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 дисциплине "Математическая статистика"

Тема Гистограмма и эмпирическая функция распределения

Студент Светличная А.А.

Группа ИУ7-63Б

Преподаватель Власов П.А.

Москва — 2023 г.

1 Задание на лабораторную работу

Цель работы: построение гистограммы и эмпирической функции распределения.

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 $\mathrm{M}X$ и дисперсии $\mathrm{D}X$;

   – 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.

Содержание отчета

1. формулы для вычисления величин M_{max}, M_{min}, R, $\hat{\mu}$, S^2;

2. определение эмпирической плотности и гистограммы;

3. определение эмпирической функции распределения;

4. текст программы;

5. результаты расчетов для выборки из индивидуального варианта.

2 Теоретичсеская часть

2.1 Формулы для вычисления величин

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

Максимальное и минимальное значения выборки:

$$M_{max} = x_{(n)} = \max\{x_1, ..., x_n\}, \tag{2.1}$$

$$M_{min} = x_{(1)} = \min\{x_1, ..., x_n\}, \tag{2.2}$$

где $x_{(1)}, x_{(n)}$ — крайние члены вариационного ряда, отвечающего выборке $\vec{x}$.

Размах выборки:

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

Оценка математического ожидания — выборочное среднее:

$$\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{2.4}$$

Оценка дисперсии — исправленная выборочная дисперсия:

$$S^2(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2, \tag{2.5}$$

где $\overline{x} = \hat{\mu}$.

2.2 Определение эмпирической плотности и гистограммы

Пусть $\vec{x} = (x_1, ..., x_n)$ — выборка объема из генеральной совокупности X. Если объем этой выборки n достаточно велик, то значения x_i группируют в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ разбивают на m равновеликих промежутков.

Для выбора m используют формулу:

$$m = [\log_2 n] + 2, \tag{2.6}$$

где n — размер выборки.

Ширина каждого из них:

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

Полагают

$$J_i = [x_{(1)} + (i-1)\cdot\Delta, x_{(1)} + i\cdot\Delta), i = \overline{1; m-1} \tag{2.8}$$

$$J_m = [x_{(1)} + (m-1)\cdot\Delta, x_{(n)}] \tag{2.9}$$

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

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

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

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

Эмпирической плотностью распределения, соответствующей выборке $\vec{x}$, называется функция:

$$f_n(x) = \begin{cases} \frac{n_i}{n\Delta}, \text{если } x \in J_i, \\ 0, \text{иначе.} \end{cases} \tag{2.10}$$

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

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

$$F_n : R \to R, \tag{2.11}$$

определенную правилом

$$F_n(t) = \frac{n(t, \vec{x})}{n} \tag{2.12}$$

3 Практическая часть

3.1 Текст программы

```
    X = [12.80,10.91,10.69,10.46,13.08,11.35,13.07,12.46,13.29,11.09,...
        10.34,12.57,12.99,12.28,12.62,12.43,12.06,13.72,11.64,11.81,...
        10.47,13.94,13.31,13.31,11.11,11.06,11.84,12.28,13.14,9.80,...
        11.27,12.56,11.15,11.65,11.92,11.31,12.07,12.06,11.53,12.04,...
        12.44,13.89,10.72,11.75,11.14,12.70,13.64,12.64,11.37,10.12,...
        11.87,13.86,12.53,12.52,10.65,9.84,11.15,9.54,11.39,11.57,...
        14.00,10.98,10.81,12.11,12.24,10.42,15.21,13.00,11.63,11.40,...
        13.41,10.78,11.98,12.51,12.95,11.59,11.04,10.99,11.91,10.72,...
        11.74,11.82,12.45,11.70,12.49,12.10,10.99,12.58,12.99,10.09,...
        12.00,13.19,12.89,11.72,13.80,11.37,12.15,10.98,11.93,10.83,...
        12.41,12.52,12.21,12.82,12.38,11.59,13.60,12.82,13.52,13.73,...
        12.53,10.78,13.69,10.41,11.13,11.28,11.28,10.40,10.47,10.51];

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

% Максимальное и минимальное значение
    M_max = X(n);
    M_min = X(1);
    fprintf("\nМаксимальное значение = %.4f\n", M_max);
    fprintf("Минимальное значение = %.4f\n", M_min);

% Размах выборки
    R = M_max - M_min;
    fprintf("\nРазмах выборки = %.4f\n", R);

% Оценка математического ожидания и дисперсии
    mu = sum(X) / n;
    S2 = sum((X-mu).^2) / (n - 1);
    fprintf("\nОценка математического ожидания = %.4f\n", mu);
    fprintf("Оценка дисперсии = %.4f\n", S2);

% Группировка значений выборки в интервалы
    m = floor(log2(n)) + 2;
    fprintf("\nЧисло интервалов = %d\n\n", m);

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

    J = [];

    for i = 1 : m
      J(i, 1) = X(1) + (i - 1) * d;
```

```
      J(i, 2) = X(1) + i * d;
    end

    N = zeros(m);

    for i = 1 : n
      for j = 1 : m
        if X(i) >= J(j, 1) && X(i) < J(j, 2)
          N(j) = N(j) + 1;
        end
      end
    end

    N(m) = N(m) + 1;

    for i = 1 : m - 1
      fprintf("[%.4f,%.4f) - %d\n", J(i, 1), J(i, 2), N(i));
    end

    fprintf("[%.4f,%.4f] - %d\n", J(m, 1), J(m, 2), N(m));

%Построение графиков
    f = [];

    for i = 1 : m
      f(i, 1) = (J(i, 1) + J(i, 2)) / 2;
      f(i, 2) = N(i) / (n * d);
    end

    args = X(1):1e-4:X(n);
    sigma = sqrt(S2);
    f_normal = normpdf(args, mu, sigma);

    bar(f(:,1), f(:,2), 1, "g");
    hold on;
    plot(args, f_normal, 'b', 'LineWidth', 2);
    grid;


    F = [];

    F(1, 1) = X(1) - 1;
    F(1, 2) = 0;


    for i = 2 : (n + 1)
      cnt = 0;
      for j = 1 : n
```

```
        if X(j) <= X(i - 1)
          cnt = cnt + 1;
        end
      end

      F(i, 1) = X(i - 1);
      F(i, 2) = cnt / n;
    end

    F(n + 2, 1) = X(n) + 1;
    F(n + 2, 2) = 1;

    F_normal = normcdf(args, mu, sigma);

    figure;
    stairs(F(:,1), F(:,2), "r");
    hold on;
    plot(args, F_normal, "b");
    grid;
```

3.2 Результаты расчетов для выборки из индивидуального варианта

```
Максимальное значение = 15.2100
Минимальное значение = 9.5400

Размах выборки = 5.6700

Оценка математического ожидания = 11.9532
Оценка дисперсии = 1.1818

Число интервалов = 8

[9.5400,10.2487) - 5
[10.2487,10.9575) - 17
[10.9575,11.6662) - 28
[11.6662,12.3750) - 24
[12.3750,13.0838) - 28
[13.0838,13.7925) - 12
[13.7925,14.5013) - 5
[14.5013,15.2100] - 1
```

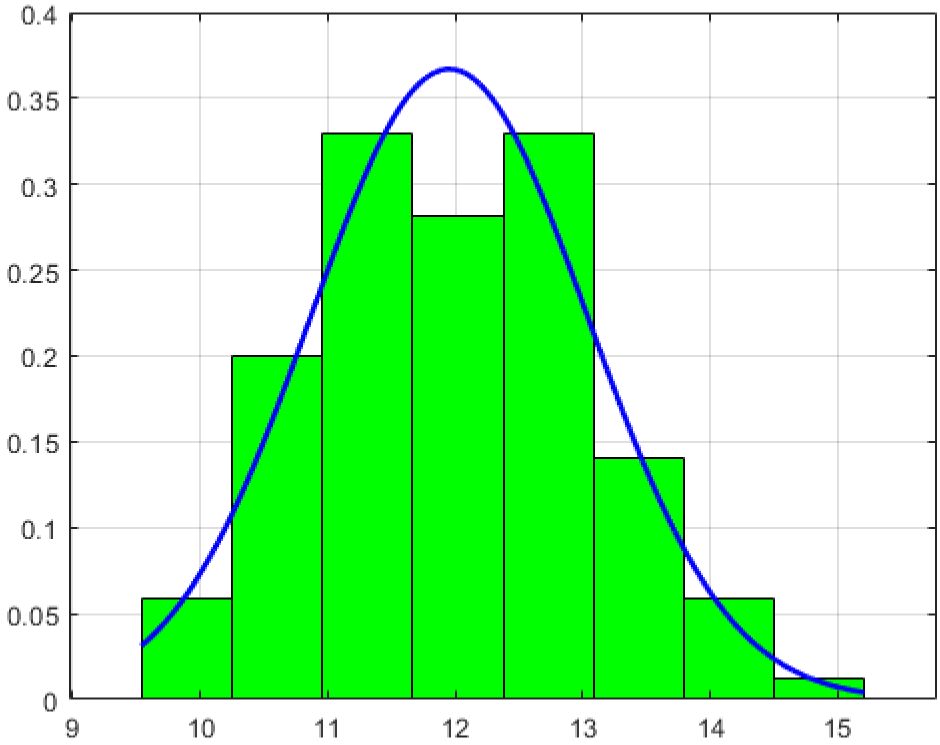

Рисунок 3.1 – Гистограмма и график функции плотности распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

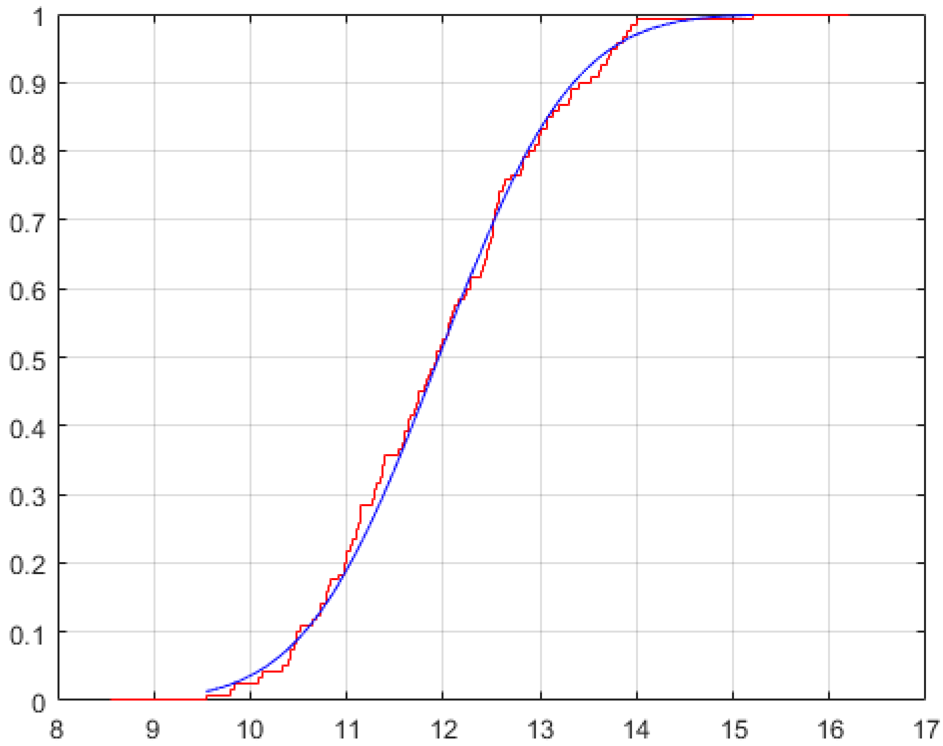

Рисунок 3.2 – График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2