



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»

Лабораторная работа № 1

Тема Гистограмма и эмпирическая функция распределения

Студент Белоусова Ю.С.

Группа ИУ7-61Б

Оценка (баллы) _____

Преподаватель Саркисян П.С.

Москва, 2020

Формулы для вычисления

$\vec{x} = (x_1, \dots, x_n)$ - реализация случайной выборки

- 1) Максимальное значение выборки: $M_{\max} = \max(x_1, \dots, x_n)$
- 2) Минимальное значение выборки: $M_{\min} = \min(x_1, \dots, x_n)$
- 3) Размах выборки: $R = M_{\max} - M_{\min}$
- 4) Оценка MX (выборочное среднее): $\hat{\mu}(\vec{x}) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- 5) Оценка DX (выборочная дисперсия): $S^2(\vec{x}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Определения

- 1) Эмпирической плотностью (отвечающей выборке \vec{x}) называют функцию $\hat{f}_n(x) = \begin{cases} \frac{n_i}{n \Delta}, & x \in J_i, i \in \overline{1, p} \\ 0, & \text{иначе} \end{cases}$, где

$$\begin{aligned} p &= \lfloor \log_2 n \rfloor + 1, & J_i &= [a_{i-1}, a_i), \quad J_p = [a_{p-1}, a_p] \\ \Delta &= \frac{|J|}{p} = \frac{x_{(p)} - x_{(1)}}{p}; & i &= \overline{1, p-1} \\ & & a_i &= x_{(i)} + i \Delta, \quad i = \overline{0, p} \end{aligned}$$

n_i - число элементов выборки, принадлежащих J_i .

- 2) Гистограммой называется график эмпирической плотности.

- 3) Эмпирической функцией распределения называют функцию $F_n: \mathbb{R} \rightarrow \mathbb{R}$, определенную условием $F_n(x) = \frac{n(x, \vec{x})}{n}$, где $n(x, \vec{x})$ - число элементов вектора \vec{x} , которые имеют значение меньше x .

Примеч. если все элементы выборки \vec{x} попарно различны, то

$$F_n(x) = \begin{cases} 0, & x \leq x_{(1)} \\ \frac{i}{n}, & x_{(i)} < x \leq x_{(i+1)}, \quad i = \overline{1, n-1} \\ 1, & x > x_{(n)} \end{cases}$$

Текст программы

```
1. function lab1()
2.     x = importdata('data1.txt');
3.     x = sort(x);
4.
5.     Mmax = x(end);
6.     fprintf("Mmax = %.2f \n", Mmax);
7.
8.     Mmin = x(1);
9.     fprintf("Mmin = %.2f \n", Mmin);
10.
11.     R = Mmax - Mmin;
12.     fprintf("R = %.2f \n", R);
13.
14.     mu = find_expected_value(x);
15.     fprintf("mu = %.4f \n", mu);
16.
17.     S2 = find_dispersion(x);
18.     fprintf("S2 = %.4f \n", S2);
19.
20.     m = find_n_intervals(x);
21.     fprintf("m = %d \n", m);
22.
23.     intervals_table = group(x, m);
24.     for i=1:m-1
25.         fprintf("[%5.2f;%5.2f) ", intervals_table(1,i), intervals_table(1,i+1));
26.     end
27.     fprintf("[%5.2f;%5.2f]\n",intervals_table(1,m), Mmax);
28.     for i=1:m
29.         fprintf("%13d ", intervals_table(2,i));
30.     end
31.     fprintf("\n\n");
32.
33.     grmin = -4;
34.     grmax = 4;
```

```

35. figure(1);
36. grid;
37. hold on;
38. hist_and_f(x,mu,S2,m,intervals_table,grmin,grmax)
39.
40. figure(2);
41. grid;
42. hold on;
43. empir_and_F(x,mu,S2,grmin,grmax)
44.end
45. % оценка мат ожидания
46.function mu = find_expected_value(x)
47.    n = length(x);
48.    mu = sum(x)/n;
49.end
50. % оценка дисперсии
51.function S2 = find_dispersion(x)
52.    n = length(x);
53.    mu = find_expected_value(x);
54.    S2 = sum((x - mu).^2) / n;
55.end
56. % кол-во интервалов
57.function m = find_n_intervals(x)
58.    n = length(x);
59.    m = floor(log2(n)) + 2;
60.end
61. % группировка значений выборки в m интервалов
62.function intervals_table = group(x, m)
63.    n = length(x);
64.    intervals_table = zeros(2,m);
65.    delta = (x(end) - x(1))/m;
66.    for i = 0:m-1
67.        intervals_table(1, i+1) = x(1) + delta*i;
68.    end
69.

```

```

70. count = 0;
71. for i = 1:n
72.     for j = 1:m-1
73.         if intervals_table(1,j) <= x(i) && x(i) < intervals_table(1, j+1)
74.             intervals_table(2, j) = intervals_table(2, j) + 1;
75.             count = count + 1;
76.             break;
77.         end
78.     end
79. end
80. intervals_table(2, m) = n - count;
81. end
82. % построение гистограммы и графика плотности
83. function hist_and_f(x,mu,S2,m,intervals_table,grmin,grmax)
84.     n = length(x);
85.     delta = (x(n)- x(1))/(n-1);
86.
87.     graph = zeros(2,n+2);
88.     graph(1,1) = grmin;
89.     graph(2,1) = 0;
90.     for i = 1:n
91.         X = x(1) + delta*(i-1);
92.         graph(1,i+1) = X;
93.         graph(2,i+1) = f_normal(X,mu,S2);
94.     end
95.     graph(1,n+2) = grmax;
96.     graph(2,n+2) = 0;
97.
98.     xx = zeros(m+4);
99.     yy = zeros(m+4);
100.     ddelta = (x(n) - x(1)) / m;
101.     xx(1) = grmin;
102.     yy(1) = 0;
103.
104.     for i = 1:m

```

```

105.     xx(i+1) = intervals_table(1,i);
106.     yy(i+1) = intervals_table(2,i) / (n * ddelta);
107.
108.     end
108.     xx(m+2) = xx(m+1)+(xx(m+1)-xx(m));
109.     yy(m+2) = yy(m+1);
110.     xx(m+3) = xx(m+2);
111.     yy(m+3) = 0;
112.     xx(m+4)= grmax;
113.     yy(m+4)= 0;
114.
115.     stairs(xx, yy, 'b'), grid;
116.
117.     plot(graph(1,:), graph(2,:), 'r'),grid;
118. end
119. % плотность нормального распределения
120. function y = f_normal(x,mx,dx)
121.     y = exp(-((x-mx)^2)/2/dx)/sqrt(2*pi*dx);
122. end
123. % построение графиков эмпирической функции и функции
    распределения
124. function empir_and_F(x,mu,S2,grmin,grmax)
125.     n = length(x);
126.     delta = (grmax-grmin)/(n-1);
127.
128.     graph = zeros(2,n);
129.     for i = 1:n
130.         X = grmin + delta*(i-1);
131.         graph(1,i) = X;
132.         graph(2,i) = F_normal(X,mu,S2);
133.     end
134.
135.     F_empir = zeros(n+2);
136.     for i = 1:n
137.         F_empir(i+1) = empiric_F(x(i), x, n);
138.     end

```

```

139.
140.    x =[grmin x grmax];
141.    F_empir(n+2) = 1;
142.    stairs(x, F_empir, 'b'),grid;
143.
144.    plot(graph(1,:), graph(2,:), 'r'),grid;
145. end
146. % функция распределения нормальной случайной величины
147. function y = F_normal(x,mx,dx)
148.    y = 1/2 * (1 + erf((x - mx) / sqrt(2*dx)));
149. end
150. % эмпирическая функция распределения
151. function Fi = empiric_F(X, x, n)
152.    count = 0;
153.    for i = 1:n
154.        if (x(i) <= X)
155.            count = count + 1;
156.        else
157.            break;
158.        end
159.    end
160.    Fi = count/n;
161. end

```

Результаты расчетов

$M_{\max} = 2.92$

$M_{\min} = -2.77$

$R = 5.69$

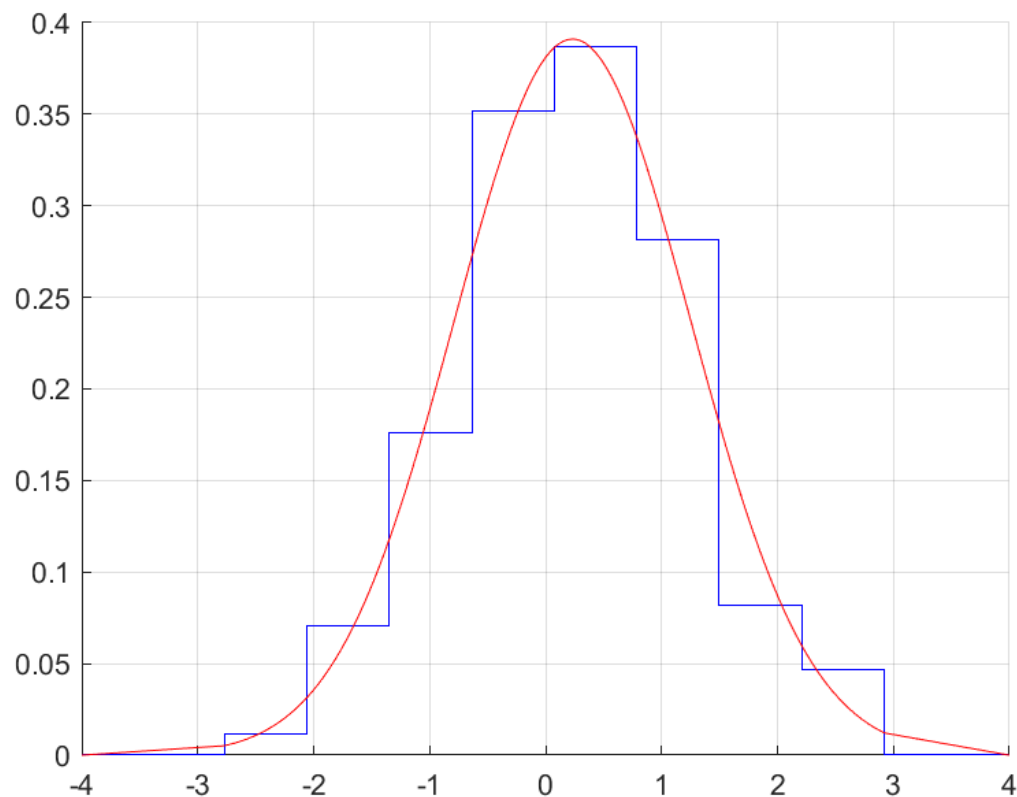
$\hat{\mu} = 0.2323$

$S^2 = 1.0319$

Группировка значений выборки в $m = 8$ интервалов:

| J_i | [-2.77;-2.06) | [-2.06;-1.35) | [-1.35;-0.64) | [-0.64;0.07) | [0.07;0.79) | [0.79;1.50) | [1.50;2.21) | [2.21; 2.92] |
|-------|---------------|---------------|---------------|--------------|-------------|-------------|-------------|---------------|
| n_i | 1 | 6 | 15 | 30 | 33 | 24 | 7 | 4 |

Построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2 :



Построение на другой координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2 :

