



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 Цветков И.А.

Группа ИУ7-63Б

Оценка (баллы) _____

Преподаватель Власов П. А.

Москва — 2022 г.

1 Задание на лабораторную работу

Цель работы: построение гистограммы и эмпирической функции распределения.

1.1 Содержание работы

1. Для выборки объема 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.

2 Теоретическая часть

2.1 Формулы для вычисления величин

Максимальное M_{\max} и минимальное M_{\min} значения выборки:

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

Размах R выборки:

$$R = M_{\max} - M_{\min}. \quad (2.2)$$

Оценки математического ожидания и дисперсии:

- выборочное среднее:

$$\hat{\mu}(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.3)$$

- исправленная выборочная дисперсия:

$$S^2(\vec{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (2.4)$$

где $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

2.2 Определение эмпирической плотности и гистограммы

Интервальный статистический ряд

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

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

J_1	J_2	\dots	J_m
n_1	n_2	\dots	n_m

$$J_i = [x_{(1)} + (i - 1) \cdot \Delta, x_{(1)} + i \cdot \Delta), i = \overline{1; m - 1} \quad (2.5)$$

Причем:

$$J_m = [x_{(1)} + (m - 1) \cdot \Delta, x_{(n)}] \quad (2.6)$$

Величина Δ при этом равна:

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}, \quad (2.7)$$

где $m = [\log_2 n] + 2$, $x_{(n)} = \max(\vec{x})$, $x_{(1)} = \min(\vec{x})$.

Эмпирическая плотность

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:

$$f_n(x) = \begin{cases} \frac{n_i}{n \cdot \Delta}, x \in J_i, i = \overline{1; m} \\ 0, \text{ иначе} \end{cases} \quad (2.8)$$

Гистограмма

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

$$F_n : R \rightarrow R, \tag{2.9}$$

определенную правилом:

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

3 Практическая часть

3.1 Код программы

```
1 function lab_01()
2     X = csvread('data_test.csv');
3
4     X = sort(X);
5     n = length(X);
6
7     fprintf("\nn = %g\n", n);
8
9     fprintf("\nВыполнение заданий.\n\n");
10
11     % Задание 1
12     fprintf("\nn1) Вычислить максимальное и минимальное значение:\n", n);
13
14     % Минимальное значение
15     minX = min(X);
16     fprintf("\nМинимальное значение = %.4f", minX);
17
18     % Максимальное значение
19     maxX = max(X);
20     fprintf("\nМаксимальное значение = %.4f\n", maxX);
21
22
23     % Задание 2
24     fprintf("\nn2) Вычислить размах R:\n", n);
25
26     % Размах R
27     R = maxX - minX;
28     fprintf("\nR = %.4f\n", R);
29
30
31     % Задание 3
32     fprintf("\nn3) Вычислить оценки математического ожидания и дисперсии:\n", n);
33     % Выборочное среднее
34     mu = sum(X) / n;
35     fprintf("\nВыборочное среднее (оценка мат ожидания) = %.4f\n", mu);
36
37     % Исправленная выборочная дисперсия
38     s_quad = sum((X - mu).^2) / (n - 1);
39     fprintf("Исправленная выборочная дисперсия (оценка дисперсии) = %.4f\n",
40             s_quad);
41
```

```

42 % Задание 4
43 fprintf("\n4) Группировка значений выборки в  $m = \lceil \log_2 n \rceil + 2$  интервала
    :\n\n", n);
44
45 % Вычисление  $m = \lceil \log_2 n \rceil + 2$  промежутков
46 m = floor(log2(n)) + 2;
47 fprintf("Интервалов m = %3d\n", m);
48
49 % Ширина интервала
50 delta = (X(n) - X(1)) / m;
51
52 % Поиск интервалов
53 % Интервалы хранятся в виде матрицы m на 2
54 intervals = zeros(m, 3);
55
56 % Вычисление границ интервалов
57 for ind = 1 : m - 1
58     intervals(ind, 1) = X(1) + (ind - 1) * delta;
59     intervals(ind, 2) = X(1) + ind * delta;
60 endfor
61
62 % Вычисление границ интервалов для m
63 intervals(m, 1) = X(1) + (m - 1) * delta;
64 intervals(m, 2) = X(n);
65
66 % Вычисление кол-ва значений в интервалах
67 indX = 1;
68 indIntervals = 1;
69
70 %% Задаются текущие границы интервала
71 leftBorder = intervals(indIntervals, 1);
72 rightBorder = intervals(indIntervals, 2);
73
74 %% Цикл по всем элементам выборки
75 while indX <= n
76
77     %% Если элемент попал в границы, то инкремент для данного интервала
78     if X(indX) >= leftBorder && X(indX) < rightBorder
79         intervals(indIntervals, 3) = intervals(indIntervals, 3) + 1;
80     %% Если элемент не попал в интервал
81     else
82         %% Элемент, не попавший в предыдущий интервал, должен быть
83         %% обработан для следующего интервала
84         indX == 1;
85
86         %% Для крайнего элемента выборки, который должен попасть в текущий и
87         %% нтервал
88         if indIntervals + 1 > m

```

```

88         intervals(indIntervals , 3) = intervals(indIntervals , 3) + 1;
89         break;
90     endif
91
92     %% Следующий интервал
93     indIntervals = indIntervals + 1;
94
95     %% Обновление границ текущего интервала
96     leftBorder = rightBorder;
97     rightBorder = intervals(indIntervals , 2);
98
99     endif
100
101     %% Следующий элемент выборки
102     indX = indX + 1;
103 endwhile
104
105 % Вывести полученные интервалы
106 for i = 1 : m
107     if i == m
108         fprintf("%3d. [ %3.4f ; %3.4f] : Элементов: %3d\n", i ,
109                                     intervals(i , 1) ,
110                                     intervals(i , 2) ,
111                                     intervals(i ,
112                                             3));
113
114     else
115         fprintf("%3d. [ %3.4f ; %3.4f) : Элементов: %3d\n", i ,
116                                     intervals(i , 1) ,
117                                     intervals(i , 2) ,
118                                     intervals(i ,
119                                             3));
120
121     endif
122 endfor
123
124 % Задание 5
125 fprintf("\n5) Построить гистограмму и график функции плотности \
126 \n распределения вероятностей нормальной случайной величины с \
127 \n математическим ожиданием  $\mu$  и  $\sigma^2$ \n\n");
128
129 fprintf("Результат в отдельном окне\n");
130
131 % Построение гистограммы
132
133 gistData = zeros(m, 2);
134
135 % Середины интервалов

```



```

134     for ind = 1 : m
135         gistData(ind , 1) = (intervals(ind , 1) + intervals(ind , 2)) / 2;
136     endfor
137
138     % Значения столбцов
139     %  $k = n_i / (n * \Delta)$ 
140     for ind = 1 : m
141         gistData(ind , 2) = intervals(ind , 3) / (n * delta);
142     endfor
143
144     % График функции плотности распределения
145
146     % Набор значений по X
147     % начало:шаг:конец
148     xGraph = (minX - 1):1e-3:(maxX + 1);
149
150     sigma = sqrt(s_quad);
151     % Значения функции плотности распределения
152     % normpdf – функция плотности нормального распределения
153     fNormal = normpdf(xGraph , mu, sigma);
154
155     % Отрисовка графиков
156     %% Отрисовка гистограммы
157     bar(gistData(:, 1), gistData(:, 2), 1);
158
159     %% Чтобы следующий график не стер предыдущий
160     hold on;
161
162     %% Отрисовка графика плотности нормального распределения
163     plot(xGraph , fNormal , 'r' , 'LineWidth' , 1.5);
164     %% Сетка
165     grid;
166
167
168     % Задание 6
169     fprintf("\\n6) Построить график эмпирической функции распределения и\\
170         \\n функции распределения нормальной случайной величины с \\
171         \\n математическим ожиданием  $\mu$  и  $s\_quad$ \\n\\n");
172
173     fprintf("Результат в отдельном окне\\n");
174
175     % Набор значений для эмпирической функции распределения
176     tSet = zeros(1, n + 2);
177
178
179     tSet(1) = X(1) - 1;
180
181     for ind = 2 : n + 1

```

```

182         tSet(ind) = X(ind - 1);
183     endfor
184
185     tSet(n + 2) = X(n) + 1;
186
187     % Значения эмпирической функции распределения
188     nEmperic = length(tSet);
189     Femperic = zeros(nEmperic, 1);
190
191     for i = 1 : nEmperic
192         cnt = 0;
193
194         for j = 1: n
195             %  $x(j - 1) < t \leq x(j)$ 
196             if X(j) <= tSet(i)
197                 cnt = cnt + 1;
198             endif
199         endfor
200
201         Femperic(i) = cnt / n;
202     endfor
203
204
205     % Набор значений по X
206     % начало:шаг:конец
207     xGraph = (minX - 1):1e-3:(maxX + 1);
208
209     sigma = sqrt(s_quad);
210     % Значения функции распределения
211     % normpdf – функция нормального распределения
212     Fnormal = normcdf(xGraph, mu, sigma);
213
214
215     % Отрисовка графиков
216     %% Чтобы графики были в отдельном окне
217     figure();
218     %% Отрисовка эмпирической функции распределения
219     stairs(tSet, Femperic);
220
221     %% Чтобы следующий график не стер предыдущий
222     hold on;
223
224     %% Отрисовка графика нормального распределения
225     plot(xGraph, Fnormal, 'r', 'LineWidth', 1.5);
226     %% Сетка
227     grid;
228 endfunction

```

3.2 Результат работы программы

Вывод программы

```
1 Выполнение заданий.
2
3
4 1) Вычислить максимальное и минимальное значение:
5
6 Минимальное значение = -17.2900
7 Максимальное значение = -12.9100
8
9 2) Вычислить размах R:
10
11 R = 4.3800
12
13 3) Вычислить оценки математического ожидания и дисперсии:
14
15 Выборочное среднее (оценка мат ожидания) = -15.2209
16 Исправленная дисперсия (оценка дисперсии) = 0.9680
17
18 4) Группировка значений выборки в  $m = \lceil \log_2 n \rceil + 2$  интервала:
19
20 Интервалов m = 8
21
22 1. [ -17.2900 ; -16.7425) : Элементов: 9
23 2. [ -16.7425 ; -16.1950) : Элементов: 13
24 3. [ -16.1950 ; -15.6475) : Элементов: 19
25 4. [ -15.6475 ; -15.1000) : Элементов: 22
26 5. [ -15.1000 ; -14.5525) : Элементов: 21
27 6. [ -14.5525 ; -14.0050) : Элементов: 23
28 7. [ -14.0050 ; -13.4575) : Элементов: 11
29 8. [ -13.4575 ; -12.9100] : Элементов: 2
30
31 5) Построить гистограмму и график функции плотности
32    распределения вероятностей нормальной случайной величины с
33    математическим ожиданием  $\mu$  и  $s\_quad$ 
34
35 Результат в отдельном окне
36
37 6) Построить график эмпирической функции распределения и
38    функции распределения нормальной случайной величины с
39    математическим ожиданием  $\mu$  и  $s\_quad$ 
40
41 Результат в отдельном окне
```

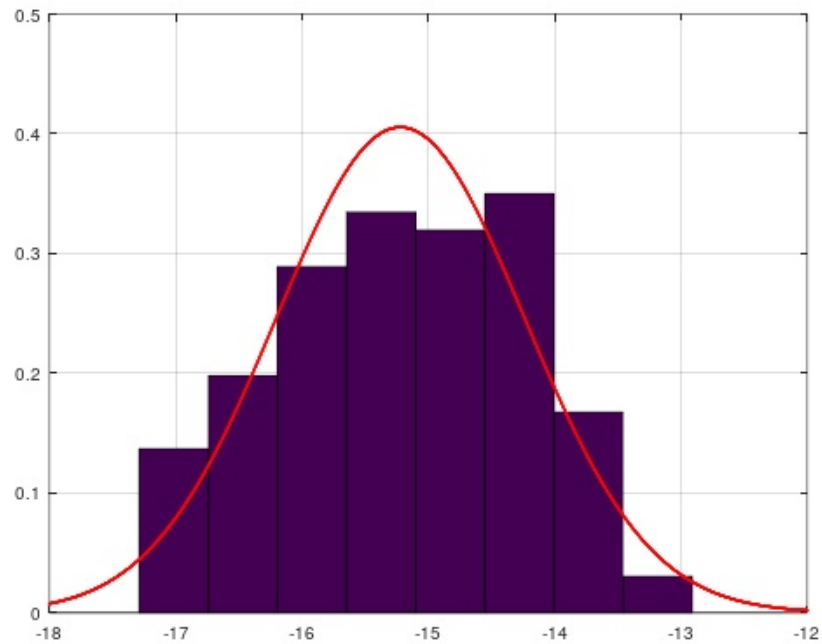


Рисунок 3.1 – Гистограмма и график функции плотности распределения нормальной случайной величины с выборочными математическим ожиданием и дисперсией

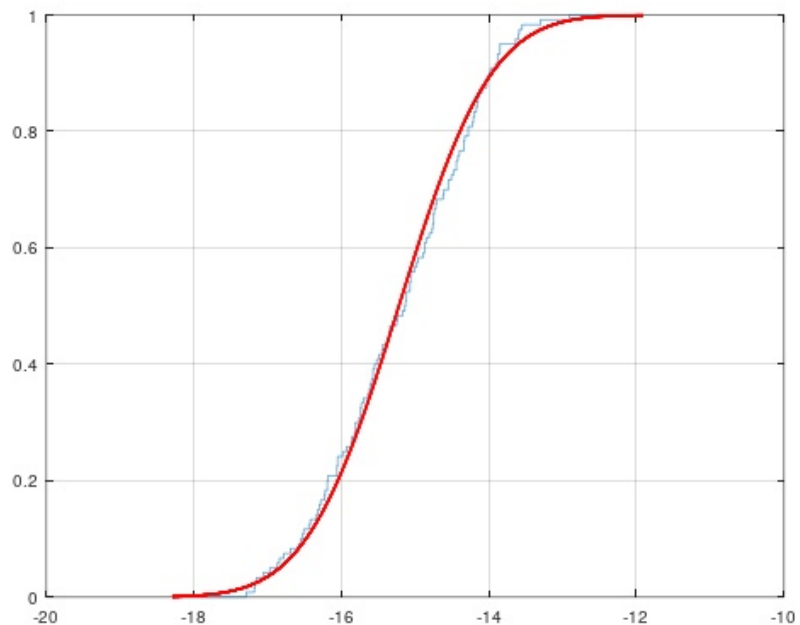


Рисунок 3.2 – График эмпирической функции распределения и функции распределения нормальной случайной величины с выборочными математическим ожиданием и дисперсией