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"

Тема Гистограмма и эмпирическая функция распределения

Студент Цветков И.А.

Группа ИУ7-63Б

Оценка (баллы) ____________________

Преподаватель Власов П. А.

Москва — 2022 г.

1 Задание на лабораторную работу

Цель работы: построение гистограммы и эмпирической функции распределения.

1.1 Содержание работы

1. Для выборки объема 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

Максимальное M_{max} и минимальное M_{min} значения выборки:

$$\begin{aligned} M_{\max} &= X_{(n)} \\ M_{\min} &= X_{(1)} \end{aligned} \tag{2.1}$$

Размах R выборки:

$$R = M_{\max} - M_{\min}. \tag{2.2}$$

Оценки математического ожидания и дисперсии:

- выборочное среднее:

$$\hat{\mu}(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.3}$$

- несмещенная оценка дисперсии:

$$S^2(\vec{X}_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \tag{2.4}$$

2.2 Определение эмпирической плотности и гистограммы

Интервальный статистический ряд

Определение: 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 = $\overline{1, m}$, где

$$J_i = [x_{(1)} + (i-1) \cdot \Delta, x_{(1)} + i \cdot \Delta), i = \overline{1; m-1} \tag{2.5}$$

Причем:

$$J_m = [x_{(1)} + (m-1) \cdot \Delta, x_{(n)}] \tag{2.6}$$

Величина Δ при этом равна:

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m} \tag{2.7}$$

Эмперическая плотность

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

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

$$\hat{f}_n(x) = \begin{cases} \frac{n_i}{n \cdot \Delta}, x \in J_i, i = \overline{1; m} \\ 0, \text{иначе} \end{cases} \tag{2.8}$$

Гистограмма

Определение: График эмпирической функции плотности называется гистограммой.

2.3 Определение эмпирической функции распределения

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

Обозначим $n(t, \vec{x})$ — числа компонент вектора $\vec{x}$, которые меньше, чем t.

Определение: Эмпирической функцией распределения, построенной по выборке $\vec{x}$, называют функцию

$$F_n : R \to R, \tag{2.9}$$

определенную правилом:

$$F_n(x) = \frac{n(t, \vec{x})}{n} \tag{2.10}$$

3 Практическая часть

3.1 Код програмы

```
function lab_01()
    X = csvread('data21.csv');

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

    fprintf("\nn = %g\n", n);

    fprintf("\nВыполнение заданий.\n\n");

    % Задание 1
    fprintf("\n1) Вычислить максимальное и минимальное значение:\n", n);

    % Минимальное значение
    minX = min(X);
    fprintf("\nМинимальное значение  = %.4f", minX);

    % Максимальное значение
    maxX = max(X);
    fprintf("\nМаксимальное значение = %.4f\n", maxX);


    % Задание 2
    fprintf("\n2) Вычислить размах R:\n", n);

    % Размах R
    R = maxX - minX;
    fprintf("\nR = %.4f\n", R);


    % Задание 3
    fprintf("\n3) Вычислить оценки математического ожидания и дисперсии:\n", n);
    % Выборочное среднее
    mu = sum(X) / n;
    fprintf("\nВыборочное среднее (оценка мат ожидания) = %.4f\n", mu);

    % Исправленная дисперсия
    s_quad = sum((X - mu) .^2) / (n - 1);
    fprintf("Исправленная дисперсия (оценка дисперсии) = %.4f\n", s_quad);


    % Задание 4
    fprintf("\n4) Группировка значений выборки в m = [log2 n] + 2 интервала:\n\n", n);

    % Вычисление m = [log2 n] + 2 промежутков
    m = floor(log2(n)) + 2;
    fprintf("Интервалов m = %3d\n\n", m);

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

    % Поиск интервалов
    % Интервалы хранятся в виде матрицы m на 2
    intervals = zeros(m, 3);

    % Вычисление границ интервалов
    for ind = 1 : m - 1
        intervals(ind, 1) = X(1) + (ind - 1) * delta;
        intervals(ind, 2) = X(1) + ind * delta;
    endfor

    % Вычисление границ интервалов для m
    intervals(m, 1) = X(1) + (m - 1) * delta;
    intervals(m, 2) = X(n);

    % Вычисление кол-ва значений в интервалах
    indX = 1;
    indIntervals = 1;

    %% Задаются текущие границы интервала
    leftBorder = intervals(indIntervals, 1);
    rightBorder = intervals(indIntervals, 2);
```

```

    %% Цикл по всем элементам выборки
    while indX <= n

        %% Если элемент попал в границы, то инкремент для данного интервала
        if X(indX) >= leftBorder && X(indX) < rightBorder
            intervals(indIntervals, 3) = intervals(indIntervals, 3) + 1;
        %% Если элемент не попал в интервал
        else
            %% Элемент, не попавший в предыдущий интервал, должен быть
            %% обработан для следующего интервала
            indX -= 1;

            %% Для крайнего элемента выборки, который должен попасть в текущий интервал
            if indIntervals + 1 > m
                intervals(indIntervals, 3) = intervals(indIntervals, 3) + 1;
                break;
            endif

            %% Следующий интервал
            indIntervals = indIntervals + 1;

            %% Обновление границ текущего интервала
            leftBorder = rightBorder;
            rightBorder = intervals(indIntervals, 2);

        endif

        %% Следующий элемент выборки
        indX = indX + 1;
    endwhile

   % Вывести полученные интервалы
    for i = 1 : m
        if i == m
            fprintf("%3d. [ %3.4f ; %3.4f] : Элементов: %3d\n", i,
                                                                    intervals(i, 1),
                                                                    intervals(i, 2),
                                                                    intervals(i, 3));
        else
            fprintf("%3d. [ %3.4f ; %3.4f) : Элементов: %3d\n", i,
                                                                    intervals(i, 1),
                                                                    intervals(i, 2),
                                                                    intervals(i, 3));
        endif
    endfor


    % Задание 5
    fprintf("\n5) Построить гистограмму и график функции плотности \
             \n распределния вероятностей нормальной случайной величины с \
             \n математическим ожиданием mu и s_quad\n\n");

    fprintf("Результат в отдельном окне\n");


    % Построение гистограммы

    gistData = zeros(m, 2);

    % Середины интервалов
    for ind = 1 : m
        gistData(ind, 1) = (intervals(ind, 1) + intervals(ind, 2)) / 2;
    endfor

    % Значения столбцов
    % k = ni / (n * delta)
    for ind = 1 : m
        gistData(ind, 2) = intervals(ind, 3) / (n * delta);
    endfor

    % График функции плотности распределения

    % Набор значений по X
    % начало:шаг:конец
    xGraph = (minX - 1):1e-3:(maxX + 1);

    sigma = sqrt(s_quad);
    % Значения функции плотности распределения
    % normpdf - функция плотности нормального распределения
    fNormal = normpdf(xGraph, mu, sigma);

```

```
    % Отрисовка графиков
    %% Отрисовка гистограммы
    bar(gistData(:, 1), gistData(:, 2), 1);

    %% Чтобы следующий график не стер предыдущий
    hold on;

    %% Отрисовка графика плоности нормального распределения
    plot(xGraph, fNormal, 'r', 'LineWidth', 1.5);
    %% Сетка
    grid;


    % Задание 6
    fprintf("\n6) Построить график эмпирической функции распределения и\
             \n функции распределения нормальной случайной величины с \
             \n математическим ожиданием mu и s_quad\n\n");

    fprintf("Результат в отдельном окне\n");

    % Набор значений для эмпирической функции распределния
    tSet = zeros(1, n + 2);


    tSet(1) = X(1) - 1;

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

    tSet(n + 2) = X(n) + 1;

    % Значения эмпирической функции распреления
    nEmperic = length(tSet);
    Femperic = zeros(nEmperic, 1);

    for i = 1 : nEmperic
        cnt = 0;

        for j = 1: n
            % x(j - 1) < t <= x(j)
            if X(j) <= tSet(i)
                cnt = cnt + 1;
            endif
        endfor

        Femperic(i) = cnt / n;
    endfor


    % Набор значений по X
    % начало:шаг:конец
    xGraph = (minX - 1):1e-3:(maxX + 1);

    sigma = sqrt(s_quad);
    % Значения функции распределения
    % normpdf - функция нормального распределения
    Fnormal = normcdf(xGraph, mu, sigma);


    % Отрисовка графиков
    %% Чтобы графики были в отдельном окне
    figure();
    %% Отрисовка эмпирической функции распределения
    stairs(tSet, Femperic);

    %% Чтобы следующий график не стер предыдущий
    hold on;

    %% Отрисовка графика нормального распределения
    plot(xGraph, Fnormal, 'r', 'LineWidth', 1.5);
    %% Сетка
    grid;
endfunction
```

3.2 Результат работы программы

Вывод программы

```
Выполнение заданий.


 1) Вычислить максимальное и минимальное значение:

Минимальное значение  = −17.2900
Максимальное значение = −12.9100

 2) Вычислить размах R:

R = 4.3800

 3) Вычислить оценки математического ожидания и дисперсии:

Выборочное среднее (оценка мат ожидания) = −15.2209
Исправленная дисперсия (оценка дисперсии) = 0.9680

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

Интервалов m =    8

     1. [  −17.2900 ; −16.7425) : Элементов:    9
     2. [  −16.7425 ; −16.1950) : Элементов:   13
     3. [  −16.1950 ; −15.6475) : Элементов:   19
     4. [  −15.6475 ; −15.1000) : Элементов:   22
     5. [  −15.1000 ; −14.5525) : Элементов:   21
     6. [  −14.5525 ; −14.0050) : Элементов:   23
     7. [  −14.0050 ; −13.4575) : Элементов:   11
     8. [  −13.4575 ; −12.9100] : Элементов:    2

 5) Построить гистограмму и график функции плотности
     распределния вероятностей нормальной случайной величины с
     математическим ожиданием mu и s_quad

Результат в отдельном окне

 6) Построить график эмпирической функции распределения и
     функции распределения нормальной случайной величины с
     математическим ожиданием mu и s_quad

Результат в отдельном окне
```

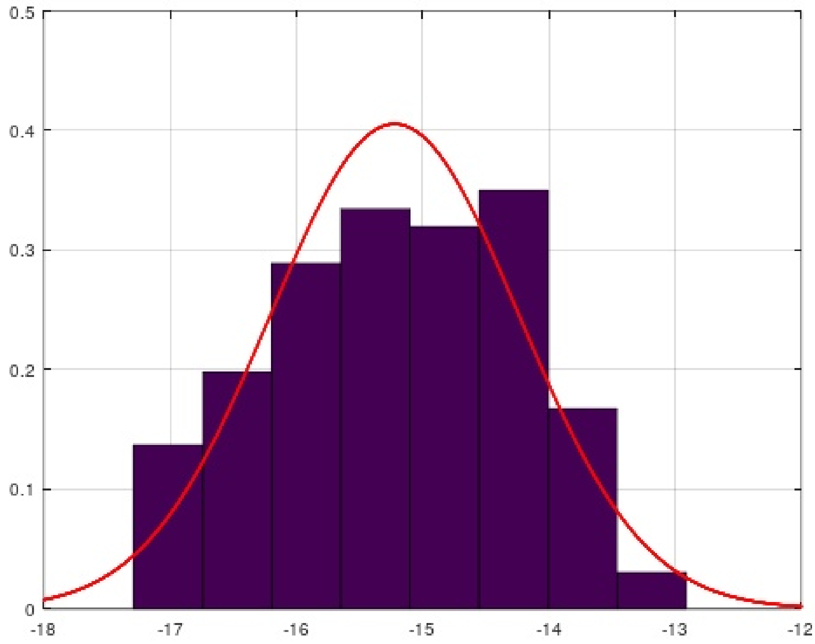

Рисунок 3.1 – Гистограмма и график функции плотности распределения нормальной случайной величины с выборочными математическим ожиданием и дисперсией

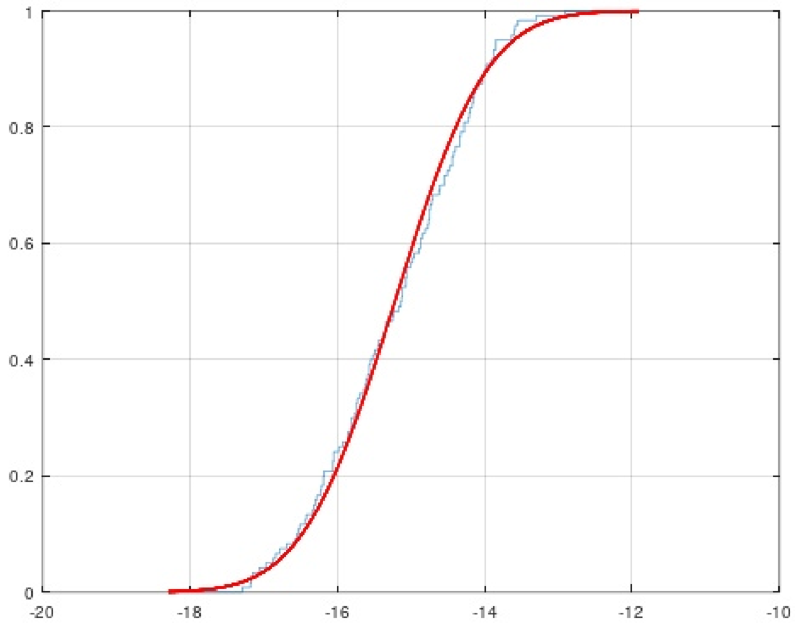

Рисунок 3.2 – График эмперической функции распределения и функции распределения нормальной случайной величины с выборочными математическим ожиданием и дисперсией