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 дисциплине «Математическая Статистика»

на тему: «Гистограмма и эмпирическая функция распределения»

Студент группы ИУ7-65Б	(Подпись, дата)	Турчанинов А. М. (Фамилия И.О.)
Преподаватель	(Подпись, дата)	Власов П. А. (Фамилия И.О.)
Преподаватель	(Подпись, дата)	Саркисян П. С. (Фамилия И.О.)

Москва — 2024 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

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 оценки $\hat{\mu}$ и S^2 математического ожидания MX и дисперсия 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.

1 Теоретическая часть

В данной части рассмотрены формулы для вычисления величин, эмпирическая плотность, эмпирическая функция распределения и гистограмма.

1.1 Формулы для вычисления величин

Минимальное значение выборки

$$M_{min} = min\{x_1, \cdots, x_n\}, \tag{1.1}$$

где $(x1, \cdots, x_n)$ – реализация случайной выборки.

Максимальное значение выборки

$$M_{max} = max\{x_1, \cdots, x_n\}, \tag{1.2}$$

где $(x1, \cdots, x_n)$ – реализация случайной выборки.

Размах выборки

$$R = M_{max} - M_{min}, \tag{1.3}$$

где M_{max} – максимальное значение выборки, M_{min} –минимальное значение выборки.

Оценка математического ожидания

$$\hat{\mu}(\vec{X_n}) = \bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.4}$$

Несмещённая оценка дисперсии

$$S^2(\vec{X_n}) = \frac{n}{n-1}\hat{\sigma} = \frac{n}{n-1}\sum_{i=1}^{n}(X_i - \bar{X_n})^2 \tag{1.5}$$

Выборочная дисперсия

$$\hat{\sigma}(\vec{X_n}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2 \tag{1.6}$$

Количество интервалов

$$m = [\log_2 n] + 2 \tag{1.7}$$

1.2 Эмпирическая плотность и гистограмма

Определение. Эмпирической плотностью распределения случайной выборки $\vec{X_n}$ называют функцию:

$$f_n(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1, m}; \\ 0, \textit{иначе}. \end{cases} \tag{1.8}$$

где $J_i, i = \overline{1, m}$ – полуинтервал из $J = [x_{(1)}, x_{(n)}]$.

$$\begin{aligned} &x_{(1)} = min\{x_1, \cdots, x_n\}, \\ &x_{(n)} = max\{x_1, \cdots, x_n\}, \\ &J_i = [x_{(1)} + (i-1)\Delta], x_{(i)} + i\Delta, i = \overline{1, m-1}, \\ &J_m = [x_{(1)} + (m-1)\Delta, x_{(1)} + m\Delta], \end{aligned}$$

m – количество полуинтервалов интервала $J = [x_{(1)}, x_{(n)}]$,
Δ – длина полуинтервала $J_i, i = 1, \bar{m}$ равная

$$\Delta = \frac{x_{(n)} - x_{(1)}}{m} = \frac{|J|}{m} \tag{1.9}$$

n_i – количество элементов выборки в полуинтервале $J_i, i = \overline{1, m}$,
n – количество элементов в выборке.

Определение. График функции $f_n(x)$ называют гистограммой.

1.3 Эмпирическая функция распределения

*Определение.*Пусть

1) $\vec{X_n} = (X_1, \cdots, X_n)$ – случайная выборка,

2) $\vec{x_n} = (x_1, \cdots, x_n)$ – реализация случайной выборки,

3) $n(x, \vec{x_n})$ – количество элементов выборки $\vec{x_n}$, которые меньше x, тогда эмпирической функцией распределения называют функцию

$$F_n : \mathbb{R} \to \mathbb{R}, \quad F_n = \frac{n(x, \vec{x_n})}{n}. \tag{1.10}$$

Замечание.

1) $F_n(X)$ обладает всеми свойствами функции распределения;

2) $F_n(X)$ кусочно-постоянна;

3) если все элементы вектора различны, то

$$F_n(x) = \begin{cases} 0, & x \le x_{(1)}; \\ \frac{i}{n}, & x_{(i)} < x \le x_{(i+1)}, \quad i = \overline{1, n-1}; \\ 1, & x > x_{(n)} \end{cases} \tag{1.11}$$

4) Эмпирическая функция распределения позволяет интерпретировать выборку $\vec{x_n}$ как реализацию дискретной случайной величины $\widetilde{X}$, ряд распределения которой имеет вид:

$\widetilde{X}$	$x_{(1)}$	$\dots$	$x_{(n)}$
P	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$

Это позволяет рассматривать числовые характеристики случайной величины $\widetilde{X}$ как приближенные значения числовых характеристик случайной величины X.

2 Практическая Часть

В данной части представлен листинг и результат работы программы.

2.1 Листинг программы

```
function lab1()
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];
X = sort(X);
Mmax = max(X);
Mmin = min(X);
fprintf('Mmin = %s\n', num2str(Mmin));
fprintf('Mmax = %s\n', num2str(Mmax));
R = Mmax - Mmin;
fprintf('R = %s\n', num2str(R));
MU = getMU(X);
fprintf('MU = %s\n', num2str(MU));
Ssqr = getSsqr(X);
fprintf('S^2 = %s\n', num2str(Ssqr));
m = getNumberOfIntervals(X);
fprintf('m = %s\n', num2str(m))
createGroup(X);
hold on;
distributionDensity(X, MU, Ssqr, m);
figure;
empiricF(X);
hold on;
distribution(X, MU, Ssqr, m);
end
```

```
function mu = getMU(X)
n = length(X);
mu = sum(X)/n;
end
function Ssqr = getSsqr(X)
n = length(X);
MX = getMU(X);
Ssqr = sum((X - MX).^2) / (n-1);
end
function m = getNumberOfIntervals(X)
m = floor(log2(length(X)) + 2);
end
function createGroup(X)
n = length(X);
m = getNumberOfIntervals(X);
intervals = zeros(1, m+1);
numCount = zeros(1, m+1);
Delta = (max(X) - min(X)) / m;
for i = 0: m
intervals(i+1) = X(1) + Delta * i;
end
j = 1;
count = 0;
for i = 1:n
if (X(i) >= intervals(j+1))
j = j + 1;
end
numCount(j) = numCount(j) + 1;
count = count + 1;
end
graphBuf = numCount(1:m+1);
for i = 1:m+1
graphBuf(i) = numCount(i) / (n*Delta);
end
stairs(intervals, graphBuf),grid;
end
```

```
function distributionDensity(X, MX, DX, m)
R = X(end) - X(1);
delta = R/m;
Sigma = sqrt(DX);
Xn = (MX - R): delta/50 :(MX + R);
Y = normpdf(Xn, MX, Sigma);
plot(Xn, Y), grid;
end
function distribution(X, MX, DX, m)
R = X(end) - X(1);
delta = R/m;
Xn = (MX - R): delta :(MX + R);
Y = 1/2 * (1 + erf((Xn - MX) / sqrt(2*DX)));
plot(Xn, Y, 'r'), grid;
end
function empiricF(X)
[yy, xx] = ecdf(X);
stairs(xx, yy), grid;
end
```

2.2 Результат работы программы

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

2.3 Графики

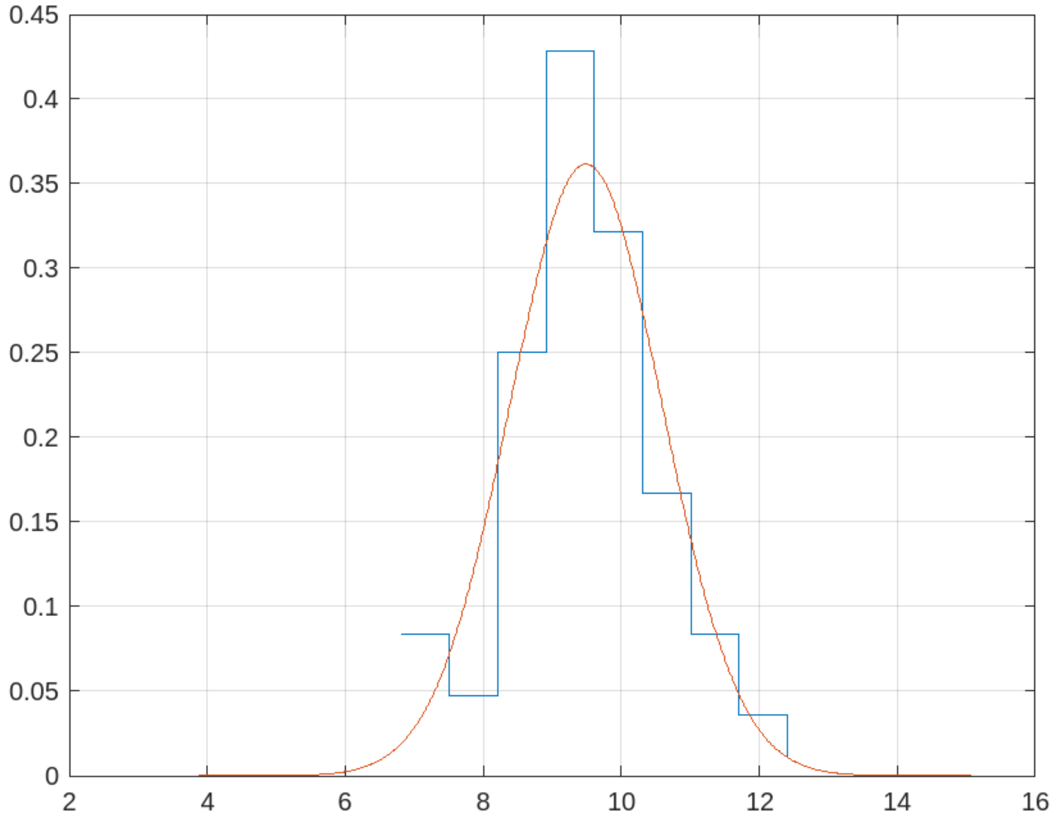

Рисунок 2.1 – Гистограмма и график функции плотности распределения нормальной случайной величины.

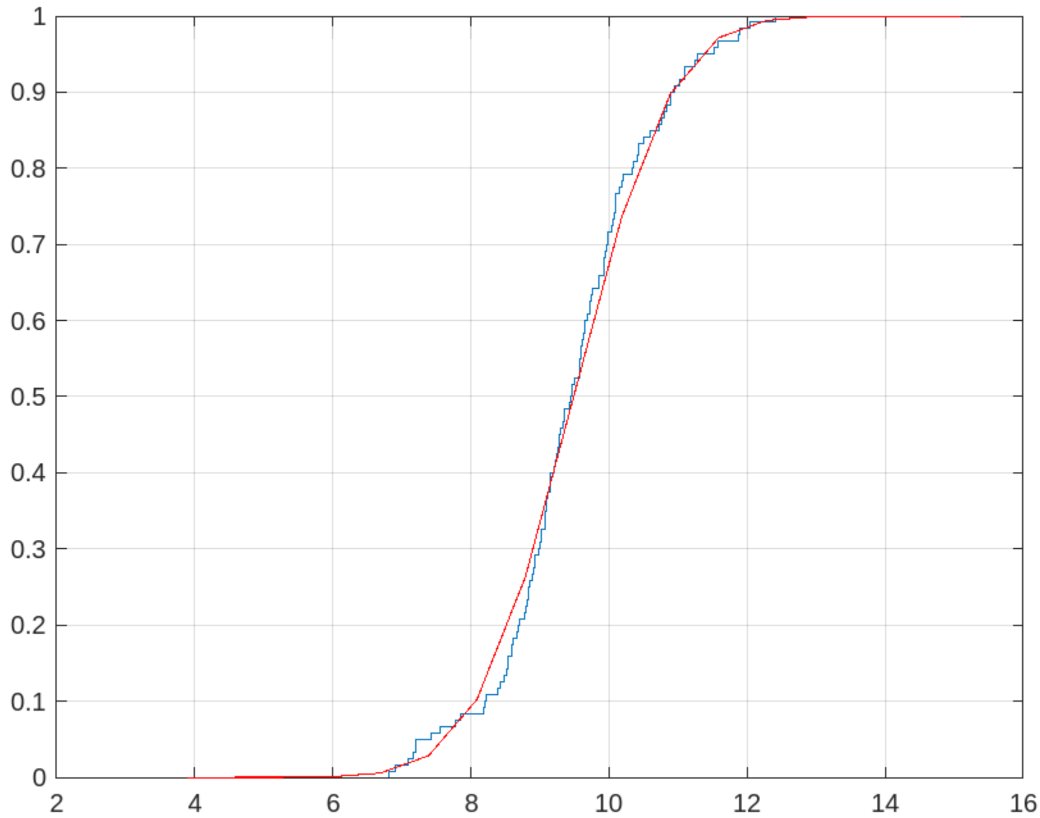

Рисунок 2.2 – График эмпирической функции распределения и функции распределения нормальной случайной величины