



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»

«Гистограмма и эмпирическая функция распределения»

Студент _____ Маслова Марина Дмитриевна

Группа _____ ИУ7-63Б

Оценка (баллы) _____

Преподаватель _____ Власов Павел Александрович

2022 г.

1 Задание

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 выборки¹ из индивидуального варианта.

2 Теоретическая часть

2.1 Формулы для вычисления величин

Пусть $\vec{x} = (x_1, \dots, x_n)$ — выборка объема n из генеральной совокупности X .

Максимальное и минимальное значения выборки:

$$M_{max} = x_{(n)} = \max\{x_1, \dots, x_n\}, \quad (2.1)$$

$$M_{min} = x_{(1)} = \min\{x_1, \dots, x_n\}, \quad (2.2)$$

где $x_{(1)}, x_{(n)}$ — крайние члены вариационного ряда, отвечающего выборке \vec{x} .

Размах выборки:

$$R = M_{max} - M_{min}. \quad (2.3)$$

Оценка математического ожидания — выборочное среднее:

$$\hat{\mu}(\vec{x}) = \frac{1}{n} \sum_{i=1}^n x_i. \quad (2.4)$$

Оценка дисперсии — исправленная выборочная дисперсия:

$$S^2(\vec{x}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (2.5)$$

где $\bar{x} = \hat{\mu}$.

2.2 Определение эмпирической плотности и гистограммы

Если объем выборки достаточно велик ($n > 50$), то элементы выборки группируют в так называемый интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ разбивают на m равновеликих промежутков. Ширина каждого из них:

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}. \quad (2.6)$$

Далее полагают:

$$\begin{aligned} J_i &= [x_{(1)} + (i-1)\Delta; x_{(1)} + i\Delta], \quad i = \overline{1, m-1}, \\ J_m &= [x_{(1)} + (m-1)\Delta; x_{(n)}]. \end{aligned} \quad (2.7)$$

Определение. Интервальным статистическим рядом, отвечающим выборке \vec{x} , называется таблица вида:

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

Здесь n_i — число элементов выборки \vec{x} , попавших в промежуток J_i , $i = \overline{1, m}$.

Количество интервалов определяется формулой:

$$m = [\log_2 n] + 2. \quad (2.8)$$

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

Определение. Эмпирической плотностью распределения (соответствующей выборке \vec{x}) называется функция:

$$f(x) = \begin{cases} \frac{n_i}{n\Delta}, & x \in J_i, \quad i = \overline{1, m}, \\ 0, & \text{иначе.} \end{cases} \quad (2.9)$$

Определение. График эмпирической функции плотности называется гистограммой.

2.3 Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, \dots, x_n)$ — 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 $F_n : \mathbb{R} \rightarrow \mathbb{R}$, определенную правилом:

$$F_n(t) = \frac{n(t, \vec{x})}{n}. \quad (2.10)$$

3 Практическая часть

Листинг 3.1 – Текст программы

```
1 x = [11.89,9.60,9.29,10.06,9.50,8.93,9.58,6.81,8.69,9.62,...
2     9.01,10.59,10.50,11.53,9.94,8.84,8.91,6.90,9.76,7.09,...
3     11.29,11.25,10.84,10.76,7.42,8.49,10.10,8.79,11.87,8.77,...
4     9.43,12.41,9.75,8.53,9.72,9.45,7.20,9.23,8.93,9.15,...
5     10.19,9.57,11.09,9.97,8.81,10.73,9.57,8.53,9.21,10.08,...
6     9.10,11.03,10.10,9.47,9.72,9.60,8.21,7.78,10.21,8.99,...
7     9.14,8.60,9.14,10.95,9.33,9.98,9.09,10.35,8.61,9.35,...
8     10.04,7.85,9.64,9.99,9.65,10.89,9.08,8.60,7.56,9.27,...
9     10.33,10.09,8.51,9.86,9.24,9.63,8.67,8.85,11.57,9.85,...
10    9.27,9.69,10.90,8.84,11.10,8.19,9.26,9.93,10.15,8.42,...
11    9.36,9.93,9.11,9.07,7.21,8.22,9.08,8.88,8.71,9.93,...
12    12.04,10.41,10.80,7.17,9.00,9.46,10.42,10.43,8.38,9.01];
13
14 n = length(x);
15
16 fprintf('Задания а-в\n');
17
18 Mmax = max(x);
19 Mmin = min(x);
20 R = Mmax - Mmin;
21 mu = sum(x) / n;
22 sqrS = sum((x - mu) .^ 2) / (n - 1);
23
24 fprintf('M_max = %6.3f\n', Mmax);
25 fprintf('M_min = %6.3f\n', Mmin);
26 fprintf('R = %5.3f\n', R);
27 fprintf('mu = %5.3f\n', mu);
28 fprintf('S^2 = %5.3f\n', sqrS);
29
30 fprintf('\nЗадание г\n');
31
32 m = floor(log2(n)) + 2;
33 limit_points = linspace(Mmin, Mmax, m + 1);
34
35 nums = zeros(m, 1);
36
37 for i = 1:m
38     left = limit_points(i);
39     right = limit_points(i + 1);
40
41     for j = 1:n
42         if x(j) >= left && x(j) < right
43             nums(i)++;
44         endif
```

```

45     endfor
46 endfor
47
48 nums(m)++;
49
50 fprintf('m = %d\n', m);
51 for i = 1:m
52     fprintf('В интервале [%6.3f, %6.3f%c %2d элементов.\n', limit_points(i),...
53         limit_points(i + 1), ifelse(i == m, ']', ')'), nums(i));
54 endfor
55
56 fprintf('\nЗадание д\n');
57 fprintf('График в окне "Задание д"\n');
58
59 centers = zeros(m, 1);
60 f_emp   = zeros(m, 1);
61
62 Delta = (Mmax - Mmin) / m;
63
64 for i = 1:m
65     centers(i) = (limit_points(i) + limit_points(i + 1)) / 2;
66     f_emp(i)   = nums(i) / (n * Delta);
67 endfor
68
69 sigma = sqrt(sqrS);
70 x_vals = (Mmin:1e-3:Mmax);
71 f_theor = normpdf(x_vals, mu, sigma);
72
73 h = figure();
74 set(h, 'numbertitle', 'off', 'name', 'Задание д');
75 hold on;
76 bar(centers, f_emp, 1, 'facecolor', 'r');
77 plot(x_vals, f_theor, 'linewidth', 2, 'color', 'b');
78 grid;
79
80 fprintf('\nЗадание е\n');
81 fprintf('График в окне "Задание е"\n');
82
83 uniq_x = unique(x);
84 uniq_n = length(uniq_x);
85 uniq_nums = histc(x, uniq_x);
86
87 t = zeros(uniq_n + 2, 1);
88 t(1) = Mmin - 1;
89 t(uniq_n + 2) = Mmax + 1;
90
91 for i = 2:uniq_n+1
92     t(i) = uniq_x(i - 1);

```

```

93 endfor
94
95 F_emp = zeros(uniq_n + 1, 1);
96 cnt = 0;
97
98 for i = 1:uniq_n
99     F_emp(i) = cnt / n;
100     cnt += uniq_nums(i);
101 endfor
102
103 F_emp(uniq_n + 1) = cnt / n;
104 F_emp(uniq_n + 2) = cnt / n;
105
106 x_vals = (Mmin-1:1e-3:Mmax+1);
107 F_theor = normcdf(x_vals, mu, sigma);
108
109 g = figure();
110 set(g, 'numbertitle', 'off', 'name', 'Задание е');
111 hold on;
112 plot(x_vals, F_theor, 'linewidth', 4, 'color', 'b');
113 stairs(t, F_emp, 'linewidth', 2, 'color', 'r');
114 grid;

```

Листинг 3.2 – Результаты расчетов для выборки из индивидуального варианта

```

1 Задания а-в
2 M_max = 12.410
3 M_min = 6.810
4 R = 5.600
5 mu = 9.487
6 S^2 = 1.217
7
8 Задание г
9 m = 8
10 В интервале [ 6.810, 7.510) 7 элементов.
11 В интервале [ 7.510, 8.210) 4 элементов.
12 В интервале [ 8.210, 8.910) 21 элементов.
13 В интервале [ 8.910, 9.610) 36 элементов.
14 В интервале [ 9.610, 10.310) 27 элементов.
15 В интервале [10.310, 11.010) 14 элементов.
16 В интервале [11.010, 11.710) 7 элементов.
17 В интервале [11.710, 12.410] 4 элементов.
18
19 Задание д
20 График в окне "Задание д"
21
22 Задание е
23 График в окне "Задание е"

```

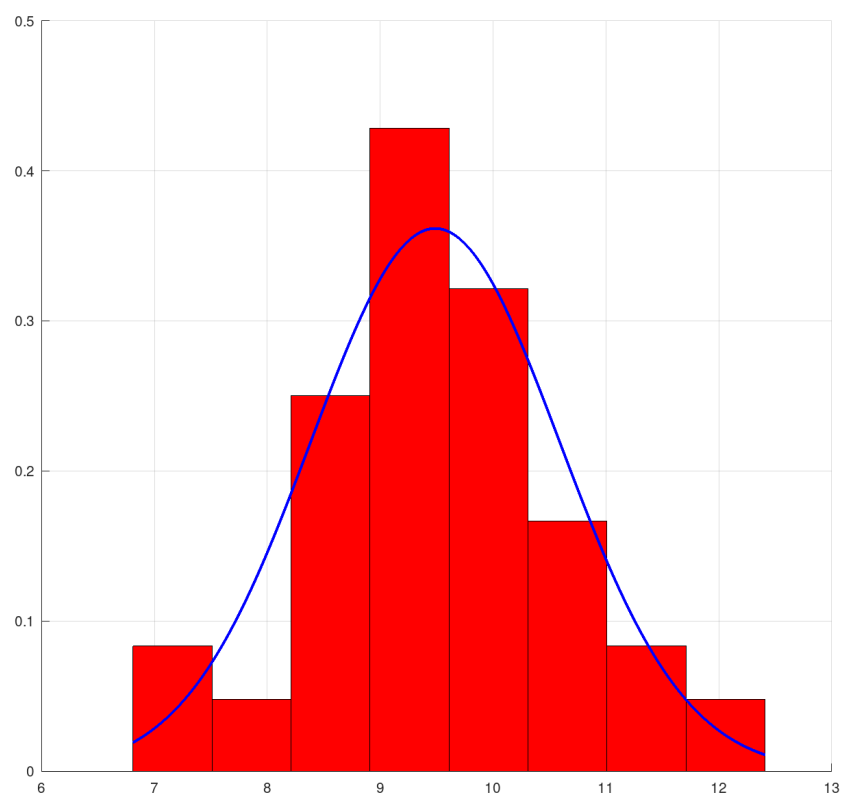


Рисунок 3.1 – Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

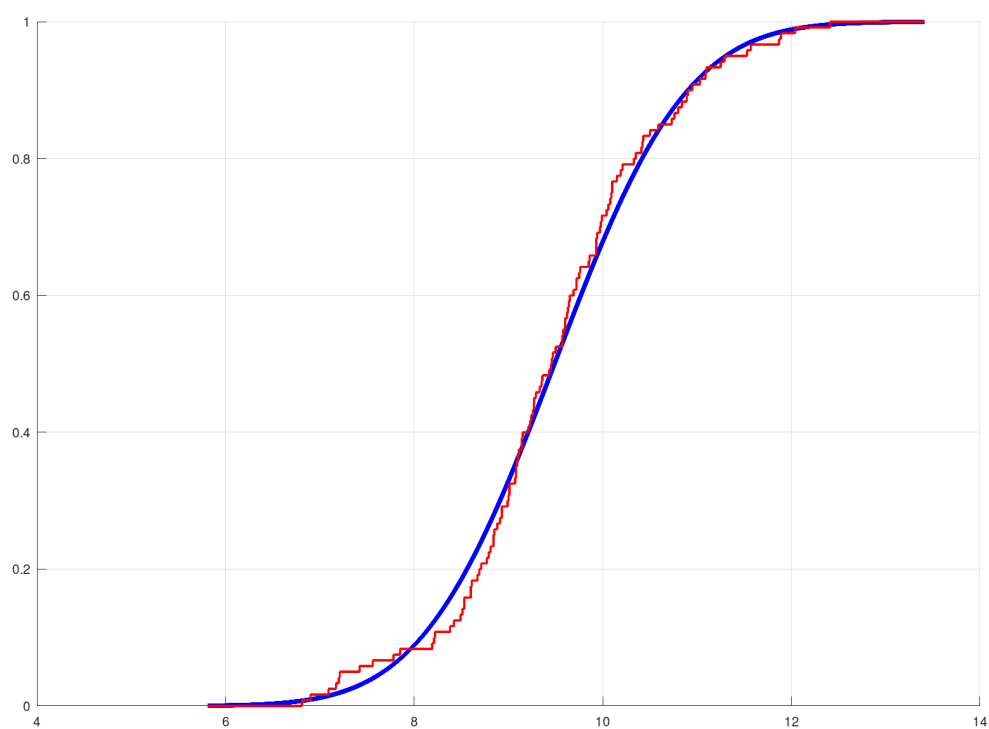


Рисунок 3.2 – График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2