**Министерство науки и высшего образования Российской Федерации**
**Федеральное государственное бюджетное образовательное учреждение**
**высшего образования**
**«Московский государственный технический университет**
**имени Н.Э. Баумана**
**(национальный исследовательский университет)»**
**(МГТУ им. Н.Э. Баумана)**

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

*Лабораторная работа №1*

*По предмету: «Математическая статистика»*

# Тема: Гистограмма и эмпирическая функция распределения
# Вариант №2

Преподаватель: Саркисян П.С.
Студент: Гасанзаде М.А.,
Группа: ИУ7-66Б

Москва, 2020 г.

# ОГЛАВЛЕНИЕ

## ВВЕДЕНИЕ

**Цель работы:** построение гистограммы и эмпирической функции распределения.

**Содержание работы:**

1. Для выборки объёма *n* из генеральной совокупности *M* реализовать в виде программы на ЭВМ

• вычисление максимального значения $M_{\max}$ и минимального значения $M_{\min}$;

• размаха *R* выборки;

• вычисление оценок $\hat{\mu}$ и $S^2$ математического ожидания M*X* и дисперсии D*X*;

• группировку значений выборки в $m = [\log_2 n\ ] + 2$ интервала;

• построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией $S^2$;

• построение на другой координатной плоскости графика эмпирической

функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией $S^2$.

2. Провести вычисления и построить графики для выборки из индивидуального варианта.

# 1. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

В данной части рассмотрены формулы для вычисления величин, эмпирическая плотность, эмпирическая функция распределения и гистограмма.

## 1.1 Формулы для вычисления величин

**Минимальное значение выборки**

$$M_{min}=\min\{x_1,\dots,x_n\}, \qquad \textbf{(1)}$$

где $(x_1,\dots,x_n)$ – реализация случайной выборки.

**Максимальное значение выборки**

$$M_{max}=\max\{x_1,\dots,x_n\}, \qquad \textbf{(2)}$$

где $(x_1,\dots,x_n)$ – реализация случайной выборки.

**Размах выборки**

$$R=M_{max}-M_{min}, \qquad \textbf{(3)}$$

где $M_{\max}$ – максимальное значение выборки,

$M_{\min}$ – минимальное значение выборки.

**Оценка математического ожидания**

$$\hat{\mu}(\vec{X}_n)=\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad \textbf{(4)}$$

**Несмещённая оценка дисперсии**

$$S^2(\vec{X}_n)=\frac{n}{n-1}\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \qquad \textbf{(5)}$$

**Выборочная дисперсия**

$$\hat{\sigma}^2(\vec{X}_n)=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \qquad \textbf{(6)}$$

**Количество интервалов**

$$m=[\log_2 n]+2 \qquad \textbf{(7)}$$

## 1.2 Эмпирическая плотность и гистограмма

*Определение*. Эмпирической плотностью распределения случайной выборки $\vec{X}_n$ называют функцию:

$$f_n(x)=\begin{cases} \dfrac{n_i}{n\Delta}, & x\in J_i, i=\overline{1,m}; \\ 0, & иначе. \end{cases} , \qquad \textbf{(8)}$$

где $J_i$, $i=\overline{1;m}$ – полуинтервал из $J=[x_{(1)}, x_{(n)}]$ ,

$$\begin{aligned} & x_{(1)}=\min\{x_1,\dots,x_n\}, \\ & x_{(n)}=max\{x_1,\dots,x_n\}, \\ & J_i=\left[x_{(1)}+(i-1)\Delta\right), x_{(i)}+i\Delta, \quad i=\overline{1,m-1}, \\ & J_m=[x_{(1)}+(m-1)\Delta, x_{(1)}+m\Delta], \end{aligned}$$

$m$ – количество полуинтервалов интервала $J=[x_{(1)}, x_{(n)}]$ ,

$\Delta$ – длина полуинтервала $J_i, i=\overline{1,m}$ равная

$$\Delta=\frac{x_{(n)}-x_{(1)}}{m}=\frac{|J|}{m}, \qquad \textbf{(9)}$$

$n_i$ – количество элементов выборки в полуинтервале $J_i$, $i=\overline{1,m}$ ,

$n$ — количество элементов в выборке.

*Определение*. График функции $f_n(x)$ называют гистограммой.

## 1.3 Эмпирическая функция распределения

*Определение*. Пусть

1. $\vec{X}_n=(X_1,\dots,X_n)$ – случайная выборка,
2. $\vec{x}_n=(x_1,\dots,x_n)$ – реализация случайной выборки,
3. $n(x,\vec{x}_n)$ – количество элементов выборки $\vec{x}_n$ , которые меньше $x$, тогда эмпирической функцией распределения называют функцию

$$\mathrm{F}_n:\Re\to\Re,\quad \mathrm{F}_n=\frac{n(x,\vec{x}_n)}{n} \qquad \textbf{(10)}$$

*Замечание.*

1. $Fn$(x) обладает всеми свойствами функции распределения;
2. $Fn$(x) кусочно-постоянна;
3. если все элементы вектора различны, то

$$F_n(x)=\begin{cases}0, & x\le x_{(1)};\\ \frac{i}{n}, & x_{(i)}<x\le x_{(i+1)}, i=\overline{1,n-1}\\ 1, x>x_{(n)}.\end{cases} \qquad \textbf{(11)}$$

4. Эмпирическая функция распределения позволяет интерпретировать выборку $\vec{x}_n$ как реализацию дискретной случайной величины $\widetilde{X}$, ряд распределения которой имеет вид:

| $\widetilde{\boldsymbol{X}}$ | $x_{(1)}$ | ... | $x_{(n)}$ |
|---|---|---|---|
| **P** | $1/n$ | ... | $1/n$ |

Это позволяет рассматривать числовые характеристики случайной величины $\widetilde{X}$ как приближенные значения числовых характеристик случайной величины $X$.

# 2. ПРАКТИЧЕСКАЯ ЧАСТЬ

В данной части представлен листинг и результат работы программы.

## 2.1Листинг программы

```
function lab1()
    clear all;
    X = [0.70,-0.35,-0.23,-1.18,-0.75,0.41,-0.71,0.97,-2.54,-1.50,1.73,-0.83,-
1.13,-0.00,-0.40,1.23,-0.14,-1.38,0.01,-1.65,0.02,-1.61,0.46,1.19,-
1.30,0.32,1.19,-0.03,-0.31,-1.64,-0.24,0.30,-0.66,-1.31,-0.65,0.63,-
0.27,1.04,0.20,0.31,0.24,1.27,-0.17,-0.62,0.03,-1.75,-2.26,-0.03,-0.27,-
0.17,0.10,-0.14,0.09,0.53,-0.78,-0.86,0.35,-0.72,-0.41,0.38,-0.91,-0.41,-
1.10,-1.00,0.39,-0.06,0.32,-1.58,-0.14,-0.90,-1.84,0.00,-0.10,-1.14,-
0.14,0.82,-2.55,-2.79,-0.02,-0.66,-0.05,-0.15,-1.68,1.62,0.21,-0.01,-
0.33,0.68,1.80,-0.29,-0.74,-0.38,-2.67,-1.53,-0.48,0.66,-
0.56,0.28,0.70,1.01,0.53,0.93,-1.27,-1.37,-0.29,-2.18,-1.02,0.21,0.19,1.75,-
0.01,0.30,-0.73,0.34,-0.23,1.13,-1.13,-0.96,0.37,0.14];
    X = sort(X);

    Mmax = max(X);
    Mmin = min(X);

    fprintf('Mmin = %s\n', num2str(Mmin));
    fprintf('Mmax = %s\n', num2str(Mmax));

    R = Mmax - Mmin;
    fprintf('R = %s\n', num2str(R));

    MU = getMU(X);
    fprintf('MU = %s\n', num2str(MU));

    Ssqr = getSsqr(X);
    fprintf('S^2 = %s\n', num2str(Ssqr));

    m = getNumberOfIntervals(X);
    fprintf('m = %s\n', num2str(m))

    createGroup(X);
    hold on;
    distributionDensity(X, MU, Ssqr, m);

    figure;
    empiricF(X);
    hold on;
    distribution(X, MU, Ssqr, m);
end

function mu = getMU(X)
    n = length(X);
    mu = sum(X)/n;
end

function Ssqr = getSsqr(X)
    n = length(X);
    MX = getMU(X);
    Ssqr = sum((X - MX).^2) / (n-1);
end
```

```
function m = getNumberOfIntervals(X)
    m = floor(log2(length(X)) + 2);
end

function createGroup(X)
    n = length(X);
    m = getNumberOfIntervals(X);

    intervals = zeros(1, m+1);
    numCount = zeros(1, m+1);
    Delta = (max(X) - min(X)) / m;

    for i = 0: m
        intervals(i+1) = X(1) + Delta * i;
    end

    j = 1;
    count = 0;
    for i = 1:n
        if (X(i) >= intervals(j+1))
            j = j + 1;
        end
        numCount(j) = numCount(j) + 1;
        count = count + 1;
    end

      graphBuf = numCount(1:m+1);
    for i = 1:m+1
        graphBuf(i) = numCount(i) / (n*Delta);
    end

    stairs(intervals, graphBuf),grid;
end

function distributionDensity(X, MX, DX, m)
    R = X(end) - X(1);
    delta = R/m;
    Sigma = sqrt(DX);

    Xn = (MX - R): delta/50 :(MX + R);
    Y = normpdf(Xn, MX, Sigma);
    plot(Xn, Y), grid;
end

function distribution(X, MX, DX, m)
    R = X(end) - X(1);
    delta = R/m;

    Xn = (MX - R): delta :(MX + R);
    Y = 1/2 * (1 + erf((Xn - MX) / sqrt(2*DX)));
    plot(Xn, Y, 'r'), grid;
end

function empiricF(X)
    [yy, xx] = ecdf(X);

    stairs(xx, yy), grid;
end
```

## 2.2 Результат работы программы

$M_{min} = -2.79$;
$M_{max} = 1.8$;
$R = 4.59$;
$\hat{\mu}(\vec{X}_n) = -0.28592$;
$S^2(\vec{X}_n) = 0.91702$.

Интервальная группировка значений выборки при m = 8:

| | | | |
|---|---|---|---|
| [-2.79; -2.22) | [-2.22; -1.64) | [-1.64; -1.07) | [-1.07; -0.50) |
| 5 | 5 | 15 | 18 |
| [-0.50; 0.08) | [ 0.08; 0.65) | [ 0.65; 1.23) | [ 1.23, 1.80] |
| 35 | 24 | 12 | 5 |

## 2.3 Графики

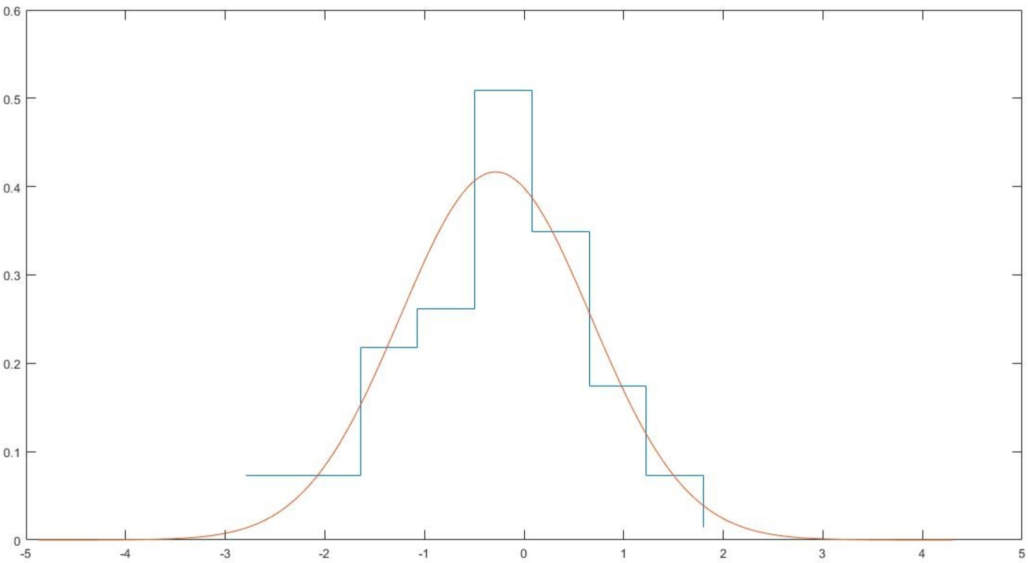


Рисунок 1. Гистограмма и график функции плотности распределения нормальной случайной величины.

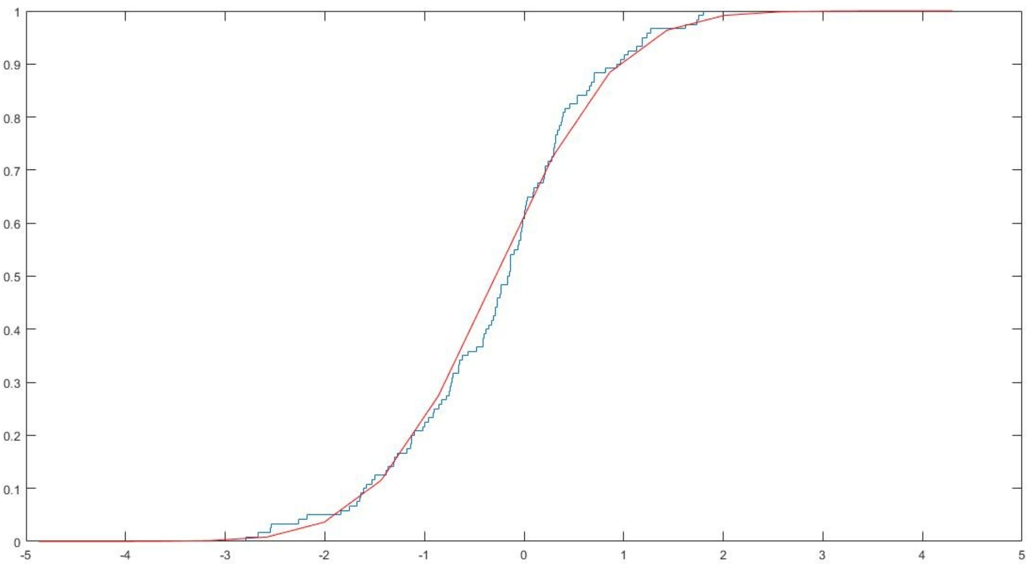


Рисунок 2. График эмпирической функции распределения и функции распределения нормальной случайной величины .