - число попаданий в интервал
%  и xm - середины интервалов
[nj,xm]=hist(X,k);

% очистили массивы для выборочной функции
% распределения
clear xfv fv ;

% -----------------------------------------
% формируем массив абсцис графика выборочной
% функции распределения

 dd=xm(2)-xm(1);     % длина интервала

 xfv = [xm-dd/2 ; xm+dd/2] ;% двумерный массив.
   %  Первая компонента - левая граница
   % разбиения. Вторая компонента - правая
   % граница разбиения.

xl= xfv(1,1);     % справа и
xr= xfv(end,end); % слева
xll = xl-dd;      % добавляем
xrr = xr+dd;      % по одному интервалу

xfv=[xll;xl;reshape(xfv,prod(size(xfv)),1)...
;xr;xrr];

% -----------------------------------------
% формируем массив ординат графика выборочной
% функции распределения
```

```
fv = nj / (n*dd); % Значение выборочной функции
               % распределения на интервале
% значения выборочной функции распределения
% в точках xfv

fv=[fv;fv];
fv=[0;0;reshape(fv,prod(size(fv)),1);0;0];

% ФОРМИРУЕМ ГРАФИКИ ПЛОТНОСТЕЙ
ТЕОРЕТИЧЕСКИХ
% РАСПРЕДЕЛЕНИЙ

clear xft ft ;    % очистили массивы

%   параметры нормального распределения
a_norm=M;
sigma_norm=sqrt(D);


%   парметр показательного распределения
alpha=1/M

%   параметры равномерного распределения
a_ravnom=M-sqrt(3*D);
b_ravnom=M+sqrt(3*D);


%   параметр распределения Релея
sigma_relea=M*sqrt(2/pi);


% массив абсцисс графиков теоретических
% плотностей распределений
xft= linspace(xll,xrr,1000);

% объём массива равен 1000 и может быть изменён
% интервал изменения абсциссы [xll,xrr]

% массив ординат графика нормального
% распределения
```

```
ftnorm=normpdf(xft,a_norm,sigma_norm);

% массив ординат графика равномерного
% распределения
 ftravnom=unifpdf(xft,a_ravnom,b_ravnom);

% массив ординат графика  распределения Релея
ftrelea=raylpdf(xft,sigma_relea);

% массив ординат графика экспоненциального
% распределения
ftexp=exppdf(xft,alpha);

% ----------------------------------------
% вывод графиков
plot(xfv,fv,xft,ftnorm,xft,ftravnom,xft,ftrelea,xft,ftexp)

% Замечание: если среди графиков теоретических
% распределений есть "дикий" - сильно
% отличающийся от выборочного, то его не
% следует выводить
```

*ПРИЛОЖЕНИЕ 4*

Вычисление статистики хи-квадрат

```
function f=lab_3_stat_chi

% путь к файлу данных студента
filestud=['d:\Lab_TV_MS\gr351\351_1.txt'];



X=load(filestud); %чтение данных из файла

% -------------------------------------
% сортировка массива данных и вычисление
% выборочного математического ожидания и
% дисперсии

X=sort(X(:)); % сортировка в порядке
          % возрастания

M=mean(X); % выборочное математическое ожидание

D=var(X); % исправленная дисперсия

n=length(X); % число заданных значений
         % случайной величины


% ----- вывод результатов ------

fprintf('число выборочных значений ...
     случайной величины %5i\n',n)
fprintf('выборочное математическое …
     ожидание M=%7.4f\n',M);
fprintf('выборочная дисперсия D=%7.4f\r\n',D);


%  ФОРМИРУЕМ ИНТЕРВАЛЫ РАЗБИЕНИЯ
```

```
% Номера интервалов, левые и правые границы
% запишем в виде таблицы Tabl[k,:], где k –
% число интервалов разбиения

% фиксируем число интервалов графика плотности
% выборочного распределения
k=round(sqrt(n));

% Замечание. Число интервалов - фиксируется как
% целая часть корня квадратного объёма выборки,
% но может быть выбрано другим

Tabl(:,1) = [1:k]'; % номера интервалов


%  вычисляются nj - число попаданий в интервал
%  и xm - середины интервалов
[nj,xm]=hist(X,k);

% ----------------------------------------
% формируем массив левых и правых границ
% интервалов

 dd=(xm(2)-xm(1))/2; % длина интервала

 Tabl(:,2) = xm'-dd; % левые границы интервалов
 Tabl(:,3) = xm'+dd ;% правые границы
                 % интервалов
 Tabl(1,2) = -inf;  % теоретическое начало
                % первого интервала
 Tabl(k,3) = inf;   % теоретический конец
                % последнего интервала


% формируется массив опытного числа попаданий
 Tabl(:,4) = nj; % опытное число попаданий

% ------------------------------------
% Формируется массив теоретического числа
% попаданий нормально распределённой случайной
```

```
% величины с параметрами M , si

 si=sqrt(D);
 for i = 1:k
    Tabl(i,5) = (normcdf(Tabl(i,3),M,si)- ...
      normcdf(Tabl(i,2),M,si))*n;
 end

 % Замечание. В случае другого вида
 % распределения параметры и функции
 % распределения берутся как в m-файле
 % Приложения 3.

% -----------------------------------------
% вычисление статистики хи-квадрат

 hi=0;
 for i = 1:k
     hi=hi+(nj(i)-Tabl(i,5))^2/Tabl(i,5);
 end


% ______ Вывод результатов ______________

fprintf(' j    aj      bj     nj   npj  \n')
fprintf('%2i %7.4f  %7.4f  %2i  %7.4f ...
   \n',Tabl');
fprintf('статистика хи-квадрат %7.4f\n',hi);
fprintf('число степеней свободы %2i\n',k-2);
```

*ПРИЛОЖЕНИЕ 5*

Генерация возмущённых значений функциональных зависимостей

```
% массив a(:) – параметры
% функциональной зависимости y=y(x)

% xn,xk - интервал изменения переменной x

% n - число значений переменной x

% Описание программы

% случайным образом выбирается i-й элемент
% массива x из интервала [xn,xk] и вычисляется
% возмущённое значение y.

% обьявление функции
function f=gener_lab4_line(a,xn,xk,n)

% путь к файлу данных
filestud=['d:\Lab_TV_MS\gr351\351_1.txt'];

delete(filestud); % очистка файла данных

fid=fopen(filestud,'w'); % открытие файла

% выбор значений переменной x случайным образом % из
интервала [xn, xk]
x=unifrnd(xn,xk,1,n)

% среднее квадратическое отклонение возмущения
sigma=5 ;

%
% ВЫЧИСЛЕНИЕ ВОЗМУЩЁННЫХ ЗНАЧЕНИЙ

% вычисление возмущённых значений y при
```

```
% полиномиальной зависимости

y=polyval(a,x)+normrnd(0,sigma,size(x));

%вычисление возмущённых значений y при
% экспоненциальной зависимости
y=a(1)*exp(a(2)*x))+normrnd(0,sigma,size(x));

% запись массивов x,y в файл студента.
fprintf(fid,'%4.3f %4.3f\r\n',[x;y]);

fclose(fid); % закрытие файла
```

Построение и графический вывод полиномиальной кривой регрессии

```
%  n - степень аппроксимирующего полинома
%  массив a(:) - коэффициенты полиномиальной
%  функции
%  y(x)=a(1)*x^n+a(2)*x^(n-1)+...+a(n)
%
%
function f=lab4_1(n) % объявление функции

% путь к файлу данных
filestud=['d:\Lab_TV_MS\gr351\351_1.txt'];

fid=fopen(filestud); % открытие файла

z = fscanf(fid, '%g'); % чтение

z = reshape(z, 2, []); % разбивка на два
                  % массива

x=z(1,:);
y=z(2,:);

fclose(fid); % закрытие файла

% построение ломаной регрессии
a=polyfit(x,y,n);


x1=sort(x); % сортировка массива данных

% построение полиномиальной кривой
func_line=polyval(a,x1);

% графический вывод
```

```
plot(x,y,'.k',x1,func_line,'-');

% печать
fprintf('Аппроксимирующий полином ...
%d-й степени:\ny(x)=',n);
fprintf('%+f6*x^%d',[a;[n:-1:0]]);
disp(' ');
```

*ПРИЛОЖЕНИЕ* 7

Построение и графический вывод экспоненциальной кривой регрессии

```
%  массив a(:) - коэффициенты
%  экспоненциальной функции
%  y(x)=a(1)*exp(a(2)*x)


function f=lab4_2 % объявление функции

% путь к файлу данных
filestud=['d:\Lab_TV_MS\gr351\351_1.txt'];

fid=fopen(filestud); % открытие файла

z = fscanf(fid, '%g'); % чтение

z = reshape(z, 2, []); % разбивка на два
                  % массива
x=z(1,:);
y=z(2,:);

fclose(fid); % закрытие файла

%  ---- построение ломаной регрессии ----
a1=nlinfit(x,y,'lab4_exp',[1 1]);

x1=sort(x);% сортировка массива данных

n=size(x1);
for i=1:n
   y1=a1(1)*exp(a1(2)*x1);
end
% -------------------------------------

% графический вывод
plot(x,y,'.k',x1,y1,'-');

% печать
fprintf('Аппроксимирующая ...
```

```
экспоненциальная функция ...
y(x)=%f6*exp(%f6*x)\n\r',a1(1),a1(2));

% вычисление экспоненцальной функции
function y=lab4_exp(a,x)
y=a(1)*exp(a(2)*x);
```

Построение таблицы совместного распределения

```
function f=lab4_3 % объявлении функции

% путь к файлу данных
filestud=['d:\Lab_TV_MS\gr351\351_1.txt'];

fid=fopen(filestud); % открытие файла

% чтение массива измерений (x,y)
z = fscanf(fid, '%g'); fclose(fid); % закрытие файла данных

% Формирование массивов переменных
z = reshape(z, 2, []);
x=z(1,:);
y=z(2,:);

% Разбиение данных на 8 интервалов
% ------------------------------------

n=length(x); % длина массива данных
xs=sort(x); ys=sort(y); % сортировка данных

% длина интервала
dx=(xs(n)-xs(1))/7; dy=(ys(n)-ys(1))/7;

for i=1:8
   xm(i)=(i-1)*dx+xs(1); % середины интервалов
   ym(i)=(i-1)*dy+ys(1);
end

for i=1:8
   xb(i)=xm(i)-dx/2; % границы интервалов
   yb(i)=ym(i)-dy/2;
end

xb(9)=xm(8)+dx; yb(9)=ym(8)+dx;
%-----------------------------------
```

```
% Построение таблицы совместного распределения
%-------------------------------------

% вероятности попадания в интервал
for i=1:8
 for j=1:8
    p(i,j)=0;
    for k=1:n
       if xs(k)>xb(i))&(xs(k)<=xb(i+1))...
          &(ys(k)>yb(j))&(ys(k)<=yb(j+1))
          p(i,j)=p(i,j)+1/n;
       end
    end
  end
end
% Печать таблицы совместного распределения
fprintf(' X/Y  %5.2f %5.2f %5.2f %5.2f ...
 %5.2f %5.2f %5.2f %5.2f\r\n',xm)
for j=1:8
    fprintf('%5.2f %5.2f %5.2f %5.2f ...
    %5.2f %5.2f %5.2f %5.2f ...
    %5.2f\r\n',ym(j),p(:,j))
end
```