|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **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

**Отчет**

**по лабораторной работе № 1**

**По курсу «Математическая статистика»**

**Тема: «Гистограмма и эмпирическая функция распределения»**

Выполнила: Овчинникова А.П.

Группа: ИУ7-65Б

Вариант: 15

Преподаватель: Волков И. К.

Москва, 2020

ОГЛАВЛЕНИЕ

[1. Постановка задачи 3](#_Toc35558308)

[2. Теоретическая часть 3](#_Toc35558309)

[2. 1. Необходимые формулы 3](#_Toc35558310)

[2. 2. Необходимые определения 4](#_Toc35558311)

[3. Практическая часть 6](#_Toc35558312)

[3. 1. Листинг программы 6](#_Toc35558313)

[3. 2. Полученные результаты 8](#_Toc35558314)

1. **Постановка задачи**

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

**Содержание работы**

1. Для выборки объема n из генеральной совокупности X реализовать в виде программы на ЭВМ:
   1. вычисление максимального значения Mmax и минимального значения Mmin;
   2. размаха R выборки;
   3. вычисление оценок и S2 математического ожидания MX и дисперсии DX;
   4. группировку значений выборки в m = [log2n] + 2 интервала;
   5. построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием и дисперсией S2;
   6. построение на другой координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием и дисперсией S2.
2. Провести вычисления и построить графики для выборки из индивидуального варианта.
3. **Теоретическая часть**

**2. 1. Необходимые формулы**

Пусть – реализация случайной выборки.

Формула для вычисления максимального значения выборки Mmax:

.

Формула для вычисления минимального значения выборки Mmin:

.

Формула для вычисления размаха выборки:

, где

– максимальное значение выборки;

– минимальное значение выборки.

Выборочным средним (выборочным математическим ожиданием) называется статистика

*.*

Выборочной дисперсией называется статистика

*.*

Формула для вычисления несмещенной оценки дисперсии:

*.*

**2. 2. Необходимые определения**

Предположим, что для выборки построен интервальный статистический ряд. Отрезок делят на p равновеликих частей:

*, i =*

, где

.

**Определение 2.1.** Эмпирической плотностью распределения случайной выборки (отвечающей выборке ) называют функцию

, где

— количество элементов выборки , которые принадлежат , ;

— количество элементов в выборке.

**Определение 2.2.** Гистограммой называют график эмпирической плотности.

Пусть – выборка из генеральной совокупности X; – реализация случайной выборки . Обозначим через число элементов вектора , которые имеют значение, меньшее x.

**Определение 2.3.** Эмпирической функцией распределения называют функцию , определенную условием .

**Замечание.**  обладает всеми свойствами функции распределения.

**Замечание.**  кусочно-постоянна и скачкообразно изменяет свои значения в точке

**Замечание.** Если элементы выборки попарно различны, то

*.*

**Замечание.** Эмпирическая функция распределения позволяет интерпретировать выборку как реализацию дискретной случайной величины , ряд распределения которой имеет вид:

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Z(1)* | *…* | *Z(m)* |
| *p* | *n1/n* | *…* | *nm/n* |

В дальнейшем это позволит рассматривать числовые характеристики случайной величины как приближенные значения числовых характеристик случайной величины X.

1. **Практическая часть**

**3. 1. Листинг программы**

Код программы представлен в листинге 1.

Листинг 1. Программа лабораторной работы 1.

function lab1\_ms()

X = readFromFile('C:\Users\novoc\Desktop\6sem\data.csv');

X = sort(X); % Вариационный ряд

Xmin = X(1);

Xmax = X(end);

R = Xmax - Xmin;

n = length(X);

mu = sum(X) / n;

ssqr = getCorrectedSampleVariance(X, mu);

p = countSubintervals(n);

fprintf('Минимальное значение выборки: %.6f\n', Xmin);

fprintf('Максимальное значение выборки: %.6f\n', Xmax);

fprintf('Размах выборки: %.6f\n', R);

fprintf('Размер выборки: %d\n', n);

fprintf('Выборочное математическое ожидание: %.6f\n', mu);

fprintf('Исправленная выборочная дисперсия: %.6f\n', ssqr);

fprintf('Всего интервалов: %d\n', p);

[J, count] = group(X, p);

drawHist(X, J, count);

hold on; % hold on сохраняет графики в текущей системе координат так,

% чтобы новые графики, добавленные к осям, не удаляли существующие графики

f(X, mu, ssqr, p);

figure; % figure создает новое окно рисунка

drawDist(X);

hold on;

F(X, mu, ssqr, p);

end

function F(X, MX, DX, p)

R = X(end) - X(1);

delta = R / p;

Xn = (MX - R) : delta / 20 : (MX + R);

Y = 0.5 \* (1 + erf((Xn - MX) / sqrt(2 \* DX)));

plot(Xn, Y, 'r');

end

function drawDist(sample)

[f, x] = ecdf(sample); % возвращает эмпирическую интегральную функцию распределения (cdf) f,

% вычисленную в точках x, используя данные в векторе y.

stairs(x, f), grid;

end

function f (X, MX, DX, m)

R = X(end) - X(1);

delta = R / m;

sigma = sqrt(DX);

Xn = (MX - R) : delta / 20 : (MX + R);

Y = normpdf(Xn, MX, sigma);

plot(Xn, Y);

end

function drawHist(sample, J, count)

p = length(count);

n = length(sample);

delta = (sample(n) - sample(1)) / p;

xes = zeros(1, p + 1);

xes(1) = 0;

for i = 2 : p

xes(i) = count(i) / (n \* delta);

end

stairs(J, xes), grid;

end

function [J, count] = group(sample, p)

delta = (sample(end) - sample(1)) / p;

count = zeros(1, p);

J = sample(1):delta:sample(end);

fprintf('Интервалы:\n');

for i = 1 : p - 1

fprintf('[%.6f; %.6f), ', J(i), J(i+1));

end

fprintf('[%.6f; %.6f]\n ', J(end - 1), J(end));

for i = 1 : length(sample)

cur = sample(i);

for j = 1 : p - 1

if ((cur >= J(j)) && (cur < J(j + 1)))

count(j) = count(j) + 1;

end

end

if ((cur >= J(end - 1)) && (cur <= J(end)))

count(end) = count(end) + 1;

end

end

end

function p = countSubintervals(sample\_size)

p = floor(log2(sample\_size)) + 1;

end

function ssqr = getCorrectedSampleVariance(sample, sample\_mean)

ssqr = sum((sample - sample\_mean).^2) / (length(sample) - 1);

end

function X = readFromFile(filename)

X = readmatrix(filename);

end

**3. 2. Полученные результаты**

Минимальное значение выборки: -6.480000

Максимальное значение выборки: -1.510000

Размах выборки: 4.970000

Размер выборки: 120

Выборочное математическое ожидание: -3.676167

Исправленная выборочная дисперсия: 0.866410

Всего интервалов: 7

Интервалы:

[-6.480000; -5.770000), [-5.770000; -5.060000), [-5.060000; -4.350000), [-4.350000; -3.640000), [-3.640000; -2.930000), [-2.930000; -2.220000), [-2.220000; -1.510000]

На рисунке 1 представлены гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием и дисперсией S2.

На рисунке 2 представлены график эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием и дисперсией S2.

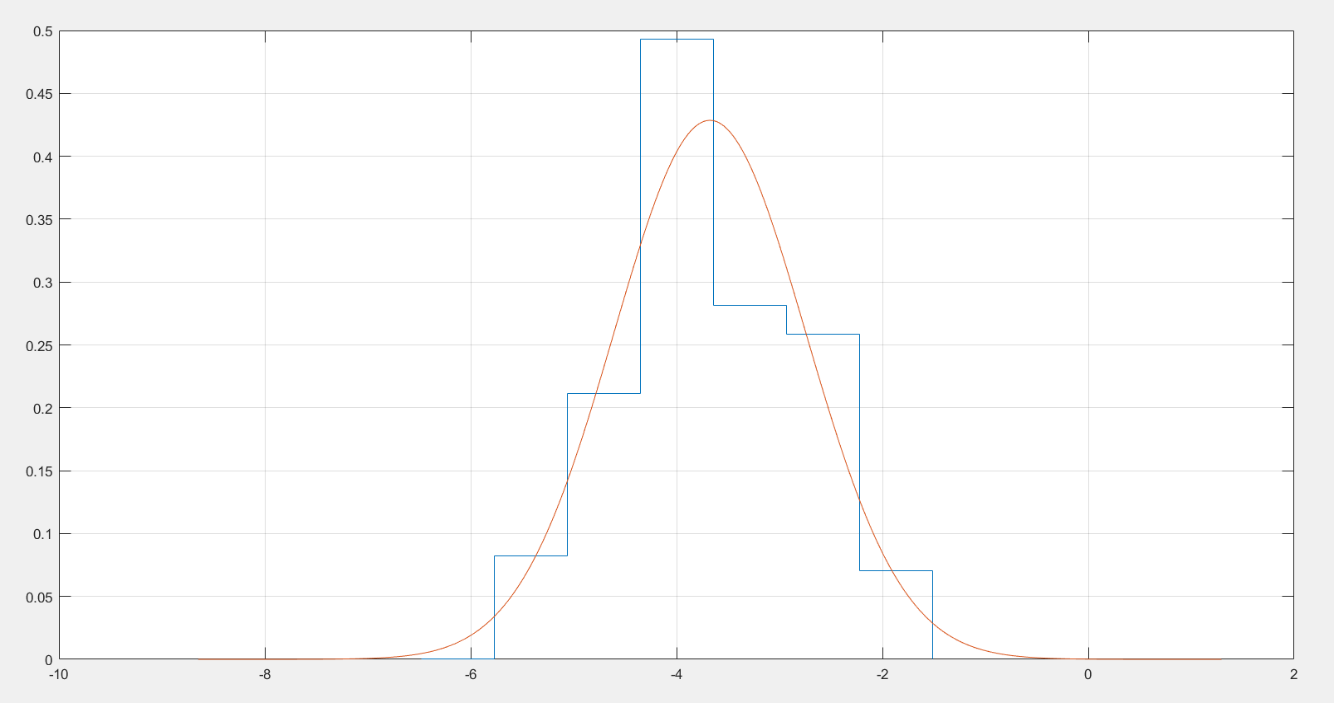


Рис. 1. Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием и дисперсией S2.

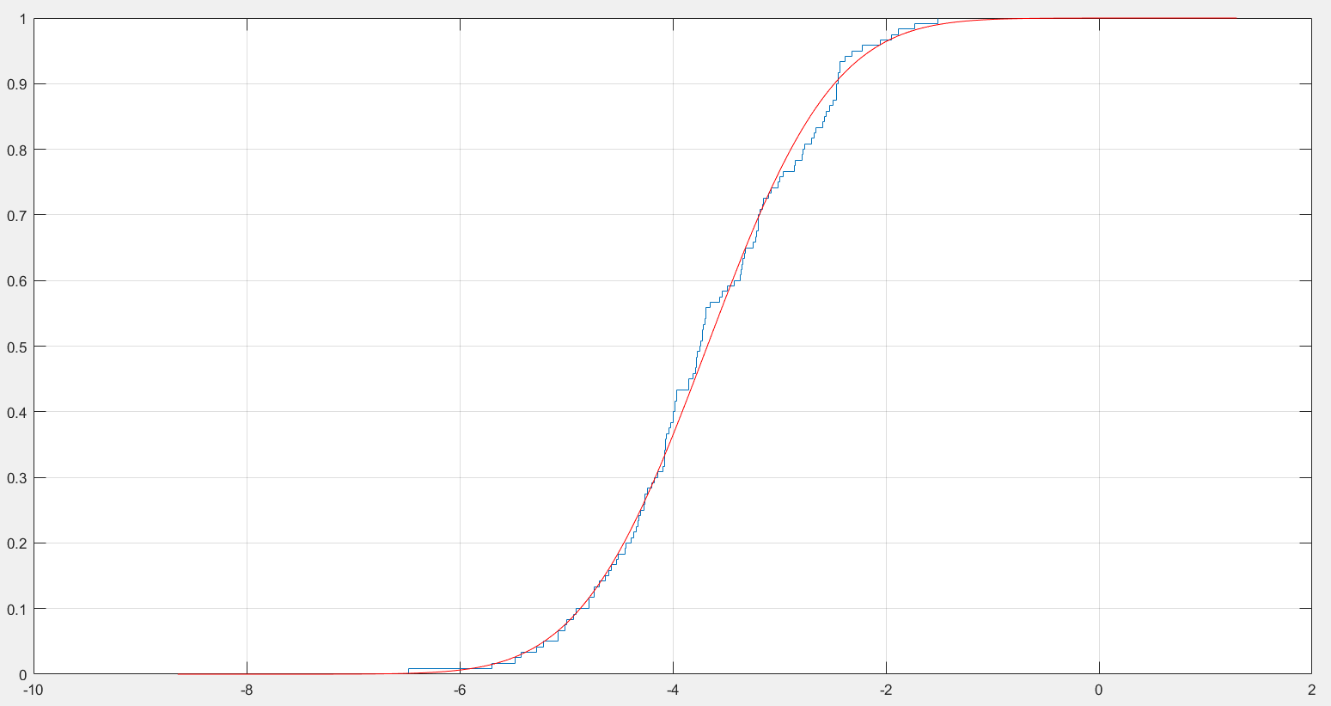


Рис. 2. График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием и дисперсией S2.