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

Название «Гистограмма и эмпирическая функция распределения»

Дисциплина «Математическая статистика»

Вариант 3

Студент ИУ7-65Б ______ Клименко А.К.

(подпись, дата) (Фамилия И.О.)

Преподаватель ______ Андреева Т.В.

(подпись, дата) (Фамилия И.О.)

Москва, 2022

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:

   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 $m = \lfloor \log_2 n \rfloor + 2$ 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.

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.

1 Выполнение работы

1.1 Расчётные соотношения

В списке ниже в скобках указаны идентификаторы соответствующих величин, рассчитываемые в программе.

- Максимальное значение выборки $\overrightarrow{x_n}$ (`M_max`)

$$M_{max} = x_{(n)} \tag{1.1}$$

- Минимальное значение выборки (`M_min`)

$$M_{min} = x_{(1)} \tag{1.2}$$

- Размах выборки (`R`)

$$R = x_{(n)} - x_{(1)} \tag{1.3}$$

- Оценка математического ожидания (`mu_hat`)

$$\hat{\mu}(\overrightarrow{x_n}) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.4}$$

- Исправленная ценка дисперсии (`S_2`)

$$S^2(\overrightarrow{x_n}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{1.5}$$

- Количество интервалов в интервальном статистическом ряду (`m`)

$$m = \lfloor \log_2 n \rfloor + 2 \tag{1.6}$$

1.2 Определение эмпирической плотности и гистограммы

Определение: интервальным статистическим рядом отвечающим выборке $\overrightarrow{x}$ называется таблица вида

J_1	J_2	$\ldots$	J_m
n_1	n_2	$\ldots$	n_m

в которой

$J_i = [x_{(1)} + (i-1)\Delta;\ x_{(1)} + i\Delta),\quad i = \overline{1,m-1}$
$J_m = [x_{(1)} + (m-1)\Delta;\ x_{(n)}]$
$\Delta = \dfrac{|J|}{m} = \dfrac{x_{(n)} - x_{(1)}}{m}$
n_i – число элементов выборки $\overrightarrow{x}$ попавших в J_i.

Определение: эмпирической плотностью распределения, соответствующей выборке $\overrightarrow{x}$, называется функция

$$f_n(x) = \begin{cases} \dfrac{n_i}{n\Delta}, & x \in J_i \\ 0, & \textit{иначе} \end{cases} \tag{1.7}$$

Определение: график эмпирической функции плотности называется гистограммой.

1.3 Определение эмпирической функции распределения

Пусть $\overrightarrow{x} = (x_1, x_2, \ldots, x_n)$ – выборка из генеральной совокупности X.

Обозначим $n(t, \overrightarrow{x})$ – число компонент вектора $\overrightarrow{x}$, которые меньше, чем t.

Определение: эмпирической функцией распределения построенной по выборке $\overrightarrow{x}$ называется функция

$$F_n : \mathbb{R} \to \mathbb{R}, \tag{1.8}$$

определённая правилом

$$F_n(t) = \frac{n(t, \overrightarrow{x})}{n}. \tag{1.9}$$

1.4 Результаты работы

Величина	**Значение**
N	120
M_min	-2.77
M_max	2.92
R	5.69
mu_hat	0.2322
S_2	1.0406
m	8

Таблица 1.1 — Результаты вычислений параметров выборки

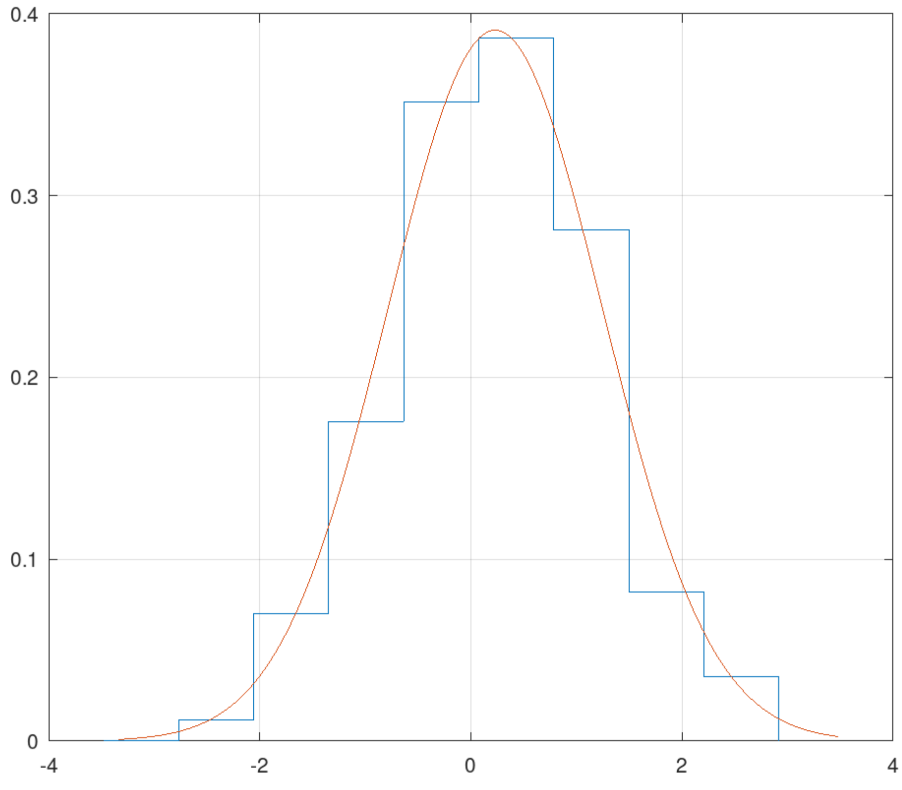

Рисунок 1.1 — Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины $Y \sim \mathbb{N}(\hat{\mu}, S^2)$

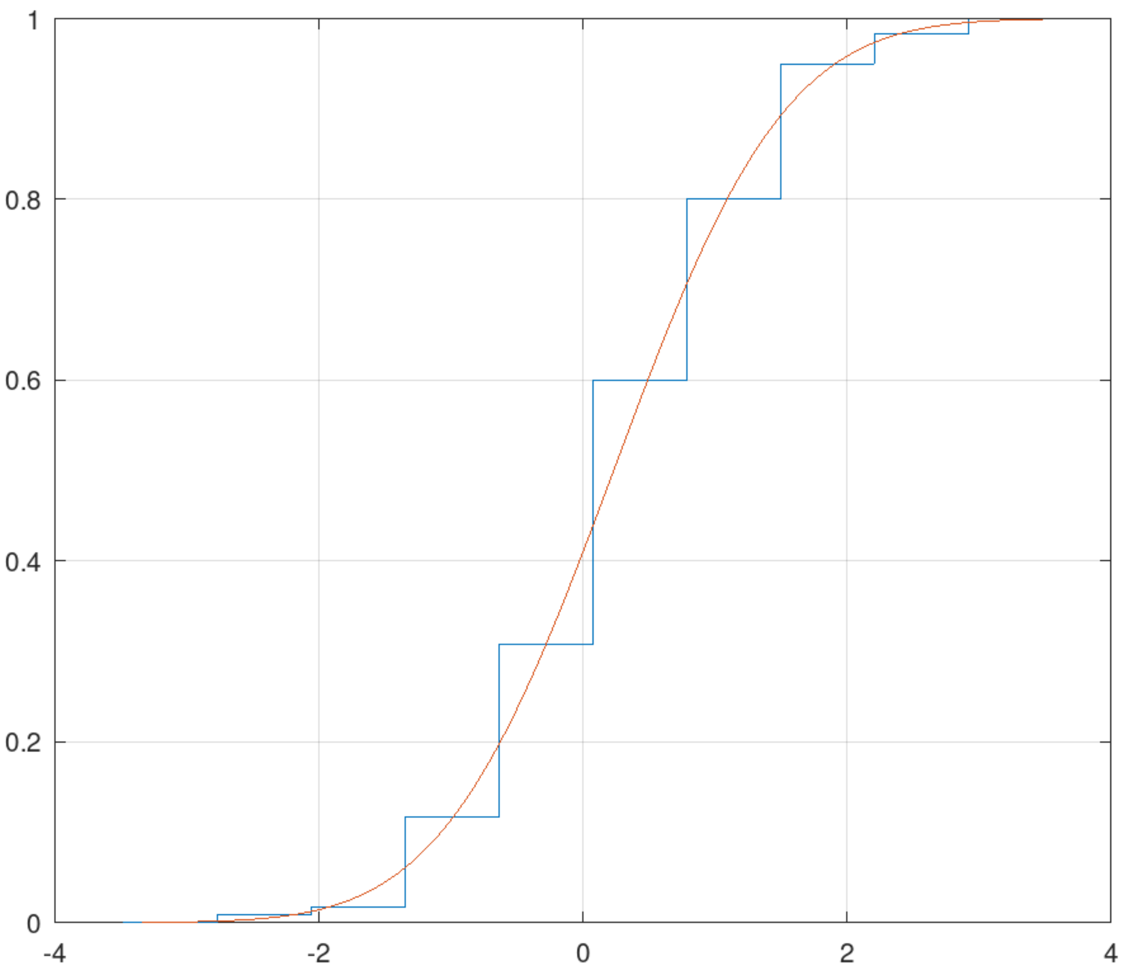

Рисунок 1.2 — График эмпирической функции распределения и функции распределения нормальной случайной величины $Y \sim \mathbb{N}(\hat{\mu}, S^2)$

1.5 Текст программы

```
#!/bin/octave -qf

pkg load statistics

X = [-0.45, -0.33, 2.92, -1.25, -1.20, 0.05, -0.53, -0.19, 1.49, 0.67, 0.22, 1.23, 0.50, -0.92,
    0.90, -1.52, -0.15, -1.24, -0.47, -0.45, 0.18, -0.05, 1.58, 1.74, 2.37, -0.24, -1.34, 1.05,
    1.28, 1.37, 1.18, 0.22, 0.11, 0.28, -0.64, -0.39, -1.77, -1.61, 0.47, 0.77, -0.27, -1.19,
    -0.25, 1.04, -0.16, 0.42, 0.29, 0.10, 1.04, 0.43, -0.67, 0.41, -0.62, -1.49, 1.46, -2.77,
    2.09, 0.88, -0.36, -0.71, -0.62, 1.34, -0.78, -0.15, 2.69, 0.92, 1.68, -0.12, 0.34, 0.74,
    1.72, 1.24, 0.23, 0.76, 0.87, -1.52, 0.63, -0.56, 0.83, 0.31, -0.18, 0.99, -1.01, 0.58,
    1.21, -1.51, 0.65, 0.35, -0.37, -0.50, -0.73, 0.63, 0.33, 1.56, -0.98, 0.85, 0.56, -1.07,
    1.47, 1.44, 1.91, 0.24, 1.34, 0.99, 1.27, 0.11, 0.22, -0.25, 0.35, -0.03, -0.56, -0.79,
    2.41, -0.45, -0.44, 0.07, 0.64, 0.69, 0.10, -0.28];

N = length(X)
[M_min, M_max] = bounds(X)          # Минимальное и максимальное значения выборки
R = range(X)                        # Размах выборки

mu_hat = sum(X) / N                 # Среднее наблюдённое значение
S_2 = sum((X - mu_hat).^2) / (N-1)  # Оценка дисперсии

m = floor(log2(N)) + 2              # Количество интервалов
intv_delta = R / m                  # Размер интервала

intv = zeros(2, m);                 # Интервалы
n = zeros(1, m);                    # Частости значений в интервалах

for k = 1:m
    intv(1, k) = X_min + (k - 1) * intv_delta;
    intv(2, k) = X_min + k * intv_delta;
    n(1, k) = sum(intv(1, k) <= X & X < intv(2, k)) / intv_delta / N;
endfor;

X_norm = linspace(M_min, M_max, 200);
Y_norm = 1 / (2 * sqrt(S_2)) * exp(-(X_norm - mu_hat) .^ 2);

stairs(intv(1, 1:m), n);            # Вывод ступенчатого графика функции плотности
hold on
plot(X_norm, Y_norm);               # Вывод графика функции плотности
hold off
pause;

n = cumsum(n) * intv_delta;
Y_norm = stdnormal_cdf((X_norm - mu_hat) / sqrt(S_2));
```

```
stairs(intv(1, 1:m), n);            # Вывод эмпирической функции распределения
hold on
plot(X_norm, Y_norm);               # Вывод функции распределения
hold off
pause;
```

Заключение

В ходе работы были приобретены навыки предварительной обработки исходных данных, вычисления различных статистик для параметров распределения случайных величин, а также были построены гистограммы эмпирической функции распределения и функции плотности распределения.

Все расчёты были проведены с использованием ПО «GNU Octave».