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»

на тему: «Гистограмма и эмпирическая функция распределения»

Вариант № 1

Студент ИУ7-63Б (Группа) ________ (Подпись, дата) Авдейкина В. П. (И. О. Фамилия)

Преподаватель ________ (Подпись, дата) Власов П. А. (И. О. Фамилия)

2024 г.

СОДЕРЖАНИЕ

1 Теоретические сведения

1.1 Формулы

Максимальное значение:

$$M_{max} = x_{(1)} = \max_{x_i \in \vec{x}} x_i. \tag{1.1}$$

Минимальное значение:

$$M_{min} = x_{(n)} = \min_{x_i \in \vec{x}} x_i. \tag{1.2}$$

Размах выборки:

$$R = M_{max} - M_{min} \tag{1.3}$$

Оценка математического ожидания (выборочное среднее):

$$\hat{\mu}(\vec{x}) = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1.4}$$

Исправленная оценка дисперсии (исправленная выборочная дисперсия):

$$S^2(\vec{x}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu}(\vec{x}))^2 \tag{1.5}$$

1.2 Определения

Пусть $\vec{x} = (x_1, \ldots, x_n)$ — реализация случайной выборки $\vec{X} = (X_1, \ldots, X_n)$ из генеральной совокупности X. Пусть также $m \in \mathbb{N}$ — количество интервалов, а отрезок $J = [x_{(1)};\ x_{(n)}]$ разбивают на m равновеликих промежутков длины

$$\Delta = \frac{x_{(n)} - x_{(1)}}{m}. \tag{1.6}$$

Интервалы определяются равенствами 1.7, 1.8:

$$J_i = [x_{(1)} + (i-1)\Delta;\ x_{(i)} + i\Delta), \quad i = 1, \ldots, m-1; \tag{1.7}$$

$$J_m = [x_{(1)} + (m-1)\Delta;\ x_{(i)} + m\Delta). \tag{1.8}$$

Интервальным статистическим рядом называют таблицу 1.9

$$\begin{array}{|c|c|c|} \hline J_1 & \ldots & J_m \\ \hline n_1 & \ldots & n_m \\ \hline \end{array} \tag{1.9}$$

где n_i — число элементов в реализации $\vec{x} = (x_1, \ldots, x_n)$, попавших в промежуток J_i, $i = 1, \ldots, m$

Предположим, что для данной реализации $\vec{x}$ построен интервальный статистический ряд.

Эмпирической плотностью распределения называют функцию:

$$f_n(t) = \begin{cases} \frac{n_i}{n \cdot \Delta}, & t \in J_i, \quad i = 1, \ldots, m, \\ 0, & t \notin J. \end{cases} \tag{1.10}$$

График функции $f_n(t)$ называют *гистограммой.*

Пусть $\vec{x} = (x_1, \ldots, x_n)$ — реализация случайной выборки $\vec{X} = (X_1, \ldots, X_n)$ из генеральной совокупности X. Обозначим с помощью $l(t, \vec{x})$ число элементов $\vec{x}$, которые меньше t ($t \in \mathbb{R}$).

Эмпирической функцией распределения называют отображение

$$F_n : \mathbb{R} \to \mathbb{R}, \tag{1.11}$$

заданное формулой

$$F_n(t) = \frac{l(t, \vec{x})}{n}. \tag{1.12}$$

2 Текст программы

На листингах 2.1, 2.2, 2.2, 2.4 представлена реализованная в ходе лабораторной работы программа (на языке Octave).

Листинг 2.1 – Исходный текст программы (часть 1)

```
warning("off", "Octave:data-file-in-path");
X = csvread('x.csv');

X = sort(X);
n = length(X);

M_max = max(X);
M_min = min(X);

fprintf("a) вычисление максимального значения M_max и\
        минимального значения M_min:\n");
fprintf("\tM_max = %+.4f\n\tM_min = %+.4f\n", M_max, M_min);

R = M_max - M_min;

fprintf("\nб) вычисление размаха R выборки:\n");
fprintf("\tR = %.4f\n", R);

Mu = sum(X) / n;
S_quad = sum((X - Mu) .^2) / (n - 1);
S = sqrt(S_quad);

fprintf("\nв) вычисление оценок Mu и S_quad математического\
        ожидания MX и дисперсии DX:\n");
fprintf("\tMu = %.4f\n\tS_quad = %.4f\n", Mu, S_quad);

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

m = floor(log2(n)) + 2;
fprintf("\tКоличество интервалов m = %d\n\n", m);
delta = (X(n) - X(1)) / m;
J_limits = M_min : delta : M_max;
ni = zeros(m, 1);
```

Листинг 2.2 – Исходный текст программы (часть 2)

```
for i = 1 : m
    count = 0;

    for x = X
        if (i == m) && (x >= J_limits(i))
           && (x <= J_limits(i + 1))
            count++;
        elseif (x >= J_limits(i)) && (x < J_limits(i + 1))
            count++;
        endif
    endfor

    if (i == m)
        fprintf("\t%d) [%+.3f; %+.3f), n%d = %d\n", i,
                J_limits(i), J_limits(i + 1), i, count);
    else
        fprintf("\t%d) [%+.3f; %+.3f], n%d = %d\n", i,
                J_limits(i), J_limits(i + 1), i, count);
    endif

    ni(i) = count;
endfor

fprintf("\nд) построение на одной координатной плоскости \
        гистограммы\ \n   и графика функции плотности расп\
        редления вероятностей \n   нормальной случайной ве\
        личины с математическим\ \n   ожиданием Mu и диспер\
        сией S_quad\n");

J_middles = zeros(m, 1);

for i = 1 : m
    J_middles(i) = (J_limits(i) + J_limits(i + 1)) / 2;
endfor

fn_values = zeros(m, 1);

for i = 1 : m
    fn_values(i) = ni(i) / (n * delta);
endfor
```

Листинг 2.3 – Исходный текст программы (часть 3)

```
step = S / 1000;
x_coords = (M_min - R) : step : (M_max + R);
f_density_normal = normpdf(x_coords, Mu, S);

figure
hold on;
bar(J_middles, fn_values, 1, 'y')
plot(x_coords, f_density_normal, 'b', 'LineWidth', 2);
grid on;
hold off;

fprintf("\ne) построение на другой координатной плоскости\
        графика \n   эмпирической функции распределения и\
        функции \n   распределения нормальной случайной ве\
        личины с \n   математическим ожиданием Mu и диспер\
        сией S_quad\n");

t_arr = zeros(n + 2, 1);
t_arr(1) = X(1) - 1;
t_arr(n + 2) = X(n) + 1;

for i = 2 : n + 1
    t_arr(i) = X(i - 1);
endfor

f_emp = zeros(length(t_arr), 1);

for i = 1 : length(t_arr)
    count = 0;

    for j = 1 : n
        if (t_arr(i) >= X(j))
            count++;
        end
    endfor

    f_emp(i) = count / n;
endfor
```

Листинг 2.4 – Исходный текст программы (часть 4)

```
xs = (M_min - R) : step : (M_max + R);
f_norm = normcdf(x_coords, Mu, S);

figure
hold on;
plot(xs, f_norm, ':r', 'linewidth', 2);
stairs(t_arr, f_emp, 'b', 'linewidth', 2);
grid on;
hold off;
```

3 Результаты расчетов для выборки из индивидуального варианта

Результаты расчетов, выполненных с помощью реализованной программы для выборки из индивидуального варианта №1, представлены в формулах 3.1, 3.2, 3.3, 3.4, 3.5, 3.6.

$$M_max = 1.4 \tag{3.1}$$

$$M_min = -4.11 \tag{3.2}$$

$$R = 5.51 \tag{3.3}$$

$$\hat{\mu}(\vec{x}) = -1.6046 \tag{3.4}$$

$$S^2(\vec{x}) = 1.0341 \tag{3.5}$$

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

Результат группировки значений выборки в $m = 8$ интервалов представлен на рисунке 3.1.

```
г) группировка значений выборки в m = [log2(n)] + 2 интервала:
       Количество интервалов m = 8

       1) [-4.110; -3.421], n1 = 4
       2) [-3.421; -2.733], n2 = 11
       3) [-2.733; -2.044], n3 = 26
       4) [-2.044; -1.355], n4 = 33
       5) [-1.355; -0.666], n5 = 26
       6) [-0.666; +0.022], n6 = 15
       7) [+0.022; +0.711], n7 = 2
       8) [+0.711; +1.400), n8 = 2
```

Рисунок 3.1 — Результат группировки значений выборки

На рисунке 3.2 представлены гистограмма и график функции плотности распредления нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.

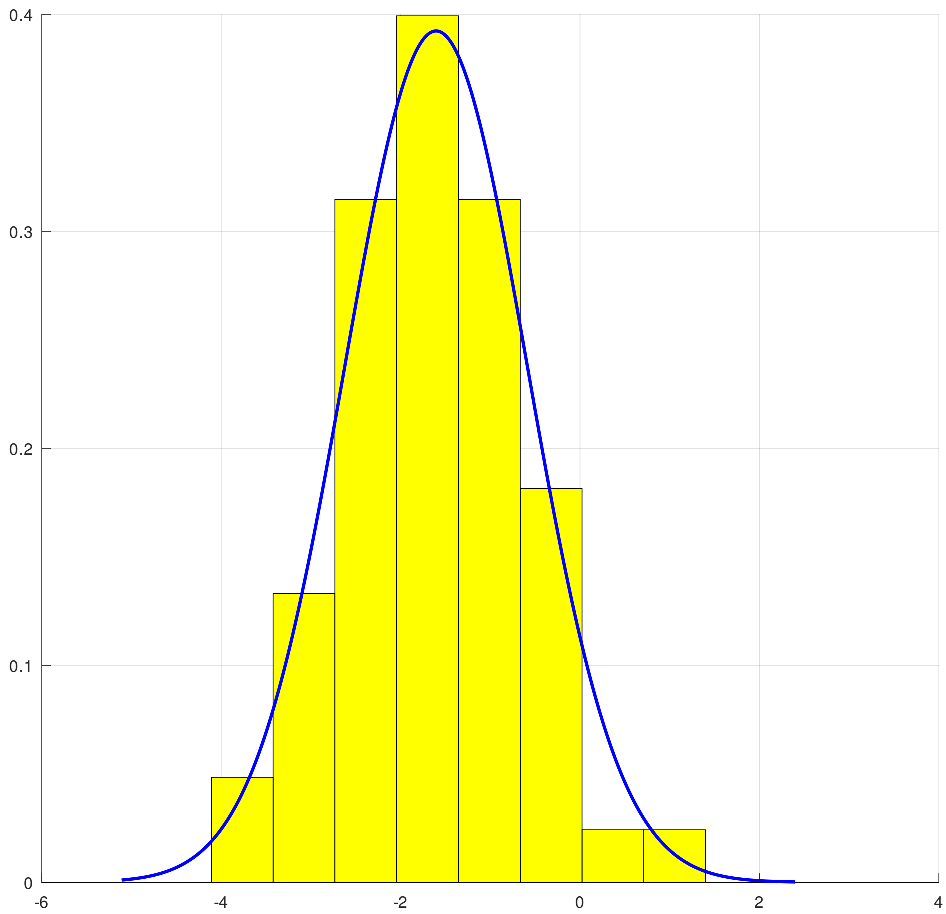

Рисунок 3.2 — Гистограмма и график функции плотности распредления нормальной случайной величины

На рисунке 3.3 представлены графики эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.

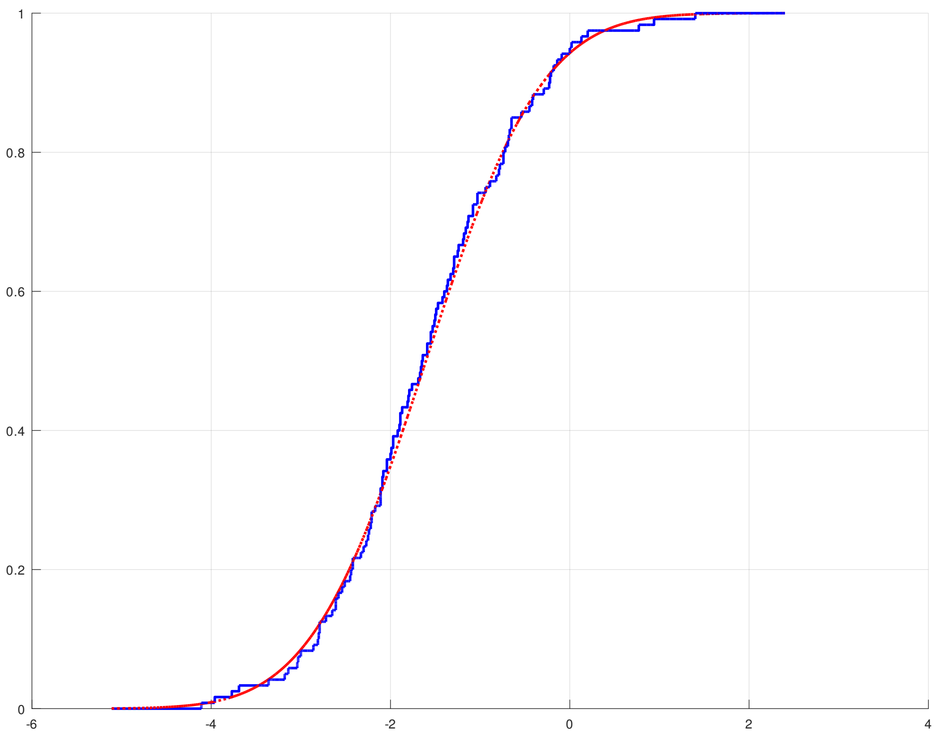

Рисунок 3.3 — Графики эмпирической функции распределения и функции распределения нормальной случайной величины