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 курсу "Математическая статистика"

Тема Гистограмма и эмпирическая функция распределения

Студент Зайцева А. А.

Группа ИУ7-62Б

Преподаватель Власов П. А.

Москва — 2022 г.

Оглавление

1 Задание

Цель работы: построение гистограммы и эмпирической функции распределения.

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.

Провести вычисления и построить графики для выборки из индивидуального варианта.

2 Теоретические сведения

2.1 Формулы для вычисления величин

Минимальное и максимальное значения выборки, соответственно:

$$\begin{aligned} M_{\max} &= X_{(n)} \\ M_{\min} &= X_{(1)} \end{aligned} \tag{2.1}$$

Размах выборки:

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

Оценки математического ожидания и дисперсии, соответственно:

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

2.2 Определение эмпирической плотности и гистограммы

Пусть $\vec{x} = (x_1, ..., x_n)$ – выборка из генеральной совокупности X, вектор $(x_{(1)}, ..., x_{(n)})$ – вариационный ряд, построенный по этой выборке. Если объем n выборки велик, то значения x_i группируют в так называемый интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}; x_{(n)}]$ разбивают на m равновеликих промежутков:

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

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

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

Определение

Интервальным статистическим рядом, отвечающим выборке $\vec{x}$ называют таблицу вида

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

где n_i – количество элементов выборки $\vec{x}$, которые $\in J_i$.
Для выбора m используют формулу $m = [\log_2 n] + 2$ или $m = [\log_2 n] + 1$.

Пусть для данной выборки $\vec{x}$ построен интервальный статистический ряд $(J_i, n_i), i = \overline{1; m}$

Определение

Эмпирической функцией плотности распределения, соответсвующей выборке $\vec{x}$, называют функцию

$$\hat{f}_n(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1; m} \\ 0, \text{иначе.} \end{cases} \tag{2.4}$$

Определение График эмпирической функции плотности называют гистограммой.

2.3 Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, ..., x_n)$ – выборка из генеральной совокупности X. Обозначим $n(t, \vec{x})$ – число компонент вектора $\vec{x}$, которые меньше, чем t.

Определение

Эмпирической функцией распределения, построенной по выборке $\vec{x}$, называют функцию $F_n : R \to R$, определенную правилом

$$F_n(t) = \frac{n(t, \vec{x})}{n} \tag{2.5}$$

3 Результат работы

3.1 Код программы

```
X =
    [7.76,6.34,5.11,7.62,8.84,4.68,8.65,6.90,8.79,6.61,6.62,7.13,6.75,7.28,7.74,7.08,5.57,8.20,7.7

%
M_max = max(X);
M_min = min(X);

%
R = M_max - M_min;

%
MX = mean(X);
DX = var(X); % sigma == std == sqrt(var(arg))

%
m = floor(log2(length(X))) + 2;
h = histogram(X, m);
%disp(h);

%
sigma = std(X);
x = (M_min - 1):(sigma / 100):(M_max + 1);
f = normpdf(x, MX, sigma); % normal probability distribution function
figure;
heights = h.Values / (sum(h.Values) * h.BinWidth);
centers = [];
for i = 1:(length(h.BinEdges) - 1)
    centers = [centers, (h.BinEdges(i + 1) + h.BinEdges(i)) / 2];
end
%disp(centers);
hold on;
bar(centers, heights, 1); %  :)
plot(x, f, 'g', 'LineWidth', 2);

% )
F = normcdf(x, MX, sigma); % normal cumulative distribution function
figure;
hold on;
ecdf(X); % empiric cumulative distribution function
plot(x, F, 'r');
```

3.2 Результаты расчётов

$$M_{\min} = -4{,}33$$

$$M_{\max} = 0{,}01$$

$$R = 4{,}34$$

$$\hat{\mu}(\vec{x}_n) = 2{,}0585$$

$$S^2(\vec{x}_n) = 0{,}994$$

$$m = 8$$

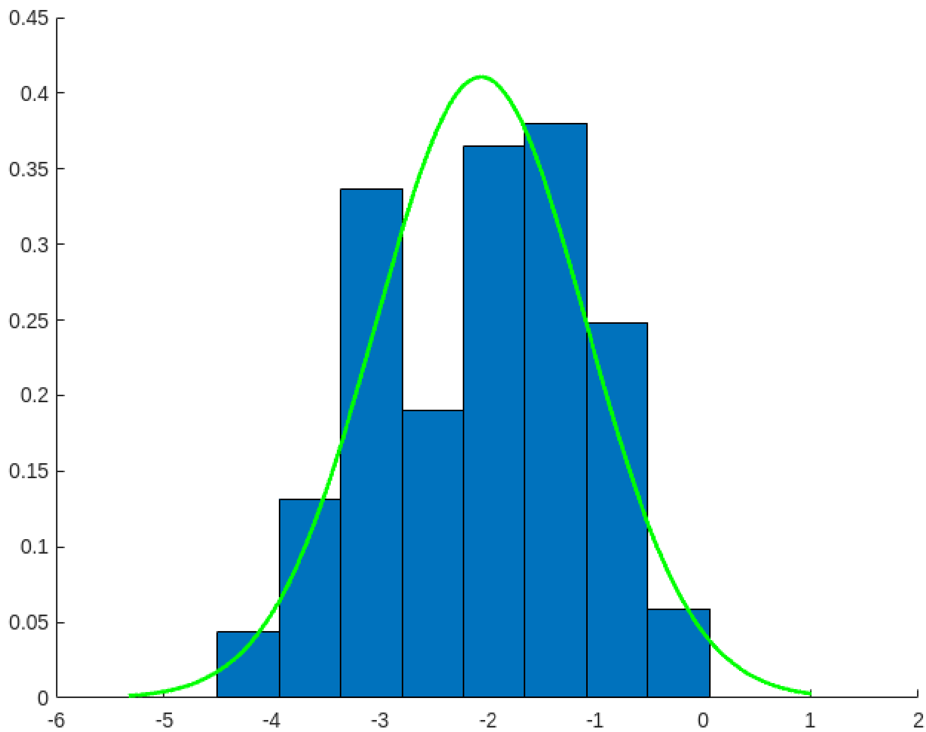

Рис. 3.1: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочными мат. ожиданием и дисперсией

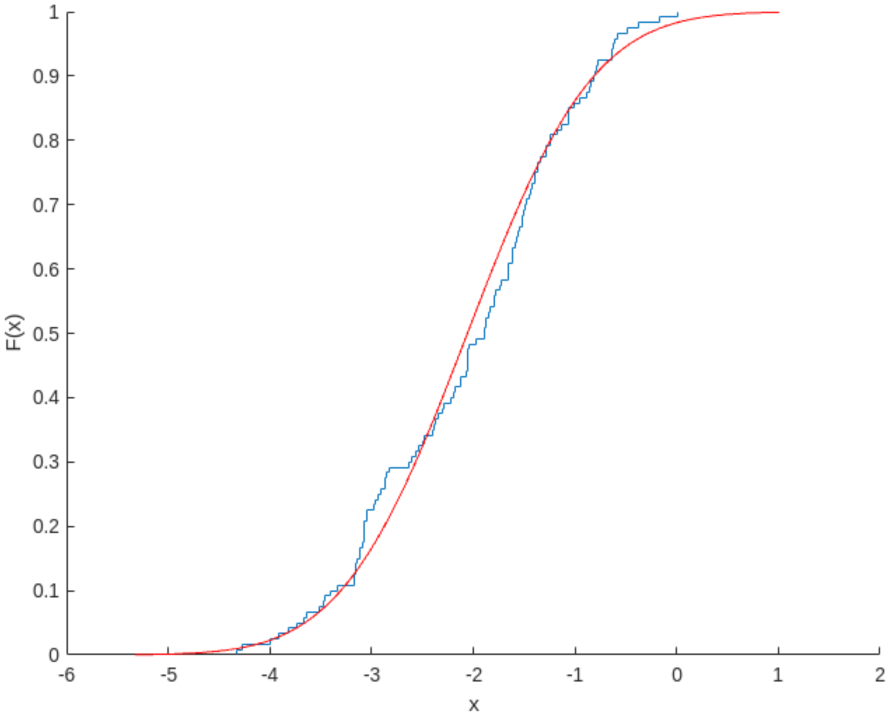

Рис. 3.2: График эмперической функции распределения и функции распределения нормальной случайной величины с выборочными мат. ожиданием и дисперсией