

Московский государственный технический университет
имени Н. Э. Баумана



Факультет: Информатика и системы управления

Кафедра: Программное обеспечение ЭВМ и информационные технологии

Дисциплина: Математическая статистика

Лабораторная работа №1
Гистограмма и эмпирическая функция распределения

Выполнил:

Группа:

Вариант: 15

Москва, 2015 г.

Содержание

1	Постановка задачи	3
2	Отчёт	4
2.1	Формулы для вычисления величин	4
2.1.1	Минимальное значение выборки	4
2.1.2	Максимальное значение выборки	4
2.1.3	Размах выборки	4
2.1.4	Оценка математического ожидания	4
2.1.5	Выборочная дисперсия	4
2.1.6	Несмещённая оценка дисперсии	4
2.2	Эмпирическая плотность и гистограмма	4
2.3	Эмпирическая функция распределения	5
3	Текст программы	6
4	Результаты расчётов	9
5	Графики	9
5.1	Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2	9
5.2	График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2	10

1 Постановка задачи

Цель работы: построение гистограммы и эмпирической функции распределения.

Содержание работы:

1. Для выборки объёма n из генеральной совокупности X реализовать в виде программы на ЭВМ
 - вычисление максимального значения M_{\max} и минимального значения M_{\min} ;
 - размаха R выборки;
 - вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX ;
 - группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
 - 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 .
2. Провести вычисления и построить графики для выборки из индивидуального варианта.

2 Отчёт

2.1 Формулы для вычисления величин

2.1.1 Минимальное значение выборки

$$M_{\min} = \min\{x_1, \dots, x_n\}, \quad \text{где} \quad (1)$$

- (x_1, \dots, x_n) — реализация случайной выборки.

2.1.2 Максимальное значение выборки

$$M_{\max} = \max\{x_1, \dots, x_n\}, \quad \text{где} \quad (2)$$

- (x_1, \dots, x_n) — реализация случайной выборки.

2.1.3 Размах выборки

$$R = M_{\max} - M_{\min}, \quad \text{где} \quad (3)$$

- M_{\max} — максимальное значение выборки;
- M_{\min} — минимальное значение выборки.

2.1.4 Оценка математического ожидания

$$\hat{\mu}(\vec{X}_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (4)$$

2.1.5 Выборочная дисперсия

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2. \quad (5)$$

2.1.6 Несмещённая оценка дисперсии

$$S^2(\vec{X}_n) = \frac{n}{n-1} \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (6)$$

2.2 Эмпирическая плотность и гистограмма

Определение 2.1. Эмпирической плотностью распределения случайной выборки \vec{X}_n называют функцию

$$f_n(x) = \begin{cases} \frac{n_i}{n\Delta}, & x \in J_i, \quad i = \overline{1, m}; \\ 0, & \text{иначе.} \end{cases}, \quad \text{где} \quad (7)$$

- $J_i, i = \overline{1, m}$, — полуинтервал из $J = [x_{(1)}, x_{(n)}]$, где

$$x_{(1)} = \min\{x_1, \dots, x_n\}, \quad x_{(n)} = \max\{x_1, \dots, x_n\}; \quad (8)$$

при этом все полуинтервалы, кроме последнего, не содержат правую границу т. е.

$$J_i = [x_{(1)} + (i-1)\Delta, x_{(i)} + i\Delta), \quad i = \overline{1, m-1}; \quad (9)$$

$$J_m = [x_{(1)} + (m-1)\Delta, x_{(1)} + m\Delta); \quad (10)$$

- m — количество полуинтервалов интервала $J = [x_{(1)}, x_{(n)}]$;
- Δ — длина полуинтервала J_i , $i = \overline{1, m}$ равная

$$\Delta = \frac{x_{(n)} - x_{(1)}}{m} = \frac{|J|}{m}; \quad (11)$$

- n_i — количество элементов выборки в полуинтервале J_i , $i = \overline{1, m}$;
- n — количество элементов в выборке.

Определение 2.2. График функции $f_n(x)$ называют гистограммой.

2.3 Эмпирическая функция распределения

- $\vec{X}_n = (X_1, \dots, X_n)$ — случайная выборка;
- $\vec{x}_n = (x_1, \dots, x_n)$ — реализация случайной выборки \vec{X}_n ;
- $n(x, \vec{x}_n)$ — количество элементов выборки \vec{x}_n , которые меньше x .

Определение 2.3. Эмпирической функцией распределения называют функцию

$$F_n: \mathfrak{R} \rightarrow \mathfrak{R}, \quad F_n(x) = \frac{n(x, \vec{x}_n)}{n} \quad (12)$$

Замечание. $F_n(x)$ обладает всеми свойствами функции распределения. При этом она кусочно-постоянна и принимает значения

$$0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{(n-1)}{n}, 1$$

Замечание. Если все элементы вектора \vec{x}_n различны, то

$$F_n(x) = \begin{cases} 0, & x \leq x_{(1)}; \\ \frac{i}{n}, & x_{(i)} < x \leq x_{(i+1)}, \quad i = \overline{1, n-1}; \\ 1, & x > x_{(n)}. \end{cases} \quad (13)$$

Замечание. Эмпирическая функция распределения позволяет интерпретировать выборку \vec{x}_n как реализацию дискретной случайной величины \tilde{X} ряд распределения которой

\tilde{X}	$x_{(1)}$	\dots	$x_{(n)}$
P	$1/n$	\dots	$1/n$

Это позволяет рассматривать числовые характеристики случайной величины \tilde{X} как приближённые значения числовых характеристик случайной величины X .

3 Текст программы

```
1 function lw1()
2     clear all;
3
4     X = readFromFile();
5     X = sort(X);
6
7     minX = X(1);
8     fprintf('Mmin = %s\n', num2str(minX));
9
10    maxX = X(end);
11    fprintf('Mmax = %s\n', num2str(maxX));
12
13    R = maxX - minX;
14    fprintf('R = %s\n', num2str(R));
15
16    mu = expectation(X);
17    fprintf('mu = %s\n', num2str(mu));
18
19    sigmaSqr = populationVariance(X);
20    fprintf('sigma^2 = %s\n', num2str(sigmaSqr));
21
22    sSqr = unbiasedSampleVariance(X);
23    fprintf('S^2: %s\n', num2str(sSqr));
24
25    m = numberOfSubintervals(length(X));
26    fprintf('m = %s\n ', num2str(m));
27
28    intervals(X, m);
29    hold on;
30    f(X, mu, sSqr);
31    figure;
32
33    F(X, mu, sSqr);
34    hold on;
35    empiricF(X);
36 end
37
38 function X = readFromFile()
39     X = csvread('data.txt');
40 end
41
42 function m = numberOfSubintervals(size)
43     m = floor(log2(size) + 2);
44 end
45
46 function mu = expectation(X)
47     n = length(X);
48     sum = 0;
```

```

49
50     for i = 1:n
51         sum = sum + X(i);
52     end
53
54     mu = sum / n;
55 end
56
57 function sigmaSqr = populationVariance(X)
58     n = length(X);
59     sum = 0;
60
61     for i = 1:n
62         sum = sum + (X(i))^2;
63     end
64
65     mu = expectation(X);
66
67     sigmaSqr = sum / n - mu^2;
68 end
69
70 function sSqr = unbiasedSampleVariance(X)
71     sigmaSqr = populationVariance(X);
72     n = length(X);
73
74     sSqr = n / (n - 1) * sigmaSqr;
75 end
76
77 function intervals(X, m)
78     count = zeros(1, m+1);
79     delta = (X(end) - X(1)) / m;
80
81     J = X(1):delta:X(end);
82     j = 1;
83     n = length(X);
84
85     for i = 1:n
86         if (j ~= m)
87             if ((not (X(i) >= J(j) && X(i) < J(j+1))))
88                 j = j + 1;
89                 fprintf( '[%.2f;%.2f]\t', J(j-1), J(j));
90             end
91         end
92         count(j) = count(j) + 1;
93     end;
94     fprintf( '[%.2f;%.2f]\n', J(m), J(m + 1));
95
96     Xbuf = count(1:m+1);
97     for i = 1:m+1
98         Xbuf(i) = count(i) / (n*delta);

```

```

99      end
100
101      stairs(J, Xbuf), grid;
102 end
103
104 function f(X, MX, DX)
105     Y = 1 / sqrt(2*pi*DX) * exp( -power((X - MX), 2) / (2*DX));
106     plot(X, Y);
107 end
108
109 function F(X, MX, DX)
110     Y = 1/2 * (1 + erf((X - MX) / sqrt(2*DX)));
111
112     plot(X, Y, 'r');
113 end
114
115 function empiricF(X)
116     n = length(X);
117     Y = zeros(1, n);
118
119     for i = 1:n
120         Y(i) = i / n;
121     end;
122
123     stairs(X, Y), grid;
124 end

```


4 Результаты расчётов

$$\begin{aligned}M_{\min} &= -6.48; \\M_{\max} &= -1.51; \\R &= 4.97; \\\hat{\mu}(\vec{X}_n) &= -3.6762; \\S^2(\vec{X}_n) &= 0.86641.\end{aligned}$$

Интервальная группировка значений выборки при $m = 8$:

$$\begin{aligned}[-6.48, -5.86), [-5.86, -5.24), [-5.24, -4.62), [-4.62, -4.00), [-4.00, -3.37), \\[-3.37, -2.75), [-2.75, -2.13), [-2.13, -1.51].\end{aligned}$$

5 Графики

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

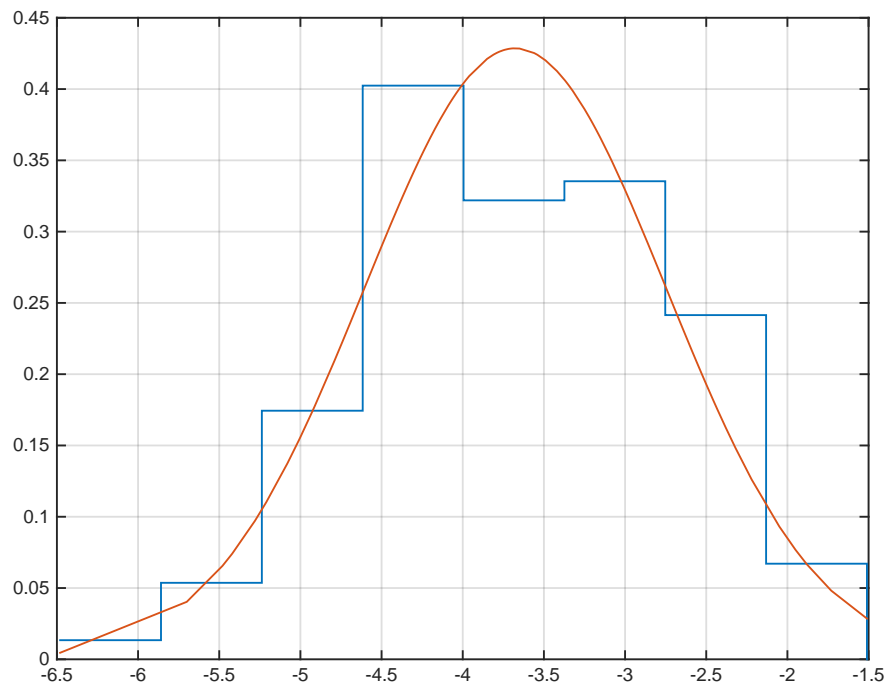


Рис. 1: Гистограмма и график функции плотности распределения нормальной случайной величины.

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

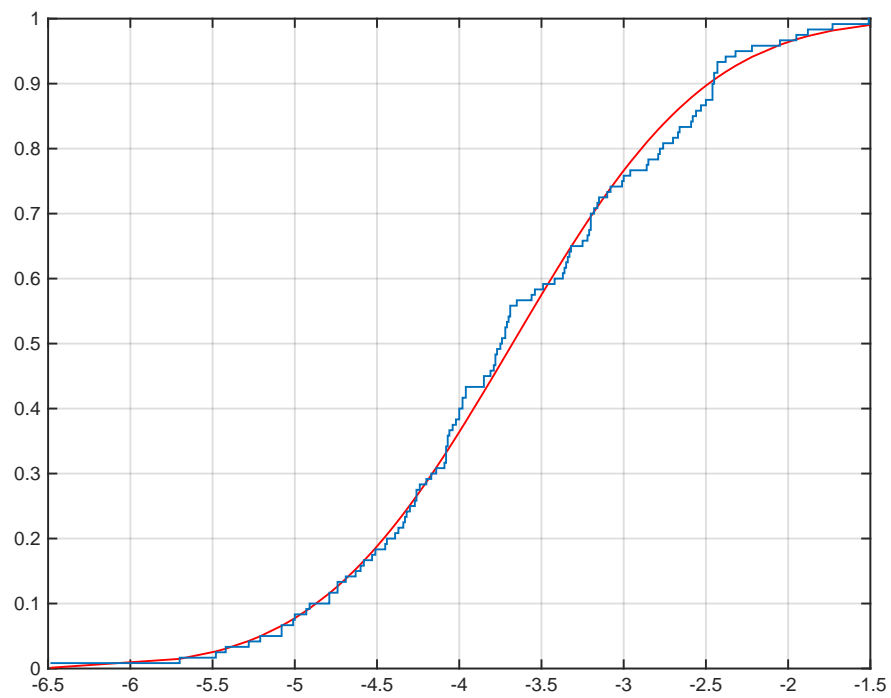


Рис. 2: График эмпирической функции распределения и функции распределения нормальной случайной величины.