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

О т ч е т

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

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:
    a. вычисление максимального значения M_{max} и минимального значения M_{min};
    b. размаха R выборки;
    c. вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;
    d. группировку значений выборки в m = [log2n] + 2 интервала;
    e. построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
    f. 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.

2. Теоретическая часть

2. 1. Необходимые формулы

Пусть $(x_1, \ldots, x_n)$ – реализация случайной выборки.

Формула для вычисления максимального значения выборки M_{max}:

$$M_{max} = max(x_1, \dots, x_n).$$

Формула для вычисления минимального значения выборки M_{min}:

$$M_{min} = min(x_1, \dots, x_n).$$

Формула для вычисления размаха выборки:

$$R = M_{max} - M_{min}, \text{ где}$$

M_{max} – максимальное значение выборки;

M_{min} – минимальное значение выборки.

Выборочным средним (выборочным математическим ожиданием) называется статистика

$$\hat{\mu}(\overrightarrow{X_n}) = \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Выборочной дисперсией называется статистика

$$\widehat{\delta^2}(\overrightarrow{X_n}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_n})^2.$$

Формула для вычисления несмещенной оценки дисперсии:

$$S^2(\overrightarrow{X_n}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2.$$

2. 2. Необходимые определения

Предположим, что для выборки $\vec{x}$ построен интервальный статистический ряд. Отрезок $J = [X_{(1)}, X_{(n)}]$ делят на p равновеликих частей:

$$J_i = [a_i, a_{i+1}),\ i = \overline{1, p-1}$$

$$J_p = [a_{p-1}, a_p], \text{ где}$$

$$a_i = x_{(1)} + i\Delta,\ i = \overline{0, p}$$

$$\Delta = \frac{|J|}{p} = \frac{x_{(n)} - x_{(1)}}{p}.$$

Определение 2.1. Эмпирической плотностью распределения случайной выборки $\overrightarrow{X_n}$ (отвечающей выборке $\vec{x}$) называют функцию

$$\widehat{f_n} = \begin{cases} \frac{n_i}{n\Delta}, & x \in J_i, i = \overline{1, p} \\ 0, & \text{иначе} \end{cases}, \text{ где}$$

n_i — количество элементов выборки $\vec{x}$, которые принадлежат J_i, $i = \overline{1, p}$;

n — количество элементов в выборке.

Определение 2.2. Гистограммой называют график эмпирической плотности.

Пусть $\vec{X} = (X_1, \ldots, X_n)$ – выборка из генеральной совокупности X; $\vec{x} = (x_1, \ldots, x_n)$ – реализация случайной выборки $\vec{X}$. Обозначим через $n(x, \vec{x})$ число элементов вектора $\vec{x}$, которые имеют значение, меньшее x.

Определение 2.3. Эмпирической функцией распределения называют функцию $F_n: \Re \to \Re$, определенную условием $F_n = \frac{n(x,\vec{x})}{n}$.

Замечание. F_n обладает всеми свойствами функции распределения.

Замечание. F_n кусочно-постоянна и скачкообразно изменяет свои значения в точке $z_{(i)}$.

Замечание. Если элементы выборки $\vec{x}$ попарно различны, то

$$F_n = \begin{cases} 0, x \le x_{(1)} \\ \frac{1}{n}, x_{(1)} < x \le x_{(2)} \\ \cdots \\ \frac{i}{n}, x_{(i)} < x \le x_{(i+1)} \\ \cdots \\ 1, x > x_n \end{cases}, i = \overline{1, n-1}.$$

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

$\tilde{X}$	$Z_{(1)}$	…	$Z_{(m)}$
p	n_1/n	…	n_m/n

В дальнейшем это позволит рассматривать числовые характеристики случайной величины $\widetilde{X}$ как приближенные значения числовых характеристик случайной величины X.

3. 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 $\hat{\mu}$ и дисперсией S^2.

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

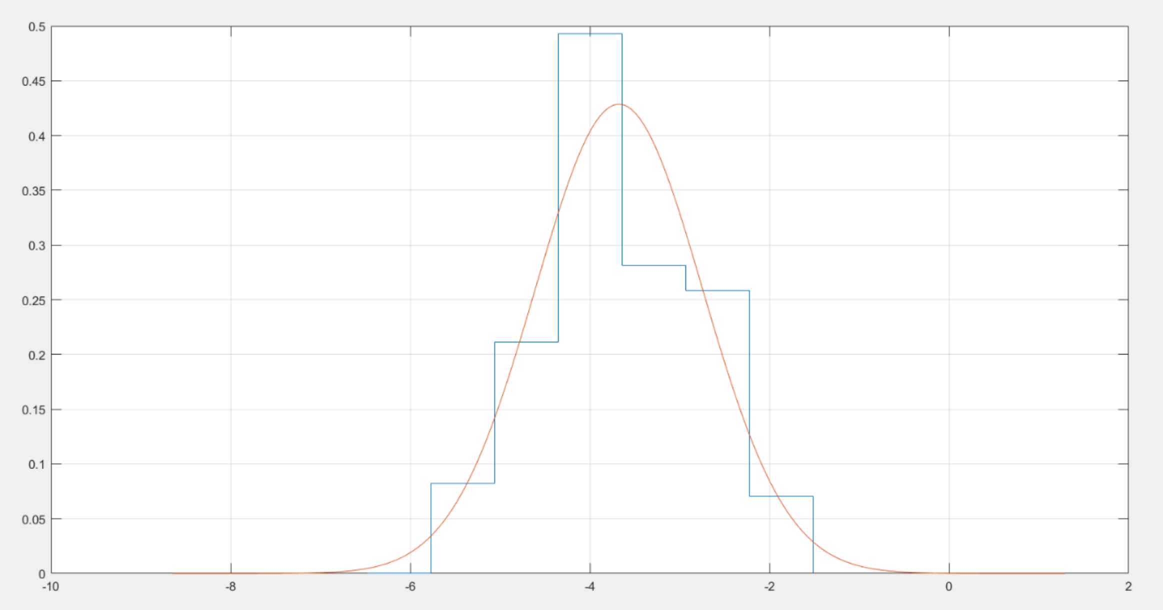

Рис. 1. Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.

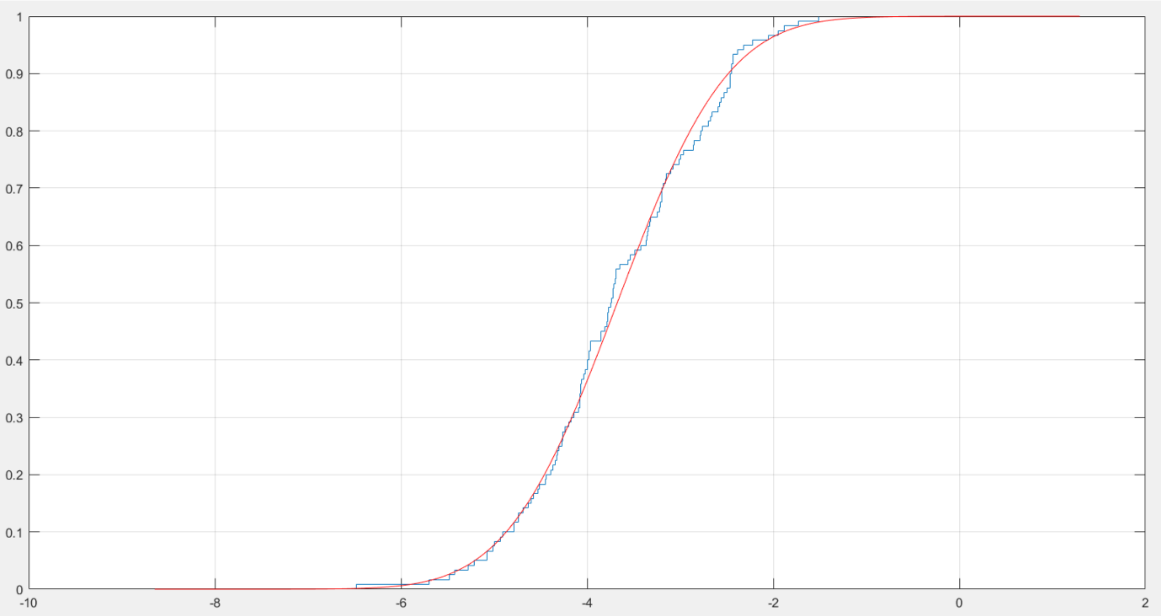

Рис. 2. График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.