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 13

$\vec{X}$=(-10.82,-9.27,-9.65,-9.36,-9.27,-11.25,-9.89,-9.26,-11.15,-8.90,-11.02,-8.28,-9.18,-10.16,-10.59,-10.82,-9.05,-9.47,-10.98,-11.50,-8.64,-10.86,-10.76,-11.49,-11.09,-9.33,-9.32,-9.66,-8.79,-10.54,-9.12,-10.40,-8.59,-10.22,-9.06,-10.59,-10.60,-10.25,-9.35,-11.44,-9.81,-9.32,-9.95,-9.33,-10.64,-9.45,-10.99,-10.15,-10.39,-10.36,-10.49,-11.67,-10.00,-10.87,-11.11,-9.68,-10.77,-9.13,-8.62,-10.33,-11.36,-10.24,-9.41,-11.05,-10.15,-9.35,-11.45,-9.87,-10.41,-10.11,-10.84,-11.48,-7.77,-10.79,-9.88,-10.70,-9.07,-9.47,-10.15,-9.93,-11.52,-9.04,-10.93,-10.13,-9.56,-11.39,-9.79,-9.19,-11.09,-9.86,-10.67,-10.26,-9.07,-10.53,-11.24,-10.16,-11.33,-8.76,-8.88,-10.53,-10.12,-8.98,-9.84,-9.90,-10.13,-9.32,-9.31,-9.99,-8.55,-11.64,-11.32,-10.51,-11.71,-10.50,-10.50,-12.20,-11.68,-10.45,-7.88,-10.84)

Теоретические сведения

Формулы для вычисления величин

Минимальное и максимальное значения выборки

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

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 = [-10.82,-9.27,-9.65,-9.36,-9.27,-11.25,-9.89,-9.26,-11.15,-8.90,-11.02,-8.28,-9.18,-10.16,-10.59,-10.82,-9.05,-9.47,-10.98,-11.50,
-8.64,-10.86,-10.76,-11.49,-11.09,-9.33,-9.32,-9.66,-8.79,-10.54,-9.12,
-10.40,-8.59,-10.22,-9.06,-10.59,-10.60,-10.25,-9.35,-11.44,-9.81,
-9.32,-9.95,-9.33,-10.64,-9.45,-10.99,-10.15,-10.39,-10.36,-10.49,
-11.67,-10.00,-10.87,-11.11,-9.68,-10.77,-9.13,-8.62,-10.33,-11.36,
-10.24,-9.41,-11.05,-10.15,-9.35,-11.45,-9.87,-10.41,-10.11,-10.84,
-11.48,-7.77,-10.79,-9.88,-10.70,-9.07,-9.47,-10.15,-9.93,-11.52,-9.04,
-10.93,-10.13,-9.56,-11.39,-9.79,-9.19,-11.09,-9.86,-10.67,-10.26,
-9.07,-10.53,-11.24,-10.16,-11.33,-8.76,-8.88,-10.53,-10.12,-8.98,
-9.84,-9.90,-10.13,-9.32,-9.31,-9.99,-8.55,-11.64,-11.32,-10.51,-11.71,
-10.50,-10.50,-12.20,-11.68,-10.45,-7.88,-10.84];

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

- $M_{\max} = -7.7700$
- $M_{\min} = -12.2000$
- $R = 4.4300$
- $\hat{\mu}(\vec{x}_n) = -10.1317$
- $S^2(\vec{x}_n) = 0.8460$
- $m = 8$

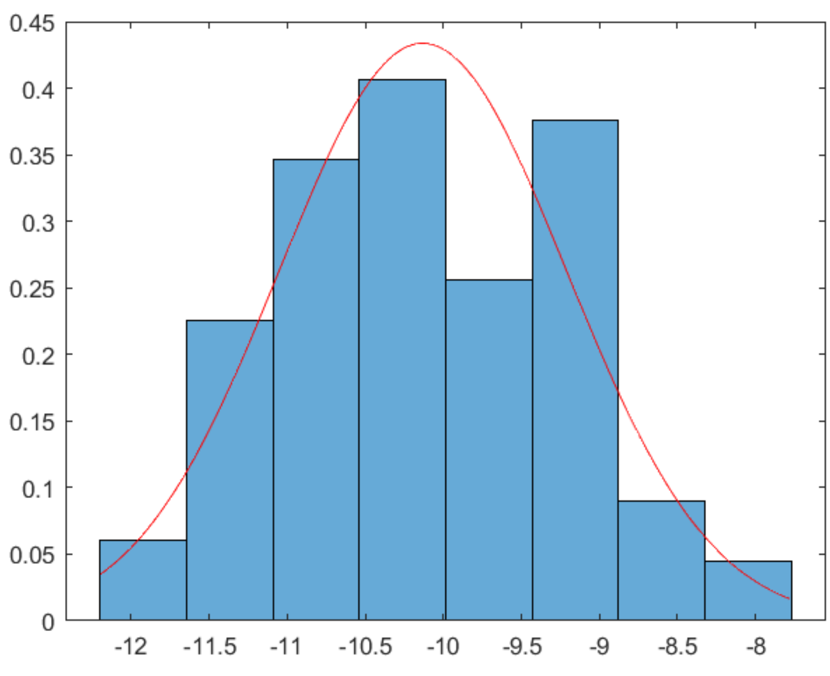

Рис. 1: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочным математическим ожиданием и исправленной выборочной дисперсией

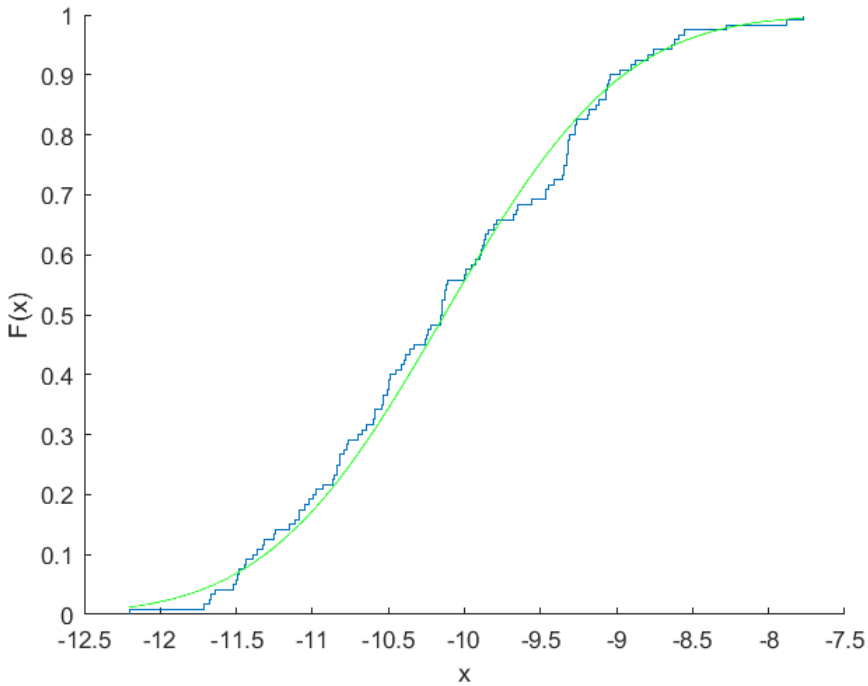

Рис. 2: График эмпирической функции распределения и функции распределения нормальной случайной величины с выборочным математическим ожиданием и исправленной выборочной дисперсией