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

Студент Егорова П.А.

Группа ИУ7-64Б

Оценка (баллы) ________________

Преподаватели Андреева Т.В.

Москва — 2023 г.

1 Задание

Цель работы: построение гистограммы и эмпирической функции распределения.

1. Для выборки объёма 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 $m = [\log_2 n] + 2$ 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 Теоретические сведения

2.1 Формулы для вычисления величин

Минимальное и максимальное значения выборки

$$\begin{gathered} M_{\max} = X_{(n)} \\ M_{\min} = X_{(1)} \end{gathered} \tag{1}$$

Размах выборки

$$R = M_{\max} - M_{\min}. \tag{2}$$

Оценки математического ожидания и исправленной дисперсии

$$\begin{aligned} \hat{\mu}(\vec{X}_n) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ S^2(\vec{X}_n) &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \end{aligned} \tag{3}$$

3 Определение эмпирической плотности и гистограммы

Пусть $\vec{x}$ – выборка из генеральной совокупности X. Если объем n этой выборки велик, то значения x_i группируют в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ делят на m равновеликих частей:

$$J_i = [x_{(1)} + (i-1) \cdot \Delta, x_{(1)} + i \cdot \Delta), i = \overline{1; m-1},$$

$$J_m = [x_{(1)} + (m-1) \cdot \Delta, x_{(n)}],$$

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

Интервальным статистическим рядом называют таблицу:

J_1	...	J_i	...	J_m
n_1	...	n_i	...	n_m

где n_i – количество элементов выборки $\vec{x}$, которые $\in J_i$.

Обычно выборку разбивают на $m = [\log_2 n] + 2$ интервалов, где n – размер выборки.

Гистограмма – это график эмпирической плотности.

Эмпирической плотностью, отвечающей выборке $\vec{x}$, называют функцию:

$$\hat{f}(x) = \begin{cases} \dfrac{n_i}{n\Delta}, & x \in J_i, i = \overline{1; m}, \\ 0, & \text{иначе}, \end{cases} \tag{4}$$

где J_i – полуинтервал статистического ряда, n_i – количество элементов выборки, входящих в полуинтервал, n – количество элементов выборки.

4 Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, ..., x_n)$ – выборка из генеральной совокупности X. Обозначим $n(x, \vec{x})$ – число элементов вектора $\vec{x}$, которые имеют значения меньше x.

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную как:

$$F_n(x) = \frac{n(x, \vec{x})}{n}. \tag{5}$$

Замечание.

1. Обладает всеми свойствами функции распределения;

2. Кусочно-постоянна;

3. Ксли все элементы вектора различны, то

$$F(x) = \begin{cases} 0, & x \le x_{(1)}, \\ \frac{i}{n}, & x_{(i)} < x \le x_{(i+1)}, \\ 1, & x > x_{(n)}. \end{cases} . \tag{6}$$

5 Результат работы

Вариант 3

5.1 Код программы

```
function lab1()

    function myhist()
        centers = zeros(1, m);
        heights = zeros(1, m);

        for i = 1:m
            heights(i) = counts(i) / (n * delta);
        end

        for i = 1:m
            centers(i) = bins(i + 1) - (delta / 2);
        end

        fprintf("Высоты столбцовгистограммы:\n");
        for i = 1:m
            fprintf("%ыйd- столбец: %f\n", i, heights(i));
        end

        set(gca, "xtick", bins);
        set(gca, "ytick", sort(heights));
        set(gca, "xlim", [min(bins) - 1, max(bins) + 1]);
        bar(centers, heights, 1);

        nodes = (m_min - 5):(S / 250):(m_max + 5);
        X_pdf = normpdf(nodes, mu, sqrt(S));
        plot(nodes, X_pdf, "r");
    end

    function [nt] = count_nt(t, X)
        tolerance = 1e-4;
        nt = 0;
        for x_index = 1:length(X)
            if (X(x_index) - t < tolerance)
                nt = nt + 1;
            end
        end
    end
```

```

function mycdf()
    x_tiks = unique(X);
    x_tiks(end + 1) = x_tiks(end) + 1;
    heights = zeros([1 length(x_tiks)]);
    for i = 1:length(x_tiks)
        heights(i) = count_nt(x_tiks(i), X) / n;
    end

    bin_Set = unique(bins);
    heights_Set = unique(heights);

    set(gca, "xtick", bin_Set);
    set(gca, "ylim", [0, max(heights_Set) + 0.3]);
    stairs(x_tiks, heights);

    nodes = (m_min - 2):(S / 250):(m_max + 2);
    X_cdf = normcdf(nodes, mu, sqrt(S));
    plot(nodes, X_cdf, "r");
end

X = [-0.45,-0.33,2.92,-1.25,-1.20,0.05,-0.53,-0.19,1.49,0.67,...
    0.22,1.23,0.50,-0.92,0.90,-1.52,-0.15,-1.24,-0.47,-0.45,...
    0.18,-0.05,1.58,1.74,2.37,-0.24,-1.34,1.05,1.28,1.37,1.18,...
    0.22,0.11,0.28,-0.64,-0.39,-1.77,-1.61,0.47,0.77,-0.27,-1.19,...
    -0.25,1.04,-0.16,0.42,0.29,0.10,1.04,0.43,-0.67,0.41,-0.62,...
    -1.49,1.46,-2.77,2.09,0.88,-0.36,-0.71,-0.62,1.34,-0.78,-0.15,...
    2.69,0.92,1.68,-0.12,0.34,0.74,1.72,1.24,0.23,0.76,0.87,...
    -1.52,0.63,-0.56,0.83,0.31,-0.18,0.99,-1.01,0.58,1.21,-1.51,...
    0.65,0.35,-0.37,-0.50,-0.73,0.63,0.33,1.56,-0.98,0.85,0.56,...
    -1.07,1.47,1.44,1.91,0.24,1.34,0.99,1.27,0.11,0.22,-0.25,0.35,...
    -0.03,-0.56,-0.79,2.41,-0.45,-0.44,0.07,0.64,0.69,0.10,-0.28];

X = sort(X);

% вычислениемаксимальногоиминимальногозначения

m_max = max(X);
m_min = min(X);
fprintf("-----------------------------------------\n");
fprintf("1. Максимальноезначениевыборки: M_max = %f.\n", m_max);
fprintf(" Минимальноезначениевыборки: M_min = %f.\n", m_min);
fprintf("-----------------------------------------\n");

% Вычислениеразмахавыборки
```

```
r = m_max - m_min;
fprintf("2. Размахвыборки: R = %f.\n", r);
fprintf("----------------------------------------\n");

% Вычислениеоценокматематическогоожиданияидисперсии

n = length(X);
mu = sum(X) / n;
S = sum((X - mu).^2) / (n - 1);
fprintf("3. Оценкаматематическогоожидания: m = %f.\n", mu);
fprintf(" Оценкадисперсии: S^2 = %f.\n", S);
fprintf("----------------------------------------\n");

% Группировказначенийвыборкивм = [log_2 n] + 2 интервала

m = floor(log2(n)) + 2;
bins = [];
cur = m_min;
delta = r / m;

for i = 1:(m + 1)
    bins(i) = cur;
    cur = cur + delta;
end

eps = 1e-6;
counts = [];

for i = 1:(m - 1)
    cur = 0;

    for j = 1:n
        if ((X(j) - eps) > bins(i) || abs(bins(i) - X(j)) < eps) && X(j) < (bins(i + 1) -
            eps)
            cur = cur + 1;
        end
    end

    counts(i) = cur;
end

cur = 0;
for i = 1:n
```

```
        if (bins(m) < X(i) || abs(bins(m) - X(i)) < eps) && (X(i) < bins(m + 1) || abs(bins(m +
            1) - X(i)) < eps)
            cur = cur + 1;
        end
    end

    counts(m) = cur;

    fprintf("4. Группировказначенийвыборкив%d интервалов:\n", m);
    for i = 1:(m)
        fprintf("Интервал №%d [%f : %f) - %d значенийизвыборки.\n", i, bins(i), bins(i + 1),
            counts(i));
    end
    fprintf("----------------------------------------\n");

    % ПостроениегистограммыифункцииплотностираспределениянормальнойСВ.

    fprintf("5. ПостроениегистограммыиграфикафункцииплотностираспределениянормальнойСВ.\n");
    figure;
    hold on;
    grid on;
    myhist();
    xlabel('x')
    ylabel('fn(x)')
    hold off;
    fprintf("----------------------------------------\n");

    % ПостроениеграфикаэмпирическойфункциираспределенияифункциираспределениянормальнойСВ.
    fprintf("6.
        ПостроениеграфикаэмпирическойфункциираспределенияифункциираспределениянормальнойСВ
        .\n");
    figure;
    hold on;
    grid on;
    mycdf();
    xlabel('x')
    ylabel('Fn(x)')
    hold off;
end
```

6 Результаты расчётов

$$M_{\min} = -2.77$$

$$M_{\max} = 2.92$$

$$R = 5.69$$

$$\hat{\mu}(\vec{x}_n) = 0.23225$$

$$S^2(\vec{x}_n) = 1.0406$$

$$m = 8$$

[-2.77; -2.06) - 1
[-2.06; -1.35) - 6
[-1.35; -0.64) - 15
[-0.64; 0.07) - 30
[0.07; 0.79) - 33
[0.79; 1.50) - 24
[1.50; 2.21) - 7
[2.21, 2.92] - 4

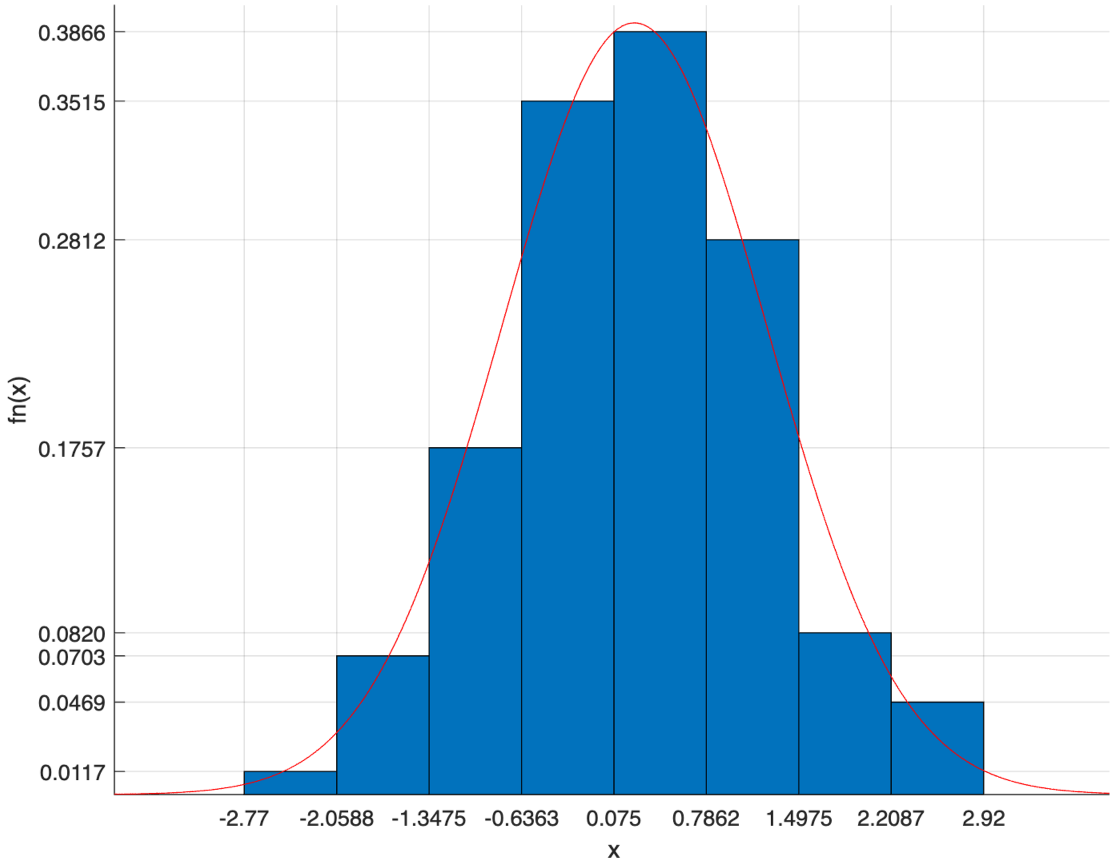

Рис. 6.1: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочными мат. ожиданием и дисперсией

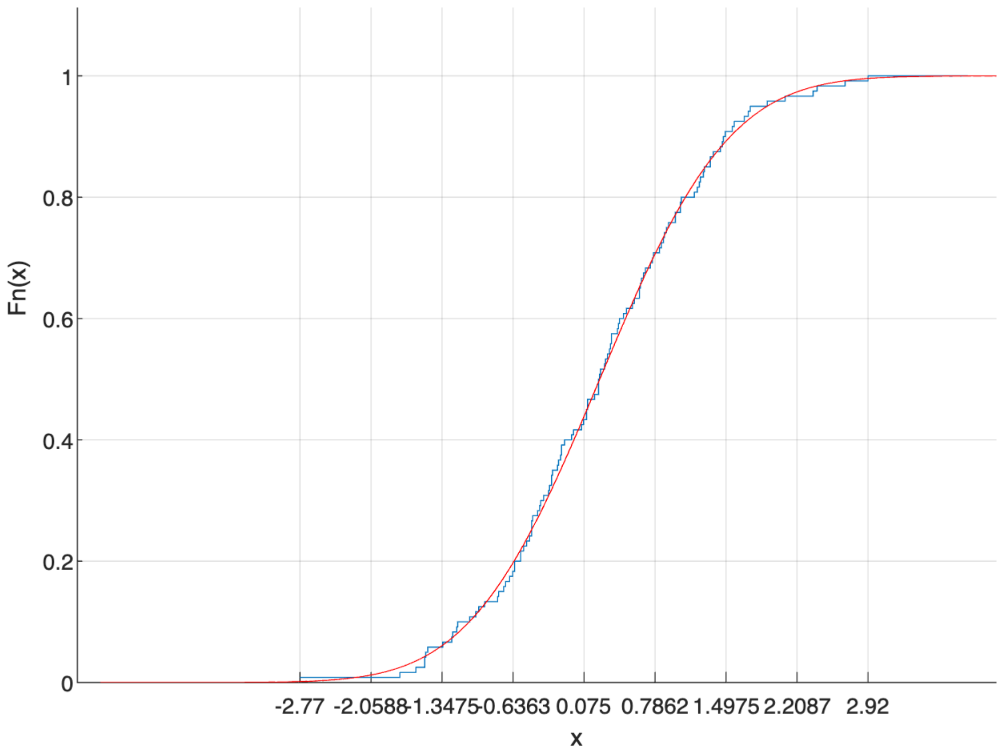

Рис. 6.2: График эмпирической функции распределения и функции распределения нормальной случайной величины с выборочными мат. ожиданием и дисперсией