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»

Лабораторная работа №1 по курсу "Математическая статистика"

Тема Гистограмма и эмпирическая функция распределения

Студент Ковалец К. Э.

Группа ИУ7-63Б

Вариант 9

Преподаватель Власов П. А.

Москва — 2022 г.

1 Содержание работы

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

   - вычисление максимального значения M_{max} и минимального значения M_{min};
   - размаха R выборки;
   - вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;
   - группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
   - построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
   - 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:

$$M_{\max} = X_{(n)} \tag{2.1}$$

- Минимальное M_{min} значение выборки:

$$M_{\min} = X_{(1)} \tag{2.2}$$

- Размах R выборки:

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

- Выборочное среднее (оценка математического ожидания):

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

- Исправленная выборочная дисперсия:

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

3 Определение эмпирической плотности, гистограммы и эмпирической функции распределения

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.

Ширина каждого из них:

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}, \tag{3.1}$$

где $m = [\log_2 n] + 2$, $x_{(1)} = min(\vec{x})$, $x_{(n)} = max(\vec{x})$.

Далее полагают m:

$$J_i = [x_{(1)} + (i-1)\cdot\Delta, x_{(1)} + i\cdot\Delta), i = \overline{1; m-1} \tag{3.2}$$

$$J_m = [x_{(1)} + (m-1)\cdot\Delta, x_{(n)}] \tag{3.3}$$

3.1 Интервальный статистический ряд

Опр. Интервальным статистическим рядом, отвечающим выборке $\vec{x}$, называется таблица вида:

J_1	J_2	...	J_m
n_1	n_2	...	n_m

Здесь n_i — число элементоав выборки $\vec{x}$, попавших в промежуток J_i, $i = \overline{1, m}$.

3.2 Эмпирическая плотность

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

Опр. Эмпирической плотностью распределния (соответствующей выборке $\vec{x}$) называется функция:

$$f_n(x) = \begin{cases} \frac{n_i}{n \cdot \Delta}, x \in J_i, i = \overline{1; m} \\ 0, \text{иначе} \end{cases} \tag{3.4}$$

3.3 Гистограмма

Опр. График эмпирической функции плотности называется гистограммой.

3.4 Эмпирическая функция распределения

Пусть $\vec{x} = (x_1, ..., x_n)$ — выборка из генеральной совокупности X.
Обозначим $n(t, \vec{x})$ — число компонент вектора $\vec{x}$, которые меньше, чем t.

Опр. Эмпирической функцией распределения, построенной по выборке $\vec{x}$, называют функцию

$$F_n : R \to R, \tag{3.5}$$

определенную правилом:

$$F_n(x) = \frac{n(t, \vec{x})}{n}. \tag{3.6}$$

4 Текст программы

```
function lab_1()
    % Выборка объема n из генеральной совокупности X
    X = csvread('X.csv');

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

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

    Mmax = max(X);
    Mmin = min(X);

    fprintf("\nMmax = %.4f\n", Mmax);
    fprintf("\nMmin  = %.4f\n", Mmin);

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

    R = Mmax - Mmin;
    fprintf("\nR = %.4f\n", R);

    fprintf("\nc) Вычисление оценок Mu и S_quad математического ожидания MX и дисперсии DX\n");

    Mu = sum(X) / n; % Выборочное среднее
    S_quad = sum((X - Mu) .^2) / (n - 1); % Исправленная выборочная дисперсия

    fprintf("\nMu = %.4f\n", Mu);
    fprintf("\nS_quad = %.4f\n", S_quad);



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

    % Поиск количества интервалов
    % floor(X) - возвращает значения, округленные до ближайшего целого <= X
    m = floor(log2(n)) + 2;
    fprintf("Кол-во интервалов m = %3d\n\n", m);

    % Ширина интервала
    delta = (X(n) - X(1)) / m;

    % Нахождение границ интервалов
    borders = Mmin : delta : Mmax;
```

```
    % Массив с кол-вом элементов выборки, попавших в i-ый промежуток
    ni_arr = zeros(m, 1);

    for i = 1 : (m)
        count = 0; % Кол-во значений в i-том интервале

        for x = X
            % Последний интервал включает в себя крайнее правое значение
            if (i == m) && (x >= borders(i)) && (x <= borders(i + 1))
                count = count + 1;
            % Остальные интервалы включают только слева
            elseif (x >= borders(i)) && (x < borders(i + 1))
                count = count + 1;
            endif
        endfor

        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);
        endif

        ni_arr(i) = count;
    endfor



    fprintf("\ne) Построение на одной координатной плоскости гистограммы\
             \n    и графика функции плотности распределния вероятностей\
             \n    нормальной случайной величины с математическим\
             \n    ожиданием Mu и дисперсие S_quad\n\n");

    % Гистограмма

    mid_intervals = zeros(m, 1);

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

    column_values = zeros(m, 1);

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

```
% Отрисовка гистограммы
bar(mid_intervals, column_values, 1, 'b');
hold on;

% График функции плотности нормального распределения

% Набор значений
x_coords = (Mmin - 1) : 1e-3 : (Mmax + 1);

% normpdf - функция плотности нормального распределения
func_density_norm = normpdf(x_coords, Mu, sqrt(S_quad));

% Отрисовка графика плоности нормального распределения
plot(x_coords, func_density_norm, 'r', 'LineWidth', 2);
grid;



fprintf("\nd) Построение на другой координатной плоскости графика\
         \n   эмпирической функции распределения и функции\
         \n   распределения нормальной случайной величины с\
         \n   математическим ожиданием Mu и S_quad\n\n");

% Эмпирической функции распределния

t_arr = zeros(1, n + 2);

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

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

% Значения эмпирической функции распреления
func_emperic = zeros(length(t_arr), 1);

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

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

    func_emperic(i) = count / n;
```

```
    endfor

    figure();

    % Отрисовка эмпирической функции распределения
    stairs(t_arr, func_emperic, 'b', 'LineWidth', 1);
    hold on;

    % График функции нормального распределения

    % Набор значений
    x_coords = (Mmin - 1) : 1e-3 : (Mmax + 1);

    % normcdf - функция нормального распределения
    func_norm = normcdf(x_coords, Mu, sqrt(S_quad));

    % Отрисовка графика нормального распределения
    plot(x_coords, func_norm, 'r', 'LineWidth', 1);
    grid;

endfunction
```

5 Результат работы программы

```
>> lab_1

a) Вычисление максимального значения Mmax и минимального значения Mmin

Mmax = -5.2000

Mmin  = -10.1100

b) Вычисление размаха R

R = 4.9100

c) Вычисление оценок Mu и S_quad математического ожидания MX и дисперсии DX

Mu = -7.6609

S_quad = 0.7779

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

Кол-во интервалов m =   8

 1. [-10.110; -9.496), кол-во элементов: 1
 2. [-9.496; -8.883), кол-во элементов: 10
 3. [-8.883; -8.269), кол-во элементов: 18
 4. [-8.269; -7.655), кол-во элементов: 32
 5. [-7.655; -7.041), кол-во элементов: 30
 6. [-7.041; -6.428), кол-во элементов: 18
 7. [-6.428; -5.814), кол-во элементов: 10
 8. [-5.814; -5.200], кол-во элементов: 1

e) Построение на одной координатной плоскости гистограммы
   и графика функции плотности распределния вероятностей
   нормальной случайной величины с математическим
   ожиданием Mu и дисперсие S_quad


d) Построение на другой координатной плоскости графика
   эмпирической функции распределения и функции
   распределения нормальной случайной величины с
   математическим ожиданием Mu и S_quad
```

Рисунок 5.1 – Результаты расчетов для выборки из индивидуального варианта

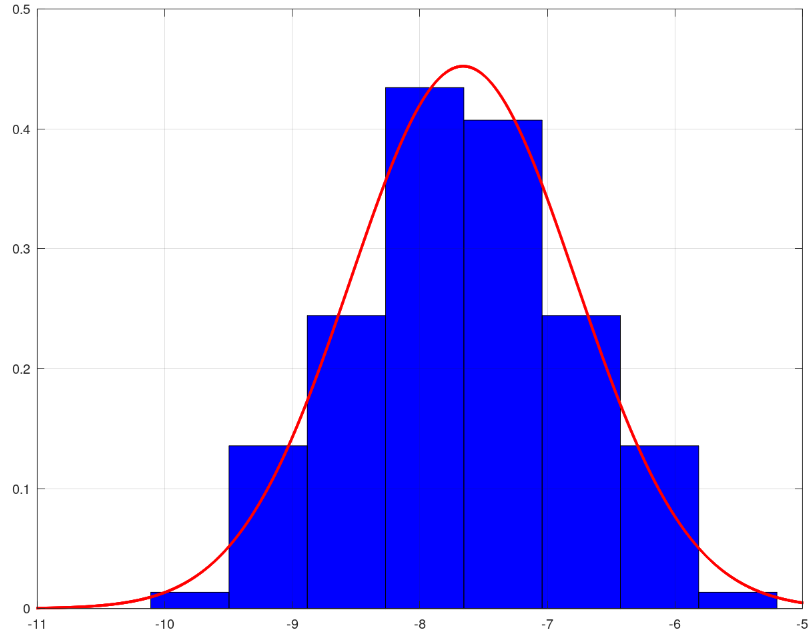

Рисунок 5.2 – Гистограмма и график функции плотности распределения нормальной случайной величины с выборочными математическим ожиданием и дисперсией

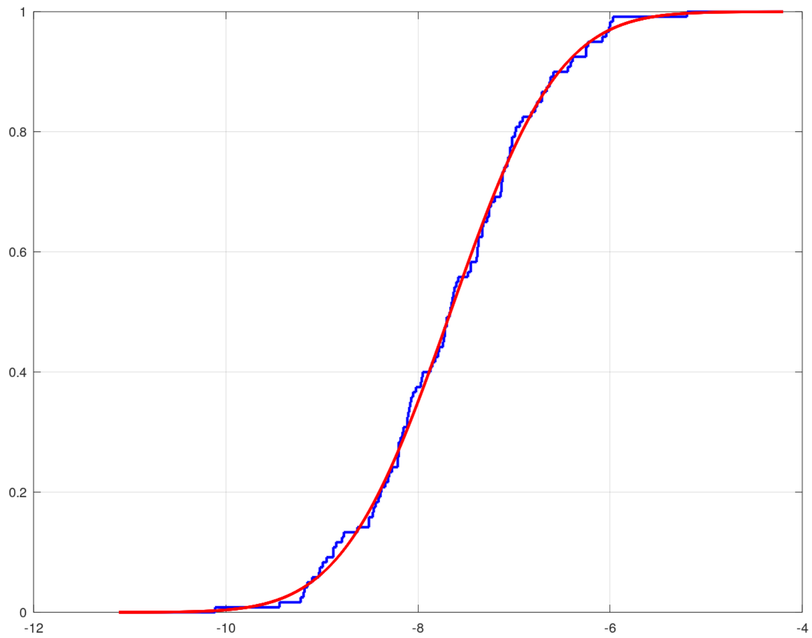

Рисунок 5.3 – График эмпирической функции распределения и функции распределения нормальной случайной величины с выборочными математическим ожиданием и дисперсией