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 Текст 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
n = length(X);


fprintf('\na)\n');
M_max = max(X);
M_min = min(X);
fprintf('M_max = %f\n', M_max);
fprintf('M_min = %f\n', M_min);


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


fprintf('\n)\n');
mu = mean(X);
s_sqr = var(X); %  var(X, 0) (  N-1), =std(X)^2
fprintf('mu = %f\n', mu);
fprintf('s_sqr = %f\n', s_sqr);


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

ni_array = zeros([1 m]);
ai_array = zeros([1 m]);
bi_array = zeros([1 m]);
for i = 1:m-1
    ai_array(i) = M_min + (i - 1) * delta;
    bi_array(i) = M_min + i * delta;
    ni_array(i) = count_ni(i, ai_array(i), bi_array(i), X);
end
ai_array(m) = M_min + (m - 1) * delta;
bi_array(m) = M_max;
ni_array(m) = count_ni(m, ai_array(m), bi_array(i) + 1, X); %  -

fprintf('\n');
```

```
for i = 1:m-1
    fprintf('J%i␣=␣[%f;␣%f)␣␣|␣', i, ai_array(i), bi_array(i));
end
fprintf('J%i␣=␣[%f;␣%f]\n', m, ai_array(m), bi_array(m));
for i = 1:m-1
    fprintf('n%i=%2i␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣|␣', i, ni_array(i));
end
fprintf('n%i=%i\n', m, ni_array(m));



fprintf('\n)\n');
figure;
hold on;
%
h_array = ni_array * (1 / (n * delta));
ai_array(end+1) = M_max;
hist = histogram('BinEdges', ai_array,'BinCounts', h_array); %histogram(X, m);
%
%
sigma = sqrt(s_sqr);
x_array = (M_min - 1):(sigma / 100):(M_max + 1);
f = normpdf(x_array, mu, sigma);
plot(x_array, f, 'r', 'LineWidth', 2);


fprintf('\ne)\n');
figure;
hold on;
%
% ecdf_values = ecdf(X); -
z_array = unique(X); %      X
z_array(end + 1) = z_array(end) + 1;
nt_array = zeros([1 length(z_array)]);
for z_index = 1:length(z_array)
    nt_array(z_index) = count_nt(z_array(z_index), X) / n;
end
stairs(z_array, nt_array, 'g', 'LineWidth', 1);
%
%
F = normcdf(x_array, mu, sigma);
plot(x_array, F, 'r');


function [ni] = count_ni(i, a, b, X)
    ni = 0;
    fprintf('J%i-␣␣␣␣{', i);
    for x_index = 1:length(X)
```

```
        if (X(x_index) >= a) && (X(x_index) < b)
            fprintf('%f, ', X(x_index));
            ni = ni + 1;
        end
    end
    fprintf('}\n');
end

function [nt] = count_nt(t, X)
    nt = 0;
    for x_index = 1:length(X)
        if (X(x_index) < t)
            nt = nt + 1;
        end
    end
end
```

3.2 Результаты расчетов

$$M_{\min} = 4.39; M_{\max} = 9.89$$

$$R = 5.5$$

$$\hat{\mu}(\vec{x}_n) = 6.9445; S^2(\vec{x}_n) = 1.171956$$

$$m = 8, \Delta = 0.6875$$

Таблица 3.1: Интервальный статистический ряд

J1 = [4.39; 5.0775)	J2 = [5.0775; 5.765)	J3 = [5.765; 6.4525)	J4 = [6.4525; 7.14)	J5 = [7.14; 7.8275)	J6 = [7.8275; 8.515)	J7 = [8.515; 9.2025)	J8 = [9.2025; 9.89]
5	12	19	38	24	10	10	2

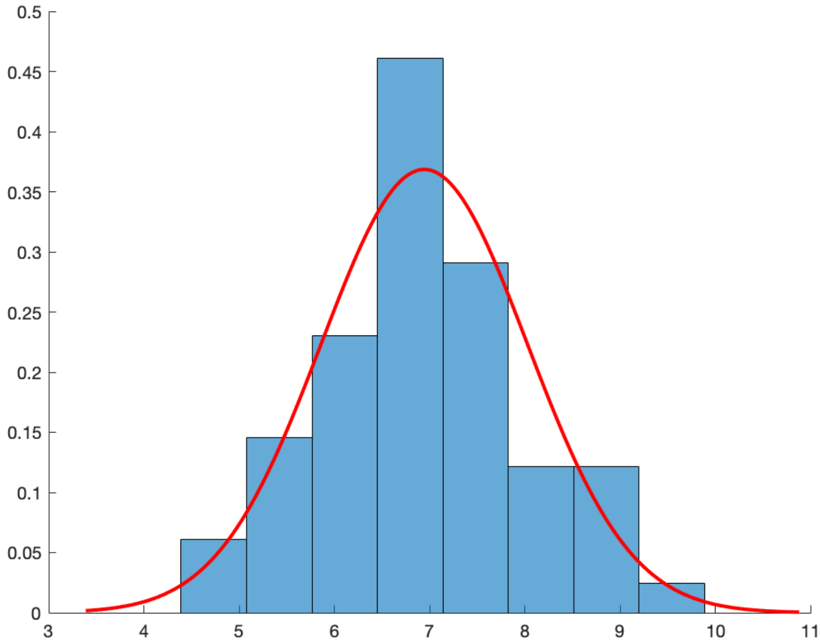

Рис. 3.1: Гистограмма (синим) и график функции плотности распределения вероятностей нормальной случайной величины с выборочными математическим ожиданием $\hat{\mu}$ и дисперсией S^2 (красным)

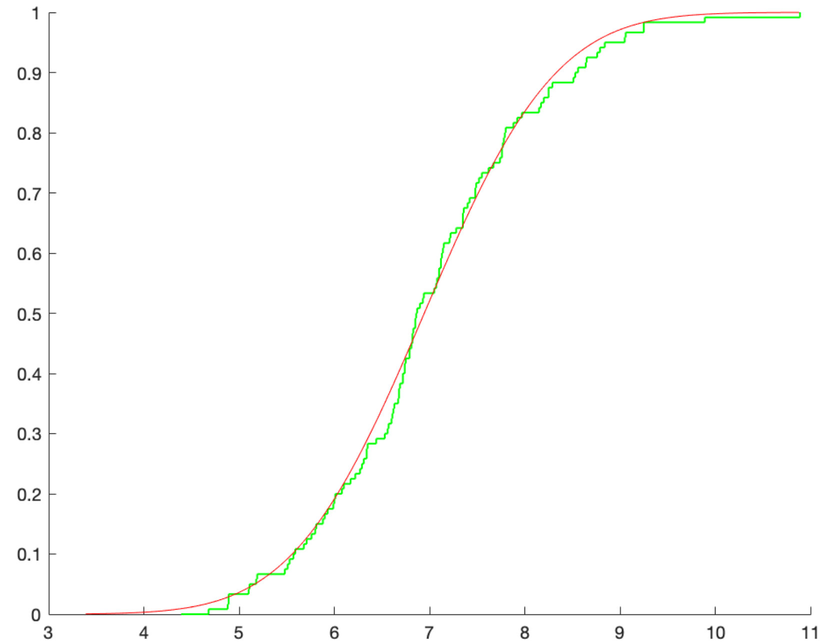

Рис. 3.2: График эмпирической функции распределения (зеленым) и функции распределения нормальной случайной величины с выборочными математическим ожиданием $\hat{\mu}$ и дисперсией S^2 (красным)