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 дисциплине «Математическая статистика»

Тема Гистрограмма и эмпирическая функция распределения

Студент Криков А.В.

Группа ИУ7-63Б

Оценка (баллы) ____________________

Преподаватель Власов П. А.

Москва — 2022 г.

Задание

Цель работы

Построение гистограммы и эмпирической функции распределения.

Постановка задачи

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.

Теоретические сведения

Формулы для вычисления величин

1. **Максимальное значение выборки** $M_{max} = max(x_1, .., x_n)$
2. **Минимальное значение выборки** $M_{min} = min(x_1, .., x_n)$
3. **Размах выборки** $R = M_{max} - M_{min}$
4. **Выборочное среднее (математическое ожидание)** $\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$
5. **Несмещённая оценка дисперсии (состоятельная оценка)** $S^2(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$

   где $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Определение эмпирической плотности и гистограммы

Эмпирической плотностью (отвечающей выборке $\vec{x}$) называют функцию

$$\hat{f}(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1; p} \\ 0, \text{ иначе} \end{cases}$$

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

Пусть $\vec{x}$ – выборка из генеральной совокупности X. Если объем n этой выборки велик, то значения x_i группируют не только в статистический ряд, но и в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ (где $x_{(1)} = min(x_1, .., x_n)$, $x_{(n)} = max(x_1, .., x_n)$) делят на m равновеликих частей:

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

$$J_m = [a_m, a_{m+1}]$$

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

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

Интервальным статистическим рядом называют таблицу

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

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

Требуемый интервал $m = [\log_2 n] + 2$

Гистограмма – это график эмпирической плотности.

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

Практическая часть

Код программы

```
% Вариант 11

function main()
    X =[
        -2.54,-0.79,-4.27,-3.09,-3.82,-0.61,...
        -0.64,-1.24,-1.73,-2.91,-1.48,-1.28,...
        -0.37,-1.88,-2.19,-1.61,-1.52,-3.17,...
        -1.36,-3.08,-3.11,-3.07,-1.57,-1.51,...
        -2.37,-0.58,-3.05,-2.93,-1.01,-1.40,...
        -2.06,-3.05,-1.84,-1.24,-1.89,-2.06,...
        -1.59,-2.83,-1.07,-2.96,-3.17,-3.08,...
        -0.49,-3.11,-3.14,-2.30,-3.99,-1.56,...
        -1.28,-3.46,-2.63,-0.82,-2.18,-0.89,...
        -3.08,-1.13,-1.62,-1.06,-2.98,-1.55,...
        -1.49,-1.65,-1.45,-2.29,-0.85,-1.44,...
        -2.87,-2.40,-2.13,-3.52,-1.42,-3.64,...
        -3.47,-2.05,-2.39,-2.07,-0.80,-1.52,...
        -3.92,-2.22,-0.78,-2.60,-1.78,-1.61,...
        -1.65,-2.06,-3.33,-3.41,-1.97,-1.74,...
        -2.04,0.01,-1.37,-3.15,-2.35,-3.66,...
        -1.79,-2.56,-1.87,-1.06,-0.64,-2.49,...
        -1.85,-1.40,-0.86,-0.17,-0.62,-2.85,...
        -2.12,-1.17,-2.48,-1.65,-3.74,-2.87,...
        -3.15,-1.89,-1.34,-4.33,-0.96,-1.79];

    % а) Максимальное и минимальное значения
    M_max = max(X);
    M_min = min(X);

    fprintf("\na) M_max = %f; M_min = %f\n", M_max, M_min);

    % б) Размах
    R = M_max - M_min;

    fprintf("б) R = %f\n", R);

    % в) Оценка мат. ожидания
    mu = sum(X) / length(X);

    % Несмещенная оценка дисперсии
    s2 = sum((X - mu).^ 2) / (length(X) - 1);

```

```
        fprintf("в) mu = %f; S^2 = %f\n", mu, s2);

        % г) Группировка значений выборки

        % Нахождение количества интервалов
        m = floor(log2(length(X))) + 2;

        fprintf("г) m = %f\n", m);


        % Группировка значений в m интервалов
        delta = R / m;
        edges = zeros(1, m + 1);
        counts = zeros(1, m + 1);

        % Указываем границы интервалов
        for i = 1:m + 1
            edges(i) = M_min + delta * (i - 1);
        end

        for i = 1:length(X)
            for j = 1:m
                if X(i) >= edges(j) && X(i) < edges(j+1)
                    counts(j) = counts(j) + 1;
                    break;
                end
            end
        end
        counts(m) = counts(m) + 1;

        % Вывод всех интервалов и количества элементов
        for i = 1:m - 1
            fprintf("[ %f ; %f ) - %d\n", edges(i), edges(i + 1), counts(i));
        end
        fprintf("[ %f ; %f ] - %d\n\n\n", edges(m), edges(m + 1), counts(m));

        % д) Построение гистограммы
        for i = 1:m + 1
            counts(i) = counts(i) / (length(X) * delta);
        end

        stairs([edges(1), edges], [0 counts], 'LineWidth', 2);

        % Построение на одной координатной плоскости
        grid on;
        hold on;

        Xn = M_min : delta / 20 : M_max;
        sigma = sqrt(s2);

        % График функции плотности распределения вероятностей нормальной случайной величины
        Y = normpdf(Xn, mu, sigma);
        plot(Xn, Y, 'LineWidth', 2);
        legend('Гистограмма', 'Ф-я плотности распрделения нсв', 'Location', 'northwest');
```

```

    % e)
    % Новая координатная плоскость
    figure;

    % Построение графика эмпрической функции распределения
    [YY, XX] = ecdf(X);
    stairs(XX, YY);
    hold on;

    % Построение графика функции распределения нормальной случайной величины
    Y = normcdf(Xn, mu, sigma);
    plot(Xn, Y, 'black');

    legend('Эмпирическая функция распределения', 'Функция распределения',
        'Location', 'northwest')
end
```

Результаты расчётов

$$M_{\min} = -4,33$$

$$M_{\max} = 0,01$$

$$R = 4,34$$

$$\hat{\mu}(\vec{x}_n) = -2,0585$$

$$S^2(\vec{x}_n) = 0,943988$$

$$m = 8$$

```
а) M_max = 0.010000; M_min = -4.330000
б) R = 4.340000
в) mu = -2.058500; S^2 = 0.943988
г) m = 8.000000
[ -4.330000 ; -3.787500 ) - 5
[ -3.787500 ; -3.245000 ) - 8
[ -3.245000 ; -2.702500 ) - 22
[ -2.702500 ; -2.160000 ) - 15
[ -2.160000 ; -1.617500 ) - 24
[ -1.617500 ; -1.075000 ) - 25
[ -1.075000 ; -0.532500 ) - 17
[ -0.532500 ; 0.010000 ] - 4
```

Рис. 1: Вывод программы

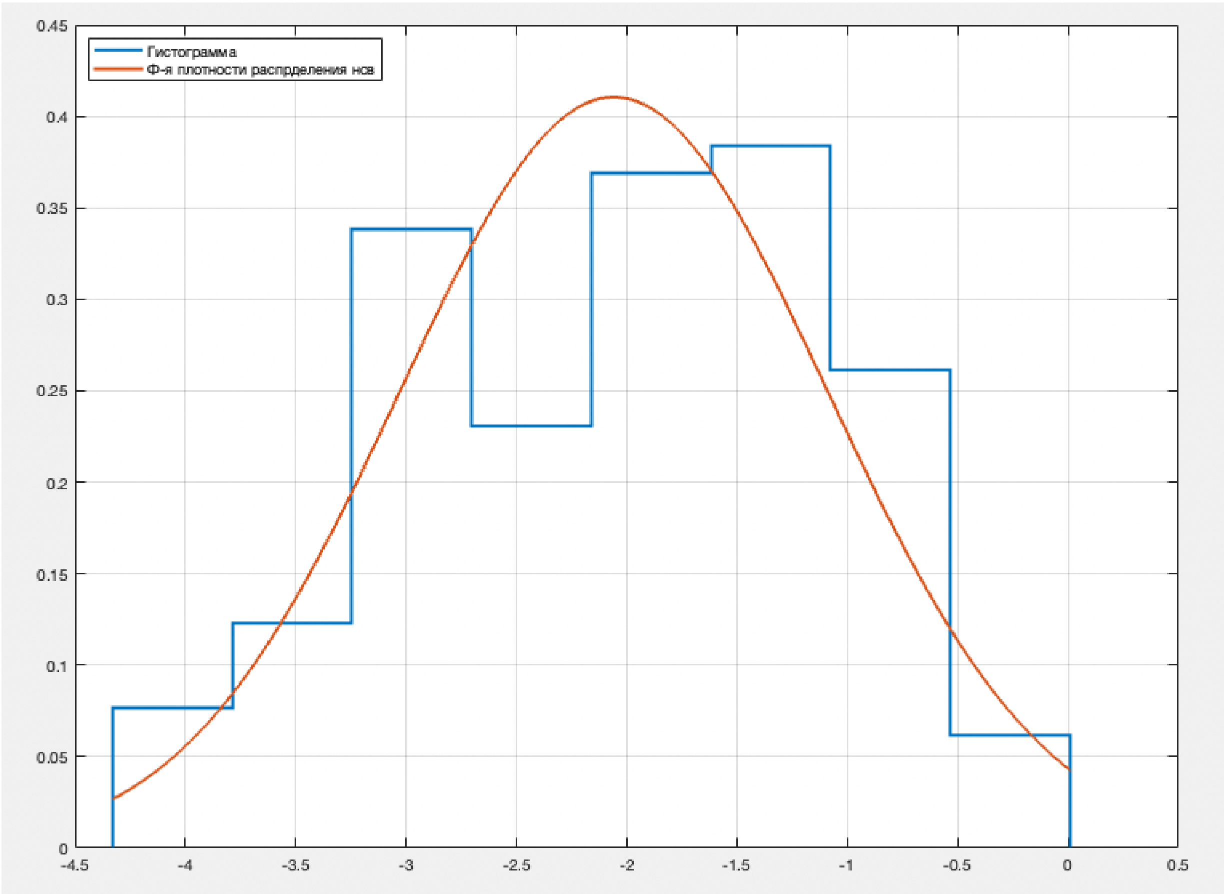

Рис. 2: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочными мат. ожиданием и дисперсией

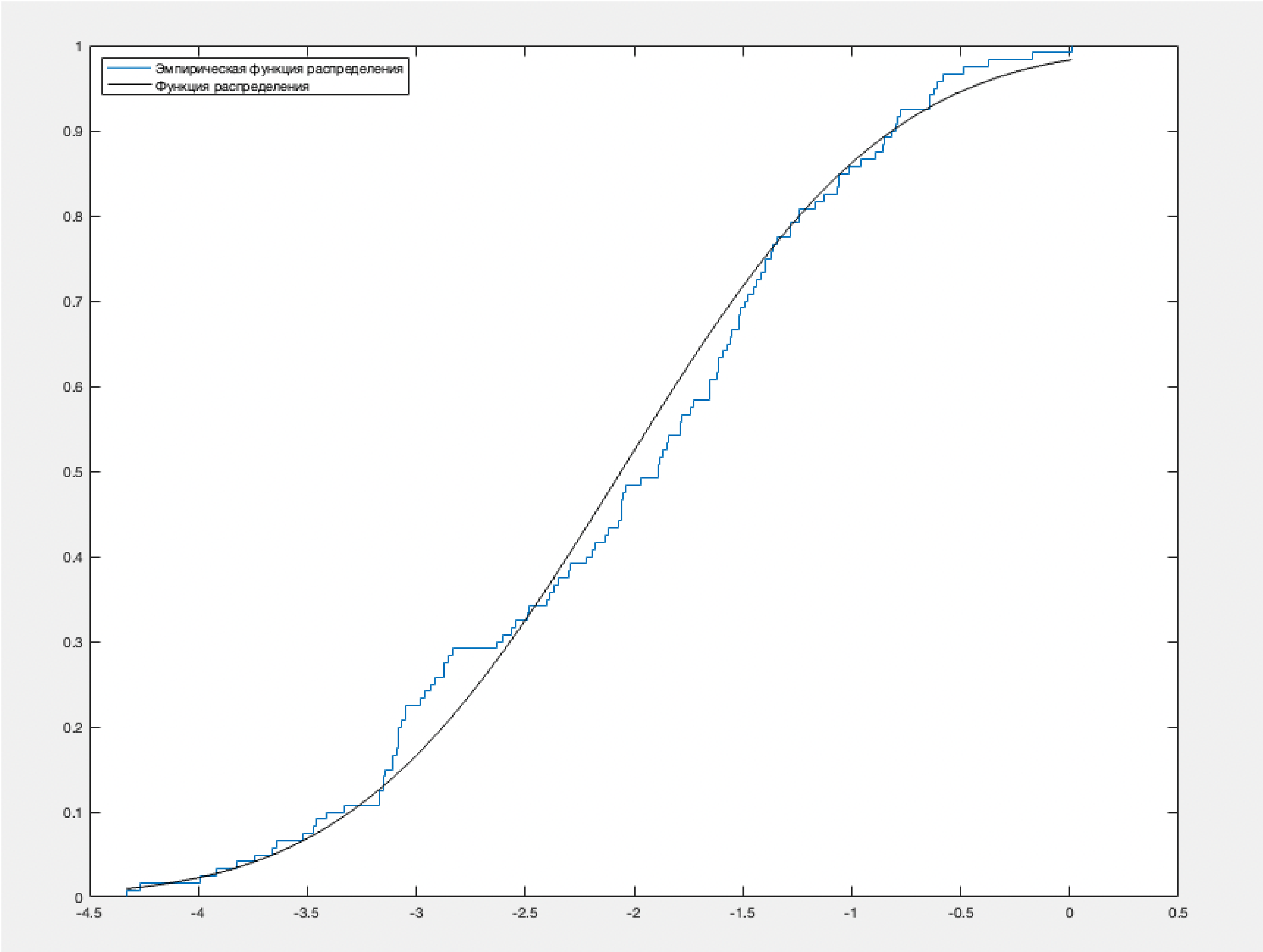

Рис. 3: График эмпирической функции распределения и функции распределения нормальной случайной величины с выборочными мат. ожиданием и дисперсией