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»

Отчёт по лабораторной работе №1 по курсу

«Математическая статистика»

Тема Гистограмма и эмпирическая функция распределения.

Студент Сироткина П.Ю.

Номер варианта 12

Группа ИУ7-66Б

Преподаватель Андреева Т.В.

Оценка ______________

Москва — 2022 г.

1. Цель работы

Построение гистограммы и эмпирической функции распределения.

2. Постановка задачи

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.

3. Данные для лабораторной работы согласно индивидуальному варианту

Листинг 1: Выборка для варианта №12

```
x = (11.89, 9.60,  9.29,  10.06, 9.50,  8.93,  9.58,  6.81,  8.69,
     9.62,  9.01,  10.59, 10.50, 11.53, 9.94,  8.84,  8.91,  6.90,
     9.76,  7.09,  11.29, 11.25, 10.84, 10.76, 7.42,  8.49,  10.10,
     8.79,  11.87, 8.77,  9.43,  12.41, 9.75,  8.53,  9.72,  9.45,
     7.20,  9.23,  8.93,  9.15,  10.19, 9.57,  11.09, 9.97,  8.81,
     10.73, 9.57,  8.53,  9.21,  10.08, 9.10,  11.03, 10.10, 9.47,
     9.72,  9.60,  8.21,  7.78,  10.21, 8.99,  9.14,  8.60,  9.14,
     10.95, 9.33,  9.98,  9.09,  10.35, 8.61,  9.35,  10.04, 7.85,
     9.64,  9.99,  9.65,  10.89, 9.08,  8.60,  7.56,  9.27,  10.33,
     10.09, 8.51,  9.86,  9.24,  9.63,  8.67,  8.85,  11.57, 9.85,
     9.27,  9.69,  10.90, 8.84,  11.10, 8.19,  9.26,  9.93,  10.15,
     8.42,  9.36,  9.93,  9.11,  9.07,  7.21,  8.22,  9.08,  8.88,
     8.71,  9.93,  12.04, 10.41, 10.80, 7.17,  9.00,  9.46,  10.42,
     10.43, 8.38,  9.01)
```

4. Выполнение лабораторной работы

4.1 Формулы для вычисления некоторых требуемых величин

1. Максимальное и минимальное значение выборки: $M_{max} = X_{(n)}$, $M_{min} = X_{(1)}$;
2. Размах R выборки: $R = M_{max} - M_{min}$;
3. Оценки $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX:
   - Выборочное среднее: $\hat{\mu}(\vec{X}) = \overline{X} = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} X_i$;
   - Выборочная дисперсия: $S^2(\vec{X}) = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} (X_i - \overline{X})^2$.

4.2 Интервальный статистический ряд

Если объем выборки достаточно велик ($n > 50$), то элементы выборки группируют в т.н. интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}; x_{(n)}]$ разбивают на m равновеликих промежутков. Ширина каждого из них определяется следующим образом:

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

Количество отрезков определяется следующей формулой: $m = [log_2 n] + 2$, где n - объем выборки.

Далее полагают:

$$J_i = [x(1) + (i-1) \cdot \Delta; x(1) + i \cdot \Delta], i = \overline{1, m}.$$

$$J_m = [x(1) + (m-1) \cdot \Delta; x(n)].$$

Интервальным статистическим рядом называют таблицу вида:

J_1	...	J_i	...	J_m
n_1	...	n_i	...	n_m

Здесь n_i - число элементов выборки $\vec{x}$, попавших в промежуток J_i.

4.3 Гистограмма

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:

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

Функция $f_n(x)$ является статистическим аналогом функции плотности.

График этой функции называется *гистограммой*.

4.4 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 t.

Эмпирической функцией распределения, построенной по выборке $\vec{x}$, называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную правилом:

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

4.5 Функция плотности и функция распределения нормальной случайной величины

Говорят, что случайная величина X распределена по нормальному закону с параметрами m и σ^2, если *функция плотности* распределения вероятностей X имеет вид:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{(2 \cdot \pi)}} \cdot e^{-\frac{(x-m)^2}{(x-m)^2}}.$$

Функция распределения случайной величины X, распределенной по нормальному закону, имеет вид:

$$F(x) = \frac{1}{\sigma \cdot \sqrt{(2 \cdot \pi)}} \cdot \int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2 \cdot \sigma^2}}$$

4.6 Код программы

```
X = [11.89,  9.60,  9.29,  10.06, 9.50,  8.93,  9.58,  6.81,
     8.69,   9.62,  9.01,  10.59, 10.50, 11.53, 9.94,  8.84,
     8.91,   6.90,  9.76,  7.09,  11.29, 11.25, 10.84, 10.76,
     7.42,   8.49,  10.10, 8.79,  11.87, 8.77,  9.43,  12.41,
     9.75,   8.53,  9.72,  9.45,  7.20,  9.23,  8.93,  9.15,
     10.19,  9.57,  11.09, 9.97,  8.81,  10.73, 9.57,  8.53,
     9.21,   10.08, 9.10,  11.03, 10.10, 9.47,  9.72,  9.60,
     8.21,   7.78,  10.21, 8.99,  9.14,  8.60,  9.14,  10.95,
     9.33,   9.98,  9.09,  10.35, 8.61,  9.35,  10.04, 7.85,
     9.64,   9.99,  9.65,  10.89, 9.08,  8.60,  7.56,  9.27,
     10.33,  10.09, 8.51,  9.86,  9.24,  9.63,  8.67,  8.85,
     11.57,  9.85,  9.27,  9.69,  10.90, 8.84,  11.10, 8.19,
     9.26,   9.93,  10.15, 8.42,  9.36,  9.93,  9.11,  9.07,
     7.21,   8.22,  9.08,  8.88,  8.71,  9.93,  12.04, 10.41,
     10.80,  7.17,  9.00,  9.46,  10.42, 10.43, 8.38,  9.01]


max_m = max(X);
min_m = min(X);

R = max_m - min_m;

MX = mean(X);
DX = var(X);

sigma = std(X);

m = floor(log2(length(X))) + 2;
delta = R / m;

x = (min_m - 1):(sigma / 250):(max_m + 1);
y1 = normpdf(x, MX, sigma);

hold on;
hist(X, m, 1 / delta);
plot(x, y1);

figure;
hold on;
y2 = empirical_cdf(x, X);
y3 = normcdf(x, MX, sigma);
plot(x, y2);
plot(x, y3);
```

4.7 Результат работы программы

Числовые характеристики:

$$M_{\min} = 6.81, \quad M_{\max} = 12.41, \quad R = 5.6, \quad m = 8, \quad \hat{\mu}(\vec{x}) = 9.4872, \quad S^2(\vec{x}) = 1.2173$$

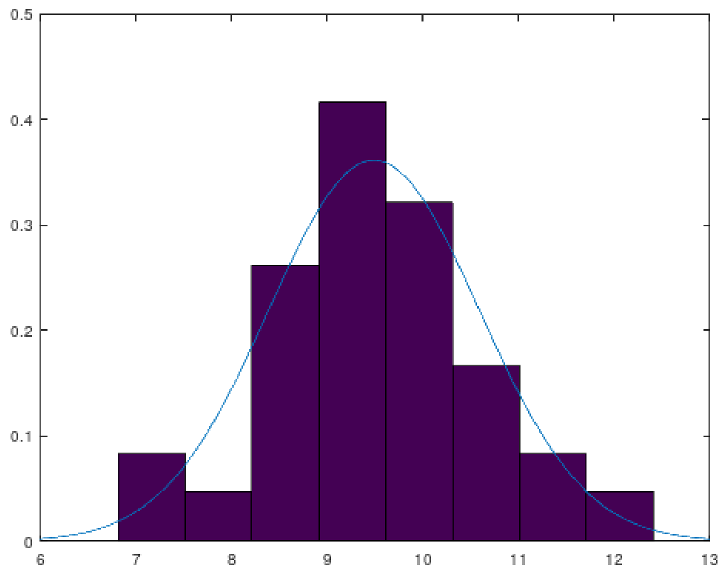

Рис. 1: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

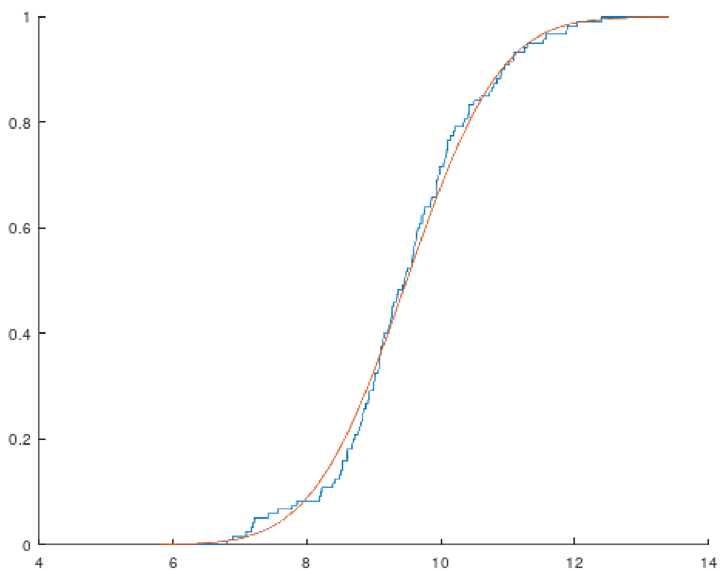

Рис. 2: График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2