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»

Отчёт

по лабораторной работе №1

Название: Гистограмма и эмпирическая функция распределения.

Дисциплина: Математическая статистика.

Студент	ИУ7-64Б (Группа)	(Подпись, дата)	Л.Е.Тартыков (И.О. Фамилия)
Преподаватель		(Подпись, дата)	М.А.Велищанский (И.О. Фамилия)

Москва, 2022

1 Задание

1.1 Цель работы

Цель работы: построение гистограммы и эмпирической функции распределения.

1.2 Содержание работы

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 Теоретическая часть

2.1 Формулы для вычисления величин M_{max} , M_{min}, R, $\hat{\mu}$, S^2

Минимальное значение выборки рассчитывается по формуле (2.1); максимальное – (2.2). Размах выборки рассчитывается по формуле (2.3); выборочное среднее – (2.4), исправленная выборочная дисперсия – (2.5).

$$M_{\min} = X_{(1)} \tag{2.1}$$

$$M_{\min} = X_{(n)} \tag{2.2}$$

$$R = M_{\max} - M_{\min}. \tag{2.3}$$

$$\hat{\mu}(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.4}$$

$$S^2(\vec{X}_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \tag{2.5}$$

2.2 Эмпирическая плотность и гистограмма

Пусть $\vec{x}$ – выборка из генеральной совокупности X.

При большом объеме n (n $>$ 50) этой выборки значения x_i группируют в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ делят на m равновеликих промежутков по формуле (2.6):

$$J_i = [x_{(1)} + (i-1)\cdot\Delta,\ x_{(1)} + i\cdot\Delta), i = \overline{1; m-1} \tag{2.6}$$

Последний промежуток определяется по формуле (2.7):

$$J_m = [x_{(1)} + (m-1)\cdot\Delta, x_{(n)}] \tag{2.7}$$

Ширина каждого из таких промежутков определяется по формуле (2.8).

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m} \tag{2.8}$$

Интервальным статистическим рядом называют таблицу 2.1:

Таблица 2.1 – Интервальный статистический ряд

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

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

Гистограмма – это график эмпирической плотности.

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

$$\hat{f}(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1; m} \\ 0 \quad , x \notin J \end{cases}$$

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

2.3 Эмпирическая функция распределения

Пусть $\vec{x} = (x_1, ..., x_n)$ – выборка из генеральной совокупности X.

Обозначим $n(t, \vec{x})$ – число элементов вектора $\vec{x}$, которые имеют значения меньше t.

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

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

3 Практическая часть

3.1 Код программы

Модуль разработанных алгоритмов представлен на листинге 3.1

Листинг 3.1 – Модуль разработанных алгоритмов

```
##Тартыков Лев ИУ7-64Б, 2022 г
pkg load statistics;
echo off all;

function main()
    [error, X] = load_data("data.txt");
    if (error == 0)
        [bins, counts, count_X, delta, Xn, Y_normpdf, Y_normcdf, Y_ecdf, M_min,
            M_max, X_without_double, J, count_elem_J] = perform_params(X);
        plot_graphs(bins, counts, count_X, delta, Xn, Y_normpdf, Y_normcdf,
            Y_ecdf, M_min, M_max, X_without_double, J, count_elem_J);
    endif
endfunction

function [error, X] = load_data(file_name)
    X = []; i = 1; error = 0;
    file = fopen(file_name, "r");
    if (file == -1)
        error = -1;
    else
        end_file = 0;
        while (end_file == 0)
            if (feof(file))
                end_file = 1;
            else
                X(i) = fscanf(file, '%f', [1,1]);
                fscanf(file, '%c', [1, 1]);
                i++;
            endif
        endwhile
    endif
    fclose(file);
endfunction

function [bins, counts, count_X, delta, Xn, Y_normpdf, Y_normcdf, Y_ecdf, M_min,
    M_max, X_without_double, J, count_elem_J] = perform_params(X)
    count_X = length(X);
    M_max = max(X)
```

```
    M_min = min(X)

    R = M_max - M_min;
    MX = find_MX(X, count_X);
    DX = find_DX(X, MX, count_X);

    m = find_m(count_X);
    output_results(M_max, M_min, R, MX, DX, m);
    [counts, bins] = hist(X, m);

    delta = R / m;
    sigma = sqrt(DX);
    abs_MX = abs(MX);
    [J, count_elem_J] = calc_hist(X, M_min, M_max, delta);
    M_min = M_min - abs_MX / 2;
    M_max = M_max + abs_MX / 2;
    Xn = M_min:delta/20:M_max;

    Y_normpdf = density_ndist(Xn, MX, sigma);
    Y_normcdf = form_normcdf(Xn, MX, sigma);
    [count_elem, X_without_double] = count_number_elems(X, count_X, M_min, M_max
        );
    Y_ecdf = form_y_ecdf(count_elem, count_X, M_min, M_max);
endfunction

function [MX] = find_MX(X, count_X)
    MX = sum(X) / count_X;
endfunction

function [DX] = find_DX(X, MX, count_X)
    DX = sum((X - MX).^2) / (count_X - 1);
endfunction

function [m] = find_m(count_X)
    m = floor(log2(count_X)) + 2;
endfunction

function [J, count_elem_J] = calc_hist(X, M_min, M_max, delta)
    J = (M_min - delta): delta: (M_max + delta);
    count_elem_J = [];
    len_J = length(J);
    len_X = length(X);
    for i = 2: len_J - 2
        count_elem = 0;
        for j = 1: len_X
            if ((X(j) > J(i)) || abs(X(j) - J(i)) <= 1e-3)
                if ((i == len_J - 2) && (X(j) < J(i + 1) ||
                    (abs(X(j) - J(i + 1))) <= 1e-3))
```

```
                    count_elem += 1;
                elseif (X(j) < J(i + 1))
                    count_elem += 1;
                endif
            endif
        endfor
        count_elem_J(i) = count_elem;
    endfor
    count_elem_J(len_J) = 0;
endfunction

function [Y_normpdf] = density_ndist(Xn, MX, sigma)
    Y_normpdf = [];
    count_X = length(Xn);
    for i = 1: count_X
        Y_normpdf(i) = \
            1 / (sqrt(2 * pi) * sigma) * exp(-(Xn(i) - MX).^2 / (2 * sigma.^2));
    endfor
endfunction

function [count_elem, X_graph] = count_number_elems(X, count_X, M_min, M_max)
    X_sort = sort(X);
    count_elem = [];
    X_without_double = [];
    i = 1;
    index_count = 1;
    while (i < count_X)
        is_all = 0;
        temp_value = X_sort(i);
        j = 1;
        while (is_all == 0)
            if (X_sort(i + j) == temp_value && i + j < count_X)
                j++;
            else
                is_all = 1;
            endif
        endwhile
        X_without_double(index_count) = temp_value;
        count_elem(index_count) = j;
        index_count += 1;
        i += j;
    endwhile

    if (X_sort(count_X) != X_sort(count_X - 1))
        count_elem(index_count) = 1;
        X_without_double(index_count) = X_sort(count_X);
    endif

```

```
    X_graph = [];
    X_graph(1) = X_without_double(1) - abs(M_min);
    X_graph(2) = X_without_double(1);
    X_graph(3) = X_graph(2);
    j = 4;
    for i = 2: length(X_without_double)
        X_graph(j) = X_without_double(i);
        X_graph(j + 1) = X_without_double(i);
        j += 2;
    endfor
    X_graph(j) = abs(M_max);
endfunction

function [Y_normcdf] = form_normcdf(Xn, MX, sigma)
    Y_normpdf = \
     @(Xn) ((1./(sqrt(2.*pi).*sigma)).*(exp((-0.5.*(Xn - MX).^2)./(sigma.^2))));

    for k = 1: length(Xn)
        Y_normcdf(k) = integral(Y_normpdf, -inf, Xn(k));
    endfor
endfunction

function [Y_ecdf] = form_y_ecdf(count_elem, count_X)
    Y_ecdf = [];
    len_celem = length(count_elem);

    Y_ecdf(1) = 0;
    Y_ecdf(2) = 0;
    Y_ecdf(3) = count_elem(1) / count_X;
    Y_ecdf(4) = Y_ecdf(3);
    j = 5;
    for i = 2: len_celem
        Y_ecdf(j) = Y_ecdf(j - 1) + count_elem(i) / count_X;
        Y_ecdf(j + 1) = Y_ecdf(j);
        j += 2;
    endfor
endfunction

function plot_graphs(bins, counts, count_X, delta, Xn, Y_normpdf, Y_normcdf,
   Y_ecdf, M_min, M_max, X_without_double)
    figure;
    subplot(1, 2, 1);
    stairs(J, count_elem_J / (count_X * delta));
    hold on;
    plot(Xn, Y_normpdf, 'LineWidth', 3, 'Color', 'green');
    xlim([M_min, M_max]);
    hold on;

```

```
    subplot(1, 2, 2);
    plot(Xn, Y_normcdf, 'LineWidth', 1, 'Color', 'red');
    xlim([M_min, M_max]);
    hold on;
    plot(X_without_double, Y_ecdf, 'LineWidth', 1, 'Color', 'blue');
    xlim([M_min, M_max]);
endfunction

function output_results(M_min, M_max, R, MX, DX, m)
    fprintf("M_min = %f,\nM_max = %f,\nR = %f,\nMX = %f,\nDX = %f,\nm = %f\n",
            M_min, M_max, R, MX, DX, m);
endfunction

main()
```

3.2 Результаты расчетов для выборки из индивидуального варианта.

Согласно варианту 15, результаты расчетов для выборки приведены на формулах (3.1), (3.2), (3.3), (3.4), (3.5), (3.6).

$$M_{\min} = -6.48 \tag{3.1}$$

$$M_{\max} = -1.51 \tag{3.2}$$

$$R = 4.97 \tag{3.3}$$

$$\hat{\mu}(\vec{x}_n) = -3.676 \tag{3.4}$$

$$S^2(\vec{x}_n) = 0.866 \tag{3.5}$$

$$m = 8 \tag{3.6}$$

На рисунке 3.1 представлены гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочными мат. ожиданием и дисперсией.

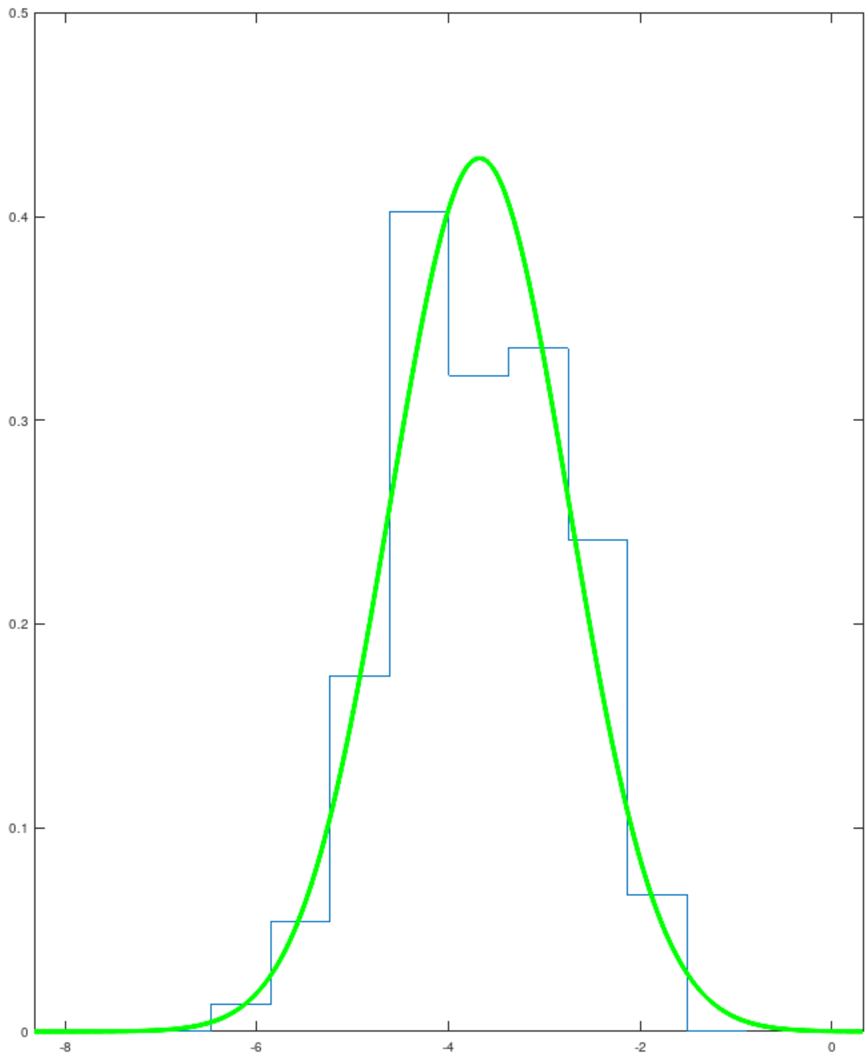

Рисунок 3.1 – Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочными мат. ожиданием и дисперсией.

На рисунке 3.2 представлены график эмпирической функции распределения и функции распределения нормальной случайной величины с выборочными мат. ожиданием и дисперсией.

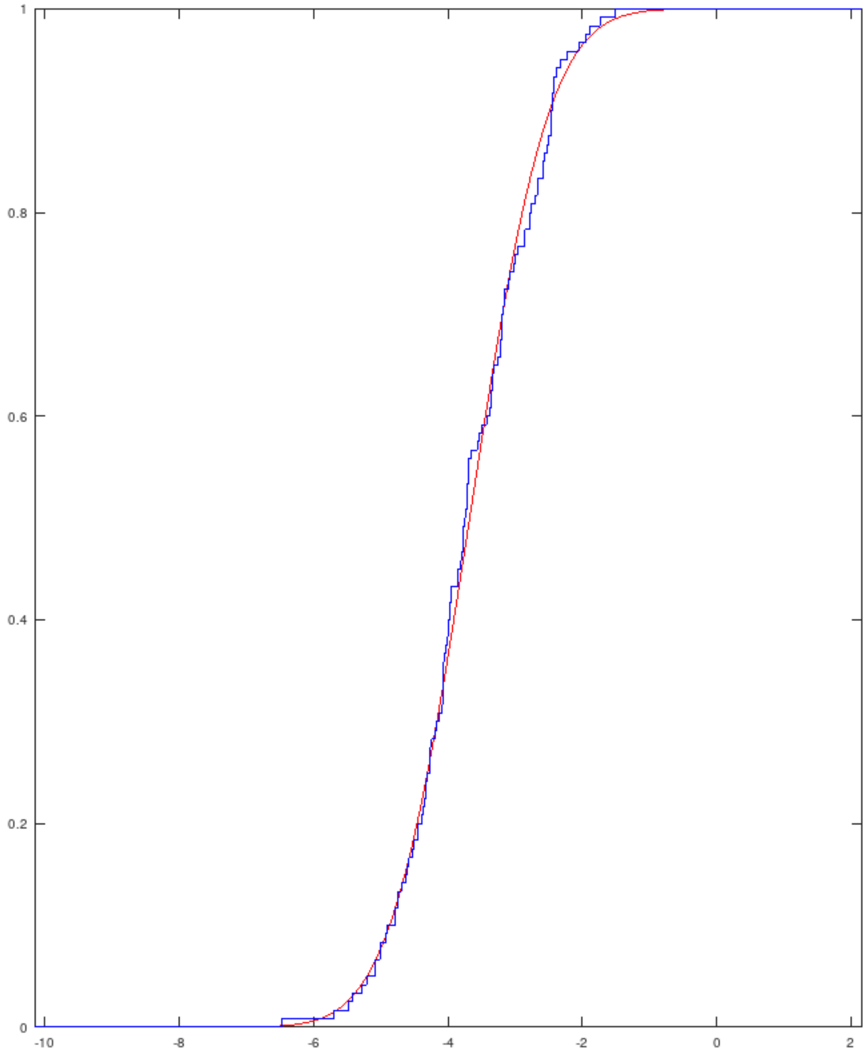

Рисунок 3.2 – График эмпирической функции распределения и функции распределения нормальной случайной величины с выборочными мат. ожиданием и дисперсией.