Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное 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»

Лабораторная работа №1
по курсу "Математическая статистика"

Тема Гистограмма и эмпирическая функция распределения

Студент Нисуев Н. Ф.

Группа ИУ7-62Б

Вариант 10

Преподаватель Власов П. А.

Москва — 2025 г.

СОДЕРЖАНИЕ

1 Содержание работы

1. Для выборки объема n из генеральной совокупности X реализовать в виде программы на ЭВМ

   а) вычисление максимального значения M_{max} и минимального значения M_{min};

   б) размаха R выборки;

   в) вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;

   г) группировку значений выборки в $m = [\log_2 n] + 2$ интервала;

   д) построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;

   е) 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 Формулы для вычисления величин

— Максимальное M_{max} значение выборки:

$$M_{\max} = X_{(n)}$$

— Минимальное M_{min} значение выборки:

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

— Размах R выборки:

$$R = M_{\max} - M_{\min}$$

— Выборочное среднее (оценка математического ожидания):

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

— Исправленная выборочная дисперсия:

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

3 Определения

Если объем выборки достаточно велик ($n > 50$), то элементы выборки группируются в так называемый статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ разбивают на m равновеликих промежутков.

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

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

где:

- $\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}$,
- $m = \lfloor \log_2 n \rfloor + 2$
- $x_{(1)} = min(\vec{x})$
- $x_{(n)} = max(\vec{x})$

3.1 Интервальный статистический ряд

Определение: Интервальным статистическим рядом называют таблицу, где n_i — число элементов выборки $\vec{x}$, попавших в J_i, $i = \overline{1, m}$.

J_1	$\cdots$	J_i	$\cdots$	J_m
n_1	$\cdots$	n_i	$\cdots$	n_m

3.2 Эмпирическая плотность

Пусть для данной выборки $\vec{x} = (x_1, \ \ldots, \ x_n)$ построен интервальный статистический ряд (J_i, n_i), $i = \overline{1; m}$

Определение: Эмпирической плотностью, отвечающей выборке $\vec{x}$, называется функция

$$f_n(x) = \begin{cases} \frac{n_i}{n \cdot \Delta}, \text{если } x \in J_i \\ 0, \text{иначе} \end{cases}$$

3.3 Гистограмма

Определение: Гистограммой называется график эмпирической функции плотности

3.4 Эмпирическая функция распределения

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

Обозначим $l(t, \vec{x})$ — число элементов выборки $\vec{x}$, которые меньше t.

Определение: Эмпирической функцией распределения, отвечающей выборке $\vec{x}$, называют отображение

$$F_n : \mathbb{R} \to \mathbb{R},$$

определенное правилом:

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

4 Текст программы

```
function lab1(X)
    X = sort(X);
    n = length(X);

    fprintf("a) Вычисление максимального значения Mmax " + ...
        "и минимального значения Mmin\n");

    Mmax = X(n);
    Mmin = X(1);

    fprintf("\nMmax = %.3f\n", Mmax);
    fprintf("Mmin  = %.3f\n", Mmin);

    fprintf("\nб) Вычисление размаха R\n");

    R = Mmax - Mmin;
    fprintf("\nR = %.3f\n", R);

    fprintf("\nв) Вычисление оценок Mu и S^2 " + ...
        "математического ожидания MX и дисперсии DX\n");

    Mu = sum(X) / n;
    S_square = sum((X - Mu) .^2) / (n - 1);

    fprintf("\nMu = %.3f\n", Mu);
    fprintf("S^2 = %.3f\n", S_square);

    fprintf("\nг) Группировка значений выборки в " + ...
        "m = [log2 n] + 2 интервала\n");

    m = floor(log2(n)) + 2;
    fprintf("\nКол-во интервалов m = %3d:\n\n", m);

    delta = (X(n) - X(1)) / m;

    borders = Mmin : delta : Mmax;

    ni_arr = zeros(m, 1);

```

```
    cur_X_pos = 1;

    for i = 1 : m
        count = 0;
        lower_bound = borders(i);
        upper_bound = borders(i+1);
        is_last_interval = (i == m);

        while cur_X_pos <= length(X) && ( ...
                (is_last_interval && X(cur_X_pos) <= upper_bound) || ...
                (~is_last_interval && X(cur_X_pos) < upper_bound))
            if X(cur_X_pos) >= lower_bound
                count = count + 1;
            end
            cur_X_pos = cur_X_pos + 1;
        end

        if i == m
            fprintf(" %d. [%.3f; %.3f], кол-во элементов: %d\n", ...
                i, borders(i), borders(i+1), count);
        else
            fprintf(" %d. [%.3f; %.3f), кол-во элементов: %d\n", ...
                i, borders(i), borders(i+1), count);
        end

        ni_arr(i) = count;
    end

    fprintf("\nд) Построение гистограммы и " + ...
        "графика плотности нормального распределения\n\n");

    mid_intervals = zeros(m, 1);

    for i = 1 : m
        mid_intervals(i) = (borders(i) + borders(i + 1)) / 2;
    end

    column_values = zeros(m, 1);

    for i = 1 : m
        column_values(i) = ni_arr(i) / (n * delta);
```

```
    end

    figure;
    bar(mid_intervals, column_values, 1);
    hold on;

    x_coords = (Mmin - 1) : 1e-3 : (Mmax + 1);
    func_density_norm = normpdf(x_coords, Mu, sqrt(S_square));
    plot(x_coords, func_density_norm, 'LineWidth', 2);
    grid;
    legend('Гистограмма', 'Плотность N(\mu, \sigma^2)');

    fprintf("\ne) Построение эмпирической функции распределения " + ...
        "и нормальной ФР\n\n");

    t_arr = [X(1) - 1, X, X(end) + 1];
    func_emperic = zeros(size(t_arr));

    for i = 1 : length(t_arr)
        func_emperic(i) = sum(X <= t_arr(i)) / n;
    end

    figure;
    stairs(t_arr, func_emperic, 'LineWidth', 1);
    hold on;

    x_coords = (Mmin - 1) : 1e-3 : (Mmax + 1);
    func_norm = normcdf(x_coords, Mu, sqrt(S_square));
    plot(x_coords, func_norm, 'LineWidth', 1);
    grid;
    legend('Эмпирическая ФР', 'Нормальная ФР N(\mu, \sigma^2)');
end
```

5 Результаты работы программы

```
а) Вычисление максимального значения Mmax и минимального значения Mmin

Mmax = 4.300
Mmin  = -0.730

б) Вычисление размаха R

R = 5.030

в) Вычисление оценок Mu и S^2 математического ожидания MX и дисперсии DX

Mu = 1.836
S^2 = 1.153

г) Группировка значений выборки в m = [log2 n] + 2 интервала

Кол-во интервалов m =   8:

 1. [-0.730; -0.101), кол-во элементов: 3
 2. [-0.101; 0.527), кол-во элементов: 13
 3. [0.527; 1.156), кол-во элементов: 14
 4. [1.156; 1.785), кол-во элементов: 27
 5. [1.785; 2.414), кол-во элементов: 25
 6. [2.414; 3.042), кол-во элементов: 24
 7. [3.042; 3.671), кол-во элементов: 9
 8. [3.671; 4.300], кол-во элементов: 5
```

Рисунок 5.1 – Результаты расчетов для выборки из 10 варианта

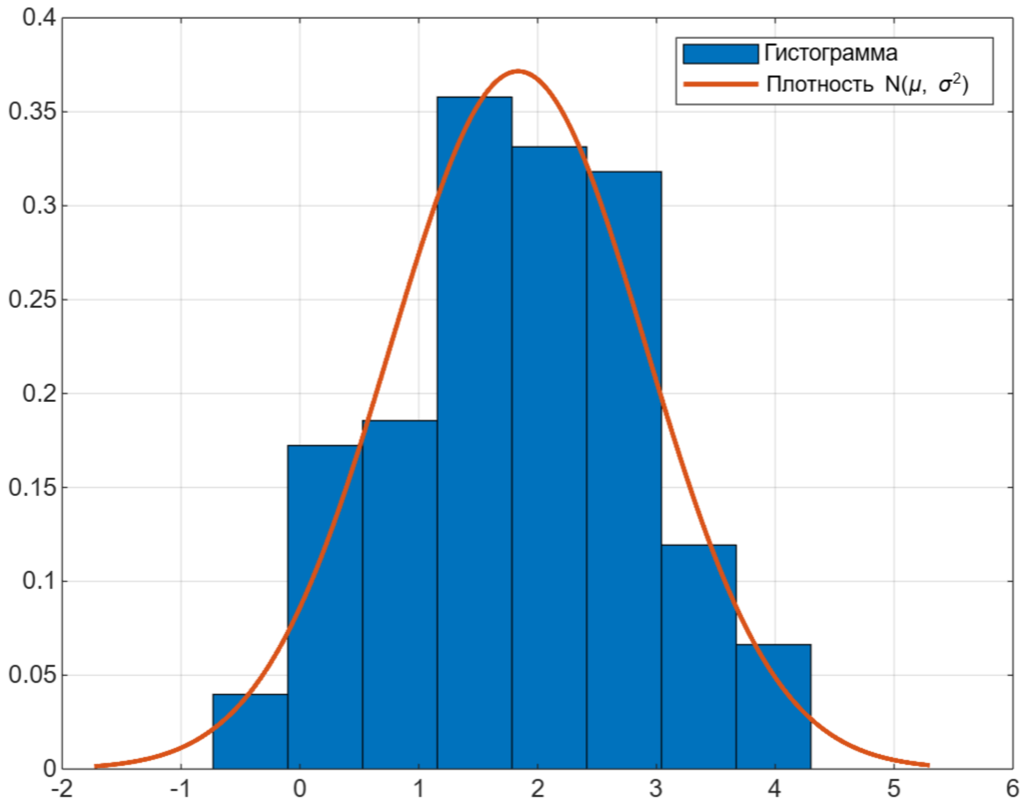

Рисунок 5.2 – Гистограмма и график функции плотности распределения нормальной случайной величины с выборочными математическим ожиданием и дисперсией

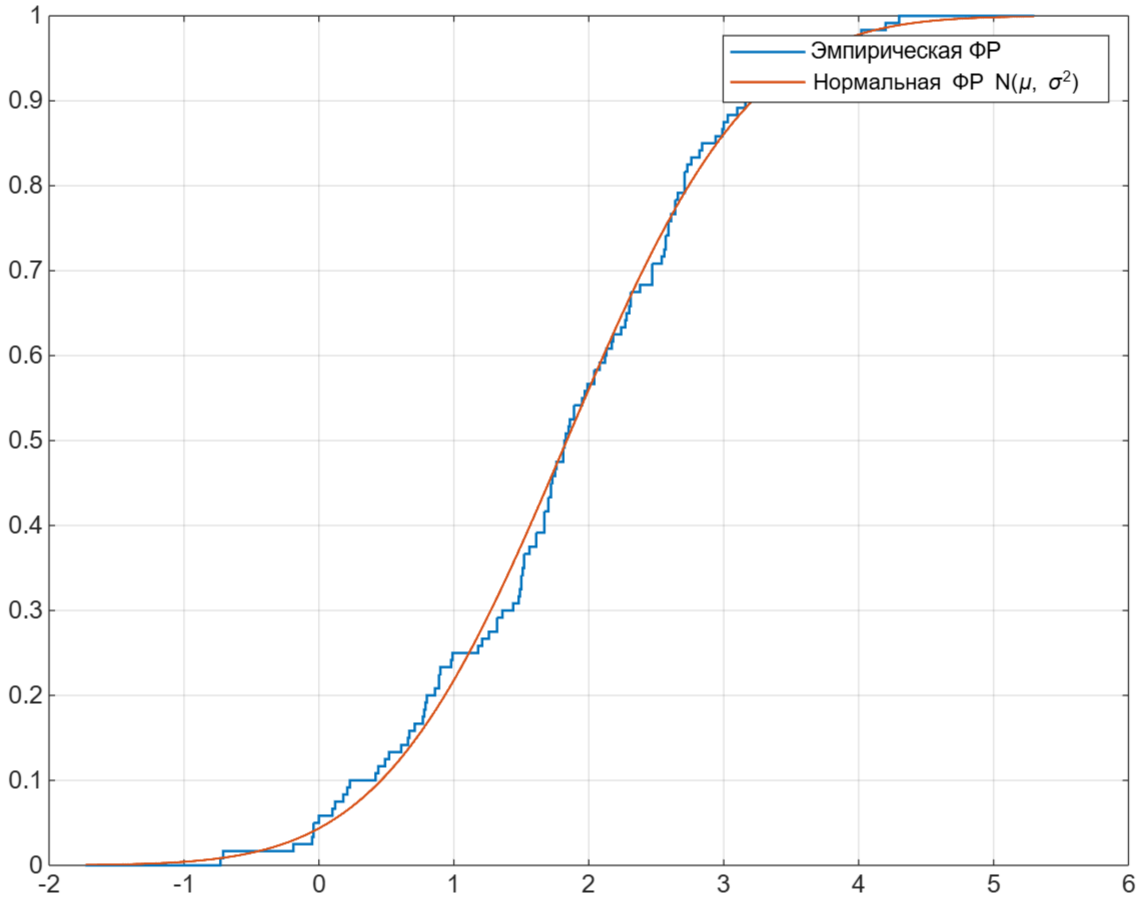

Рисунок 5.3 – График эмперической функции распределения и функции распределения нормальной случайной величины с выборочными математическим ожиданием и дисперсией