



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

Дисциплина	Математическая статистика
Тема	Гистограмма и эмпирическая функция распределения
Студент	Набиев Ф.М.
Группа	ИУ7–63Б
Преподаватель	Власов П.А.

Москва, 2020 г.

1 Теоретическая часть

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

Тогда:

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° Выборочное среднее:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

5° Состоятельная оценка дисперсии:

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



Пусть \vec{x} — выборка из генеральной совокупности X ,

$$\vec{x} = (x_1, \dots, x_n)$$

Расположим значения x_1, \dots, x_n в порядке неубывания

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

где $x_{(i)}$ — i -й элемент полученной последовательности. Такая последовательность называется вариационным рядом.

Если объем n статистической выборки \vec{x} велик ($n \geq 50$), то можно сгруппировать выборку в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ делят на $p = [\log_2 n] + 1$ (где $[a]$ — целая часть числа a) равновеликих частей:

$$J_i = [a_{i-1}, a_i), i = \overline{1, p-1}$$

$$J_p = [a_{p-1}, a_p]$$

где

$$a_i = x_{(1)} + i\Delta, i = \overline{0, p}$$

$$\Delta = \frac{|J|}{p} = \frac{x_{(n)} - x_{(1)}}{p}$$

Интервальным статистическим рядом называют таблицу

J_1	\dots	J_i	\dots	J_p
n_1	\dots	n_i	\dots	n_p

где n_i — количество элементов \vec{x} , которые $\in J_i$.

Предположим, что для выборки \vec{x} построен интервальный статистический ряд

$$(J_i, n_i), i = \overline{1; p}$$

Эмпирической плотностью (отвечающей выборке \vec{x}) называют функцию

$$\hat{f}_n(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1; p} \\ 0, \text{ иначе} \end{cases}$$

Гистограммой называют график эмпирической плотности.



Для выборки \vec{x} обозначим $n(x, \vec{x})$ — число элементов вектора \vec{x} , которые имеют значения меньше x .

Эмпирической функцией распределения называют функцию

$$F_n : \mathbb{R} \rightarrow \mathbb{R}$$

определенную условием

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



2 Практическая часть

В листинге 2.1 приведён текст программы.

Листинг 2.1 – Текст программы

```
1 function lab1 ()
2     X = [
3         -13.40, -12.63, -13.65, -14.23, -13.39, -12.36, ...
4         -13.52, -13.44, -13.87, -11.82, -12.01, -11.40, ...
5         -13.02, -12.61, -13.06, -13.75, -13.55, -14.01, ...
6         -11.75, -12.95, -12.59, -13.60, -12.76, -11.05, ...
7         -13.15, -13.61, -11.73, -13.00, -12.66, -12.67, ...
8         -12.60, -12.47, -13.52, -12.61, -11.93, -13.11, ...
9         -13.22, -11.87, -13.44, -12.70, -11.78, -12.30, ...
10        -12.89, -13.29, -12.48, -10.44, -12.55, -12.64, ...
11        -12.03, -14.60, -14.56, -13.30, -11.32, -12.24, ...
12        -11.17, -12.50, -13.25, -12.55, -12.85, -12.67, ...
13        -12.41, -12.58, -12.10, -13.54, -12.69, -12.87, ...
14        -12.71, -12.77, -13.30, -12.74, -12.73, -12.64, ...
15        -12.18, -11.20, -12.40, -13.78, -13.71, -10.74, ...
16        -11.89, -13.20, -11.31, -14.26, -10.38, -12.88, ...
17        -11.39, -11.35, -12.55, -12.84, -10.25, -12.40, ...
18        -14.01, -11.47, -13.14, -12.69, -11.92, -12.86, ...
19        -13.06, -12.57, -13.63, -12.34, -12.84, -14.03, ...
20        -13.34, -11.64, -13.58, -10.44, -11.37, -11.01, ...
21        -13.80, -13.27, -12.32, -10.69, -12.92, -13.29, ...
22        -12.58, -13.98, -11.46, -11.82, -12.33, -11.47, ...
23    ];
24    X = sort(X);
25
26    % Максимальное значение выборки
27    M_max = max(X);
28    % Минимальное значение выборки
29    M_min = min(X);
30
31    % Разброс выборки
32    R = M_max - M_min;
33
34    % Оценка мат. ожидания
35    M = mean(X);
36    % Оценка дисперсии
37    D = var(X);
38
39    fprintf( 'M_max      = %9.5f\n', M_max);
40    fprintf( 'M_min      = %9.5f\n', M_min);
41    fprintf( 'R          = %9.5f\n', R);
42    fprintf( 'mean       = %9.5f\n', M);
43    fprintf( 'variance   = %9.5f\n', D);
44    fprintf( '\n');
```

```

45
46 m = floor(log2(length(X))) + 2;
47 % Группировка значений в m интервалов
48 [N, edges] = histcounts(X, m, 'BinLimits', [M_min, M_max]);
49
50 % Вывод всех интервалов, кроме последнего,
51 % потому что последний – это отрезок,
52 % а все остальные – полуинтервалы
53 fprintf('%d intervals:\n', m);
54 for i = 1 : (length(N) - 1)
55     fprintf('%3d values in [%f,%f]\n', ...
56           N(i), edges(i), edges(i + 1));
57 end
58 % Вывод последнего отрезка
59 fprintf('%3d values in [%f,%f]\n', ...
60       N(end), edges(end - 1), edges(end));
61
62 f = normpdf(X, M, sqrt(D));
63 F = normcdf(X, M, sqrt(D));
64
65 subplot(2, 1, 1);
66 % Построение гистограммы
67 histogram(X, m, 'Normalization', 'pdf', ...
68          'BinLimits', [M_min, M_max]);
69 hold on;
70 % Построение графика функции плотности распределения
71 plot(X, f, 'LineWidth', 2);
72 hold off;
73
74 subplot(2, 1, 2);
75 % Построение графика эмпирической функции распределения
76 [YY, XX] = ecdf(X);
77 stairs(XX, YY, 'LineWidth', 2);
78 hold on;
79 % Построение графика функции распределения
80 plot(X, F, 'LineWidth', 2);
81 hold off;
82 end

```

В листинге 2.2 приведён результат выполнения описанной программы.

Листинг 2.2 – Результат программы

```

M_max    = -10.25000
M_min    = -14.60000
R        =  4.35000
mean     = -12.61483

```

```
variance = 0.86533
```

```
8 intervals:
```

```
4 values in [-14.600000, -14.056250)
18 values in [-14.056250, -13.512500)
20 values in [-13.512500, -12.968750)
36 values in [-12.968750, -12.425000)
16 values in [-12.425000, -11.881250)
14 values in [-11.881250, -11.337500)
6 values in [-11.337500, -10.793750)
6 values in [-10.793750, -10.250000]
```

Рис. 2.1 – Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

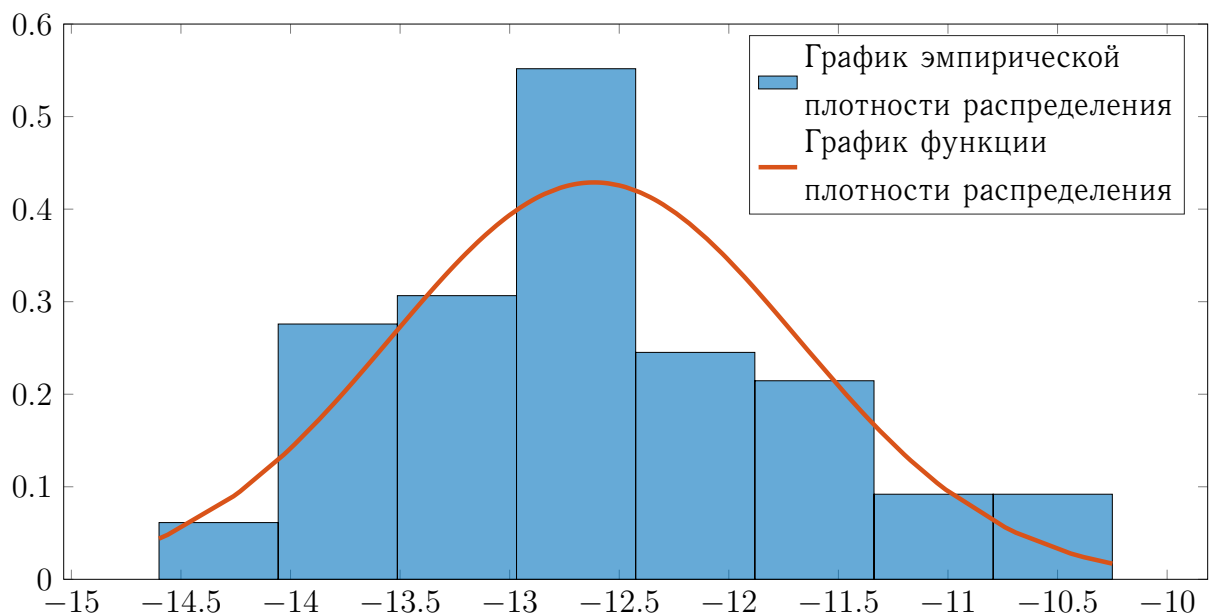


Рис. 2.2 – Графики эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

