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 Золотухин А.В.

Группа ИУ7-64Б

Преподаватель Андреева Т.В.

Москва — 2023 г.

1 | Задание

1.1 Цель работы

Цель работы: построение гистограммы и эмпирической функции распределения.

1.2 Содержание работы

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.

2 | Теоретическая часть

Пусть $\vec{x}$ – выборка из генеральной совокупности X.

2.1 Формулы для вычисления величин

2.1.1 Минимальное и максимальное значения выборки

$$\begin{gathered} M_{\max} = x_{(n)} \\ M_{\min} = x_{(1)} \end{gathered}, \tag{2.1}$$

где $x_{(n)}$ — последнее значение вариационного ряда,
$x_{(1)}$ — первое значение вариационного ряда.

2.1.2 Размах выборки

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

2.1.3 Оценки математического ожидания и дисперсии

$$\begin{gathered} \hat{\mu}(\vec{X}_n) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \\ S^2(\vec{X}_n)\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \end{gathered}, \tag{2.3}$$

где $\overline{X}_n$ — выборочное среднее.

2.2 Эмпирическая плотность и гистограмма

При большом объеме n этой выборки значения x_i группируют в интервальный статистический ряд.

Отрезок $J = [x_{(1)}, x_{(n)}]$ делят на m равновеликих частей.

$$m = [\log_2 n] + 2, \tag{2.4}$$

где n – размер выборки.

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

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

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

Интервальным статистическим рядом называют таблицу:

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

где n_i – количество элементов выборки $\vec{x}$, которые $\in J_i$.

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

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

Гистограмма – это график эмпирической плотности.

2.3 Эмпирическаяя функция распределения

Обозначим $n(x, \vec{x})$ – число элементов выборки $\vec{x}$, которые имеют значения меньше x.

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

$$F_n(x) = \frac{n(x, \vec{x})}{n} \tag{2.9}$$

3 | Практическая часть

```
n = length(X);

M_max = max(X);
M_min = min(X);
fprintf("max = %f\n", X_max)
fprintf("min = %f\n", X_min)

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

MX = mean(X);
DX = var(X);
fprintf("MX = %f    DX = %f\n", MX, DX)

m = floor(log2(n)) + 2;
delta = R / m;
fprintf('m = %i   delta = %f\n', m, delta);

h = histogram(X, m);

sigma = sqrt(DX);
x = (M_min - 1):(sigma / 100):(M_max + 1);
f = normpdf(x, MX, sigma);

figure;
delta = h.BinWidth;
heights = h.Values / (n * delta);
centers = [];
for i = 1:m
centers = [centers, (h.BinEdges(i + 1) + h.BinEdges(i)) / 2];
end
hold on;
bar(centers, heights, 1);
plot(x, f, 'LineWidth', 2);


F = normcdf(x, MX, sigma);
figure;
hold on;
[yy, xx] = ecdf(X);
```

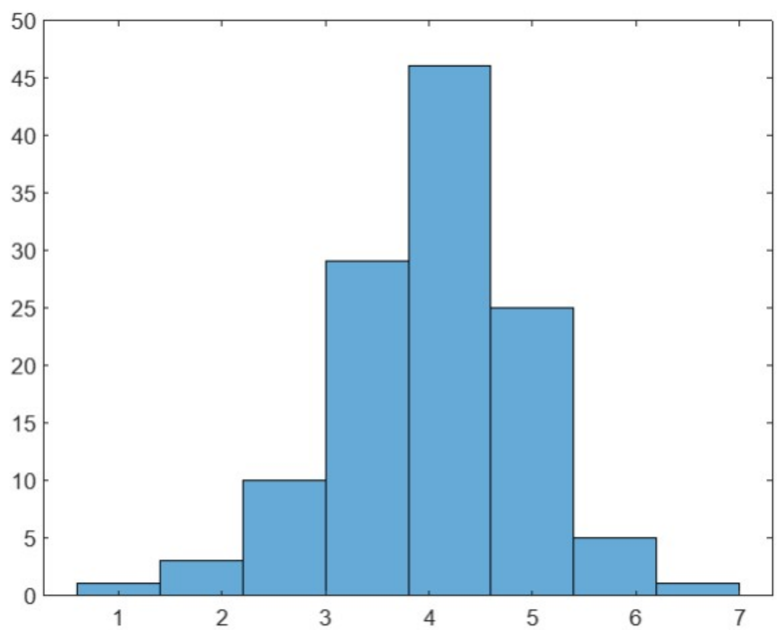
```
xx = [xx; x_max + 1];
yy = [yy; 1];
plot(x, F, 'r');
```

4 | Экспериментальная часть

4.1 Результаты расчетов

Выборка: X=[3.89,2.60,2.56,2.33,4.35,3.97,4.16,4.10,4.82,3.17,4.15,4.05,4.45,3.76,4.83,4.27,2.03,3.13, 3.69,5.44,2.61,5.20,4.03,5.26,4.42,3.67,4.68,5.22,4.11,3.35,3.83,5.34,4.69,5.12,4.33,4.41,4.82,4.65,3.12,4.71, 2.36,3.10,3.52,1.74,3.46,2.89,5.42,4.12,5.36,4.26,3.49,5.85,5.24,4.23,4.57,3.54,4.38,4.31,2.28,4.17,3.66,5.25, 3.11,4.48,3.80,4.05,3.48,1.12,2.16,4.68,4.21,3.22,4.29,4.70,4.37,4.60,4.15,4.06,3.81,3.58,4.34,3.87,4.53,6.02, 3.34,3.34,2.87,4.64,3.03,3.08,3.46,4.04,4.77,3.59,3.49,6.53,4.16,2.95,4.85,5.53,3.59,4.24,4.45,4.42,4.10,3.77, 4.56,3.89,4.13,4.00,3.95,2.73,4.76,3.19,3.66,4.87,4.18,5.30,4.97,4.56];

Результаты работы программы для выборки (Вариант 4) представлены на рисунках.



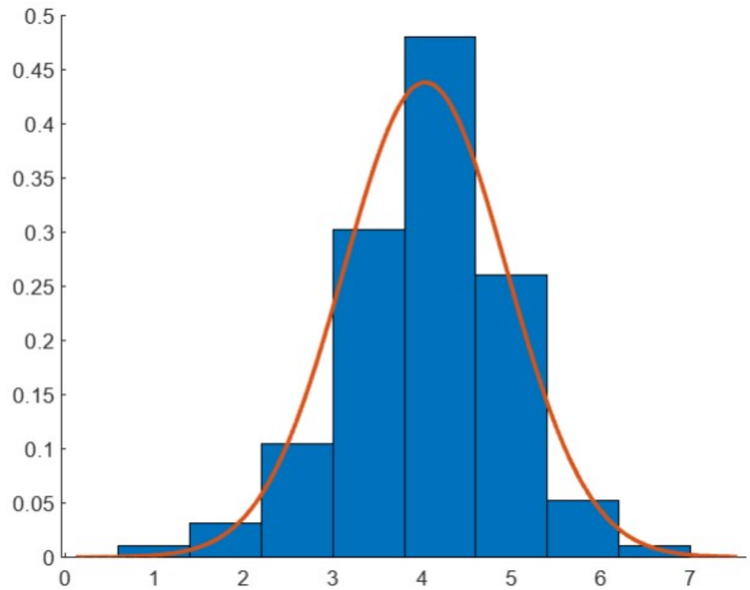


Рис. 4.1: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочными мат. ожиданием и дисперсией

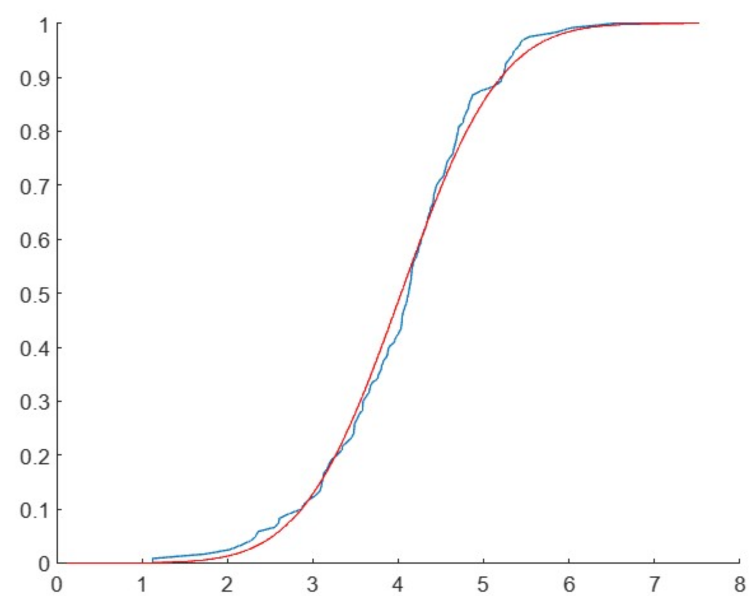


Рис. 4.2: График эмперической функции распределения и функции распределения нормальной случайной величины с выборочными мат. ожиданием и дисперсией