Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

«Московский государственный технический университет имени Н.Э. Баумана» (МГТУ им. Н.Э. Баумана)

Факультет «Информатика и системы управления»
Кафедра «Программное обеспечение ЭВМ и информационные технологии»

Отчет по лабораторной работе №1

по курсу:

«Математическая статистика»

Студент группы ИУ7-63Б: Фурдик Н. О.
(Фамилия И.О.)
Преподаватель: Власов П. А.
(Фамилия И.О.)

Москва, 2020

Оглавление

Постановка задания

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:

а) вычисление максимального значения M_{max} и минимального значения M_{min};

б) размаха R выборки;

в) вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;

г) группировку значений выборки в $m = [log_2 n] + 2$ интервала;

д) построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;

е) 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 1 из индивидуального варианта.

Теоретическая часть

Формулы для вычисления величин

Ниже представлены формулы для вычисления величин.

Минимальное значение выборки

$$M_{\min} = \min\{x_1, \ldots, x_n\}, \quad \text{где} (x_1, \ldots, x_n) - \text{реализация случайной выборки.} \tag{1}$$

Максимальное значение выборки

$$M_{\max} = \max\{x_1, \ldots, x_n\}, \quad \text{где} (x_1, \ldots, x_n) - \text{реализация случайной выборки.} \tag{2}$$

Размах выборки

$$R = M_{\max} - M_{\min}, \quad \text{где } M_{\max} - \text{максимальное значение выборки,} \tag{3}$$

$$M_{\min} - \text{минимальное значение выборки.} \tag{4}$$

Оценка математического ожидания

$$\hat{\mu}(\vec{X}_n) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{5}$$

Несмещённая оценка дисперсии

$$S^2(\vec{X}_n) = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\,. \tag{6}$$

Определения величин

Эмпирическая плотность и гистограмма

Определение 1. *Эмпирической плотностью распределения случайной выборки* $\vec{X}_n$ называют функцию

$$f_n(x) = \begin{cases} \frac{n_i}{n\,\Delta}, & x \in J_i,\ i = \overline{1,m}; \\ 0, & \text{иначе.} \end{cases}, \quad \text{где} \tag{7}$$

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

$$x_{(1)} = \min\{x_1, \ldots, x_n\}, \qquad x_{(n)} = \max\{x_1, \ldots, x_n\}; \tag{8}$$

при этом все полуинтервалы, кроме последнего, не содержат правую границу т. е.

$$J_i = [a_i, a_{i+1}), i = \overline{1; m-1}; \tag{9}$$

$$J_m = [a_m, a_{m+1}]; \tag{10}$$

где

$$a_i = x_{(1)} + (i-1)\cdot\Delta, i = \overline{1; m+1}; \tag{11}$$

2) m — количество полуинтервалов интервала $J = [x_{(1)}, x_{(n)}]$;

3) Δ — длина полуинтервала J_i, $i = \overline{1,m}$ равная

$$\Delta = \frac{x_{(n)} - x_{(1)}}{m} = \frac{|J|}{m}; \tag{12}$$

4) n_i — количество элементов выборки в полуинтервале J_i, $i = \overline{1,m}$;

5) n — количество элементов в выборке.

Определение 2. График функции $f_n(x)$ называют гистограммой.

Эмпирическая функция распределения

1) $\vec{X}_n = (X_1, \ldots, X_n)$ — случайная выборка;

2) $\vec{x}_n = (x_1, \ldots, x_n)$ — реализация случайной выборки $\vec{X}_n$;

3) $n(x, \vec{x}_n)$ — количество элементов выборки $\vec{x}_n$, которые меньше x.

Определение 3. *Эмпирической функцией распределения* называют функцию

$$F_n\colon \mathbb{R} \to \mathbb{R}, \quad F_n(x) = \frac{n(x, \vec{x}_n)}{n} \tag{13}$$

Замечание. $F_n(x)$ обладает всеми свойствами функции распределения. При этом она кусочно-постоянна и принимает значения

$$0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{(n-1)}{n}, 1$$

Замечание. Если все элементы вектора $\vec{x}_n$ различны, то

$$F_n(x) = \begin{cases} 0, & x \leq x_{(1)}; \\ \frac{i}{n}, & x_{(i)} < x \leq x_{(i+1)},\ i = \overline{1, n-1}; \\ 1, & x > x_{(n)}. \end{cases} \tag{14}$$

Листинг программы

Листинг 1: Текст программы

```
function lab1
clear all;

% считываем из файла
X = csvread('data.txt')

% получаем вариационный ряд и его длину
X = sort(X)
n = size(X, 2)

% максимальный и минимальный элементы, размах
X_min = X(1)
X_max = X(end)
R = X_max - X_min

% вычисление оценок mu и S2 математического ожидания MX и дисперсии DX
mu = mean(X)
S2 = var(X)

% группировка значений выборки в m = [log2 n] + 2 интервала
m = floor(log2(n) + 2)
h = R / m
intervals = zeros(m, 2)
```

```
% получаем границы интервалов
for i = 1:m
intervals(i,1) = X(1) + h * i;
end

% заполняем интервалы
i = 1, j = 1
border = X(1) + h

while i < n
if X(i) >= border && border < X(n) && j < m
border = border + h
j = j + 1
continue % на случай, если в интервале нет чисел
end

intervals(j, 2) = intervals(j, 2) + 1
i = i + 1
end

intervals(m, 2) = intervals(m, 2) + 1; % последний элемент

% гистограмма и ф-я плотности
gist = zeros(m, 2)
gist(1, 1) = (X(1) + intervals(1, 1))./2

% интервалы для гистограммы (значения по x)
for i = 2:m
gist(i, 1) = (intervals(i - 1, 1) + intervals(i, 1))./2
end

% значения по y для гистограммы
for i = 1:m
y = intervals(i, 2)
y = y / (n * h)
gist(i,2) = y
end

% функция плотности нормальной случайной величины
f = 1 / sqrt(2 * pi * S2) * exp(-power((X - mu), 2) / (2 * S2))

% график гистограммы и функции плотности
```

```
bar(gist(:,1), gist(:,2), 1)
hold on
plot (X, f, 'r'), grid
figure()

% эмпирическая ф-я плотности и ф-я распределения
empir = zeros(1, n+2)
empir(1) = X(1) - 1

for i = 1:n
empir(i + 1) = X(i)
end

empir(n + 2) = X(n) + 1

% вычисление интервалов для функций
n_emp = length(empir)
Mmin = min(empir)
Mmax = max(empir)
step = (Mmax - Mmin) / n_emp
Xn = Mmin : step : Mmax

% вычисление значений для эмпирической функции
E = zeros(n_emp, 1)

for i = 1:n_emp
count = 0

for j = 1:n
if X(j) <= empir(i)
count = count + 1
end
end

E(i) = count / n
end

% функция распределения нормальной случайной величины
F = 1/2 * (1 + erf((Xn - mu) / sqrt(2 * S2)))

% графики
plot(Xn, F, "--"), grid
```

```
hold on
stairs(empir, E), grid
end
```

Результаты работы и графики

$$M_{max} = -15.270 \quad (15)$$

$$M_{min} = -19.470 \quad (16)$$

$$R = 4.2 \quad (17)$$

$$\hat{\mu} = -17.589 \quad (18)$$

$$S^2 = 0.7286 \quad (19)$$

$$\text{Интервалы: } [-19.470, -18.945)[-18.945, -18.420)[-18.420, -17.895)[-17.895, -17.370)$$
$$[-17.370, -16.845)[-16.845, -16.320)[-16.320, -15.795)[-15.795, -15.270]. \quad (20)$$

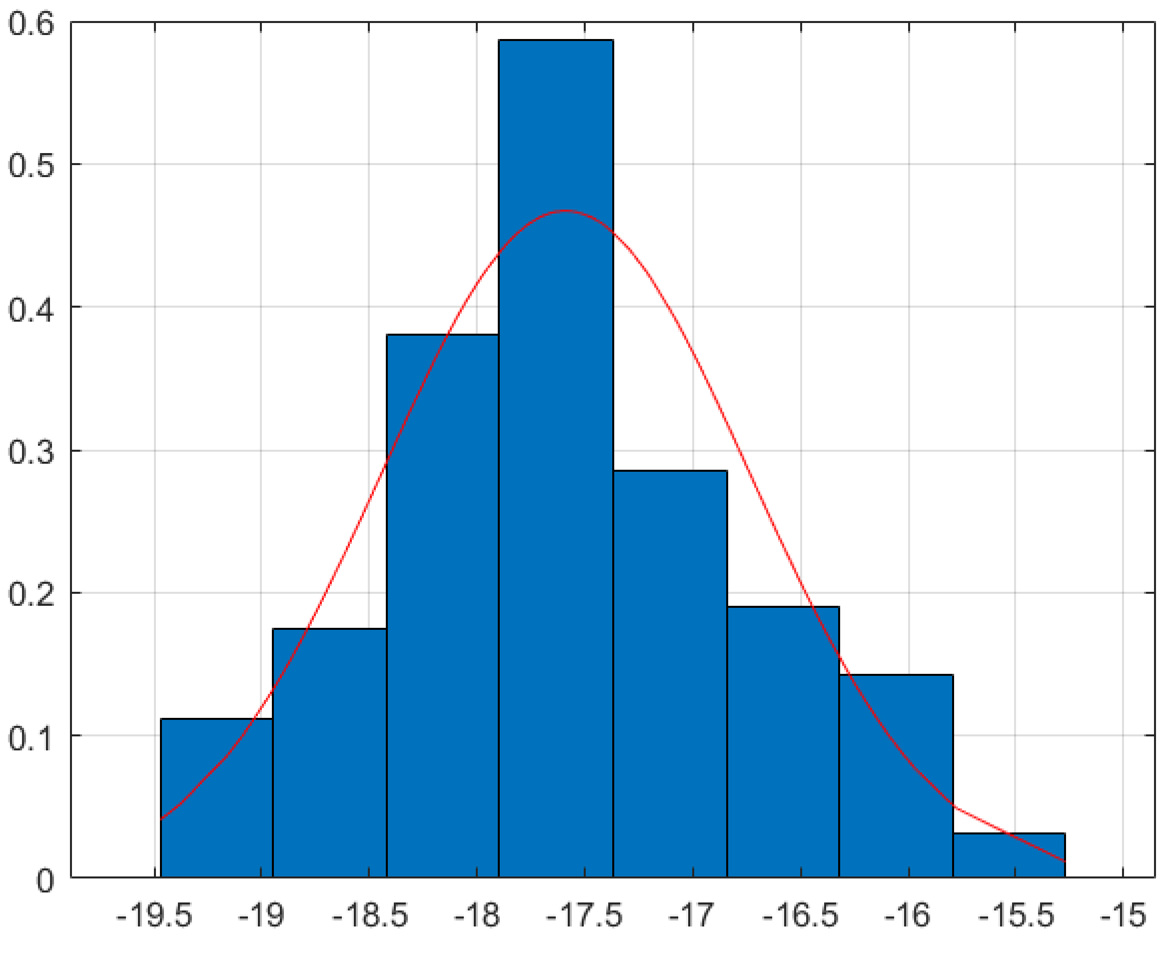

Рис. 1: Гистограмма и график функции плотности распределения вероятностей

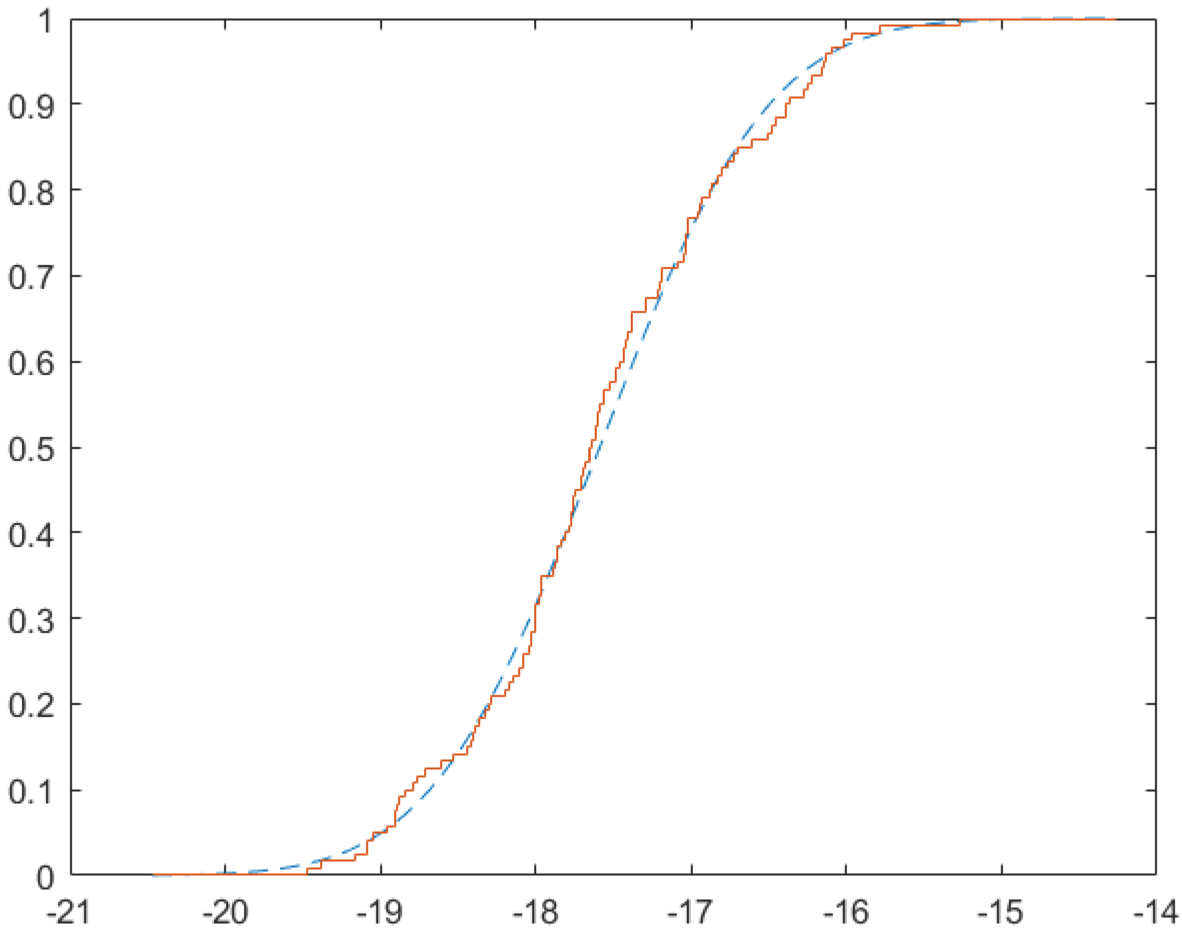

Рис. 2: График эмпирической функции распределения и функции распределения

Литература

1. Власов П.А. - Курс лекций по "Математической статистике"[Текст], Москва 2020 год.
2. Феллер В. Целочисленные величины. Производящие функции // Введение в теорию вероятностей и её приложения = An introduction to probability theory and its applicatons, Volume I second edition / Перевод с англ. Р. Л. Добрушина, А. А. Юшкевича, С. А. Молчанова Под ред. Е. Б. Дынкина с предисловием А. Н. Колмогорова. — 2-е изд. — М.: Мир, 1964. — С. 270—272.