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 дисциплине "Математическая статистика"

Тема Гистограмма и эмпирическая функция распределения

Студент Малышев И. А.

Группа ИУ7-61Б

Оценка (баллы) _______________

Преподаватель: Власов П. А.

Москва — 2022 г.

Задание

Цель работы

Построение гистограммы и эмпирической функции распределения.

Постановка задачи

1. Для выборки объёма 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.

Вариант выборки

Вариант 12

$\vec{X}$=(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)

Теоретические сведения

Формулы для вычисления величин

Минимальное и максимальное значения выборки

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

Размах выборки

$$R = M_{\max} - M_{\min}. \tag{2}$$

Оценки выборочного среднего (математического ожидания) и исправленной выборочной дисперсии

$$\begin{aligned} \hat{\mu}(\vec{X}_n) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ S^2(\vec{X}_n) &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \end{aligned} \tag{3}$$

Определение эмпирической плотности и гистограммы

Пусть $\vec{x}$ – выборка из генеральной совокупности X. Если объем n этой выборки велик, то значения x_i группируют в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ делят на m равновеликих частей:

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

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

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

Интервальным статистическим рядом называют таблицу:
где n_i – количество элементов выборки $\vec{x}$, которые $\in J_i$.
Обычно выборку разбивают на $m = [\log_2 n] + 2$ интервалов, где n – размер выборки.

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

Гистограмма – это график эмпирической плотности.

Эмпирической плотностью, отвечающей выборке $\vec{x}$, называют функцию:

$$\hat{f}(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1; m} \\ 0, \text{иначе} \end{cases} \tag{4}$$

где J_i – полуинтервал статистического ряда, n_i – количество элементов выборки, входящих в полуинтервал, n – количество элементов выборки.

Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, ..., x_n)$ – выборка из генеральной совокупности X. Обозначим $n(x, \vec{x})$ – число элементов вектора $\vec{x}$, которые имеют значения меньше x.

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную как:

$$F_n(x) = \frac{n(x, \vec{x})}{n} \tag{5}$$

Результаты работы программы

Текст 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 ];

M_max = max(X);
M_min = min(X);
fprintf("\na) M_max (максимальное значение) = %f; M_min (минимальное значение) = %f", M_max, M_min);

R = M_max - M_min;
fprintf("\nб) R (размах) = %f", R);

MX = mean(X);
DX = var(X);
fprintf("\nв) μ (оценка математическогоожидания) = %f; S^2 (оценка дисперсии) = %f", MX, DX);

m = floor(log2(length(X))) + 2;
fprintf("\nг)Группировка значений выборки в m = [log2 n] + 2 интервала: m = %f\n", m);

[counts, edges] = histcounts(X, m, 'BinLimits', [min(X), max(X)]);

delta = R / m;
hist = histogram();
hist.BinEdges = edges;
hist.BinCounts = counts / (length(X) * delta);

hold on;
```

```
sigma = sqrt(DX);
x = M_min : (sigma / 100) : M_max;
f = normpdf(x, MX, sigma);
plot(x, f, 'red');

F = normcdf(x, MX, sigma);
figure;
hold on;
ecdf(X);
plot(x, F, 'green');
```

Результаты расчётов

- $M_{\max} = 12.4100$
- $M_{\min} = 6.8100$
- $R = 5.6000$
- $\hat{\mu}(\vec{x}_n) = 9.4872$
- $S^2(\vec{x}_n) = 1.2173$
- $m = 8$

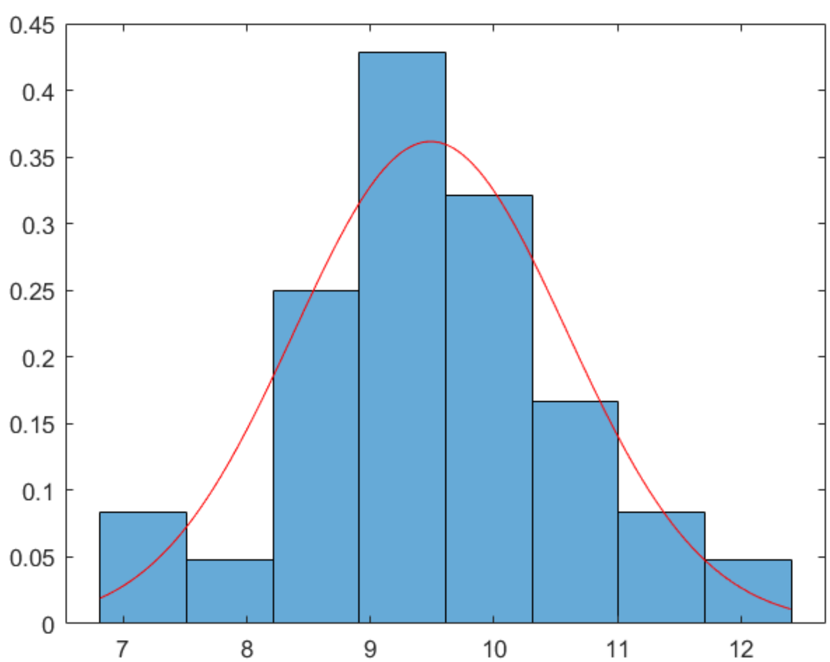

Рис. 1: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочным математическим ожиданием и исправленной выборочной дисперсией

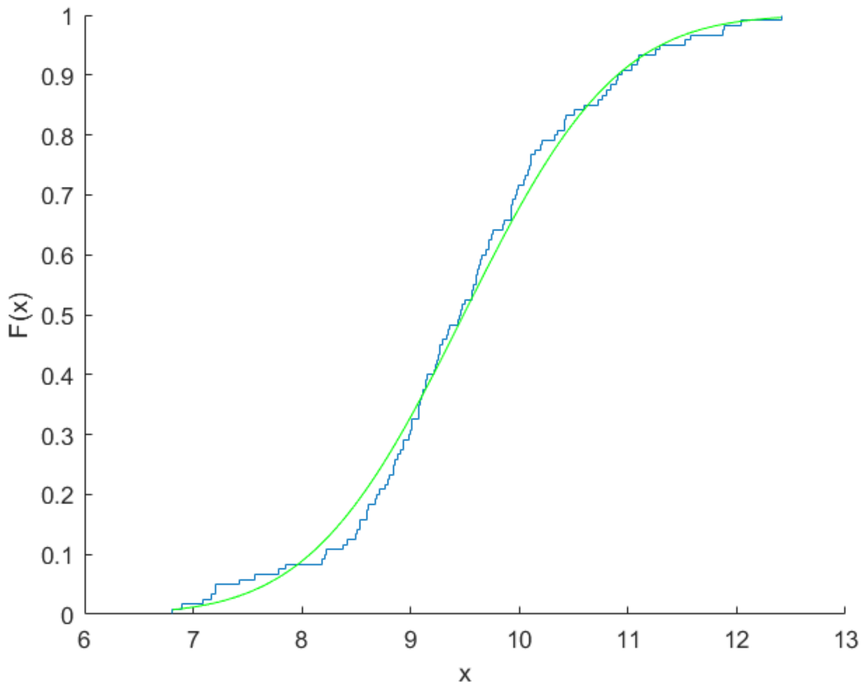

Рис. 2: График эмпирической функции распределения и функции распределения нормальной случайной величины с выборочным математическим ожиданием и исправленной выборочной дисперсией