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	Куприй А. А.
Группа	ИУ7–63Б
Преподаватель	Власов П.А.

Москва, 2020 г.

1 Теоретическая часть

1.1 Формулы для вычисления величин

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

Тогда:

1° Максимальное значение выборки:

$M_{\max} = \max\{x_1, \ldots, x_n\}$

2° Минимальное значение выборки:

$M_{\min} = \min\{x_1, \ldots, x_n\}$

3° Размах выборки:

$R = M_{\max} - M_{\min}$

4° Выборочное среднее:

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

5° Состоятельная оценка дисперсии:

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

где $\overline{x} = \hat{\mu}$

1.2 Эмпирическая плотность и гистограмма

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

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

Расположим значения $x_1, \ldots, x_n$ в порядке неубывания

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

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 $J = \left[x_{(1)}, x_{(n)}\right]$ делят на $p = [\log_2 n] + 1$ (где $[a]$ — целая часть числа a) равновеликих частей:

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

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

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	$\ldots$	J_i	$\ldots$	J_p
n_1	$\ldots$	n_i	$\ldots$	n_p

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

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

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

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.

1.3 Эмпирическая функция распределения

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

Эмпирической функцией распределения называют функцию

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

определенную условием

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

2 Практическая часть

2.1 Текст программы

В листинге 2.1 приведён текст программы.

Листинг 2.1 – Текст программы

```
function lab1()
    X=[
        -2.79, -3.01, -4.07, -2.85, -2.43, -3.20, ...
        -3.72, -4.27, -5.48, -2.38, -4.69, -4.34, ...
        -5.08, -5.01, -4.08, -4.20, -4.74, -1.88, ...
        -3.25, -2.78, -3.56, -3.54, -3.79, -3.18, ...
        -5.08, -4.30, -2.86, -2.45, -3.08, -3.22, ...
        -2.76, -3.20, -3.33, -4.91, -4.06, -3.81, ...
        -3.96, -3.65, -3.77, -4.60, -5.21, -2.67, ...
        -1.95, -2.43, -1.73, -2.50, -3.96, -3.75, ...
        -2.70, -4.26, -3.42, -4.07, -4.74, -3.00, ...
        -4.37, -5.42, -5.00, -4.08, -2.46, -4.33, ...
        -4.08, -3.72, -4.09, -2.96, -3.71, -1.51, ...
        -3.70, -6.48, -4.26, -4.39, -3.16, -4.63, ...
        -2.66, -2.22, -4.79, -2.46, -3.69, -3.35, ...
        -2.32, -4.17, -3.85, -4.93, -2.05, -3.15, ...
        -3.49, -5.70, -2.53, -3.85, -4.32, -3.37, ...
        -3.98, -3.74, -5.28, -2.56, -3.21, -3.10, ...
        -3.78, -3.36, -3.32, -2.59, -2.45, -3.34, ...
        -3.20, -4.14, -4.00, -4.79, -4.02, -4.58, ...
        -4.45, -3.69, -4.53, -3.98, -4.51, -4.44, ...
        -3.78, -4.24, -4.00, -2.46, -2.58, -4.04, ...
    ];
    X = sort(X);

    % Максимальное значение выборки
    Mmax = max(X);
    % Минимальное значение выборки
    Mmin = min(X);

    % Разброс выборки
    R = Mmax - Mmin;

    % Оценка мат. ожидания
    mu = mean(X);
    % Оценка дисперсии
    D = var(X);

    fprintf('Mmax = %f\n', Mmax);
    fprintf('Mmin = %f\n', Mmin);
    fprintf('R = %f\n', R);
```

```
    fprintf('Mean = %f\n', mu);
    fprintf('Varience = %f\n\n', D);

    % Нахождение количества интервалов
    m = floor(log2(length(X))) + 2;
    % Группировка значений в m интервалов
    [N, edges] = histcounts(X, m, 'BinLimits', [Mmin, Mmax]);
    len_c = length(N);

    % Вывод всех интервалов и количества элементов
    fprintf('%d intervals:\n', m);
    for i = 1 : (len_c - 1)
        fprintf('[%f,%f) - %d\n', edges(i), edges(i + 1), N(i));
    end
    fprintf('[%f,%f] - %d\n', edges(len_c), edges(len_c + 1), N(len_c));

    % График функции плотности распределения вероятностей нормальной
    % случайной величины
    f = normpdf(X, mu, sqrt(D));
    % Построение гистограммы
    histogram(X, m, 'Normalization', 'pdf', 'BinLimits', [Mmin, Mmax]);
    % Построение на одной координатной плоскости
    hold on;
    % Построение графика функции плотности распределения
    plot(X, f, 'LineWidth', 2);
    legend({'Гистограмма', 'Функция плотности распределения'}, ...
        'Location','northwest');
    % Новая координатная плоскость
    figure;
    % График функции распределения нормальной случайной величины
    F = normcdf(X, mu, sqrt(D));
    % Построение графика эмпирической функции распределения
    [YY, XX] = ecdf(X);
    stairs(XX, YY, 'LineWidth', 2);
    hold on;
    % Построение графика функции распределения
    plot(X, F, 'LineWidth', 2);
    legend({'Эмперическая функция распределения', ...
        'Функция распределения'}, ...
        'Location','northwest');
end
```

2.2 Результат

В листинге 2.2 приведён результат выполнения описанной программы.

Листинг 2.2 – Результат программы

```
Mmax = -1.510000
Mmin = -6.480000
R = 4.970000
Mean = -3.676167
Varience = 0.866410

8 intervals:
[-6.480000,-5.858750) - 1
[-5.858750,-5.237500) - 4
[-5.237500,-4.616250) - 13
[-4.616250,-3.995000) - 30
[-3.995000,-3.373750) - 24
[-3.373750,-2.752500) - 25
[-2.752500,-2.131250) - 18
[-2.131250,-1.510000] - 5
```

На рисунке 2.1 представлена гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

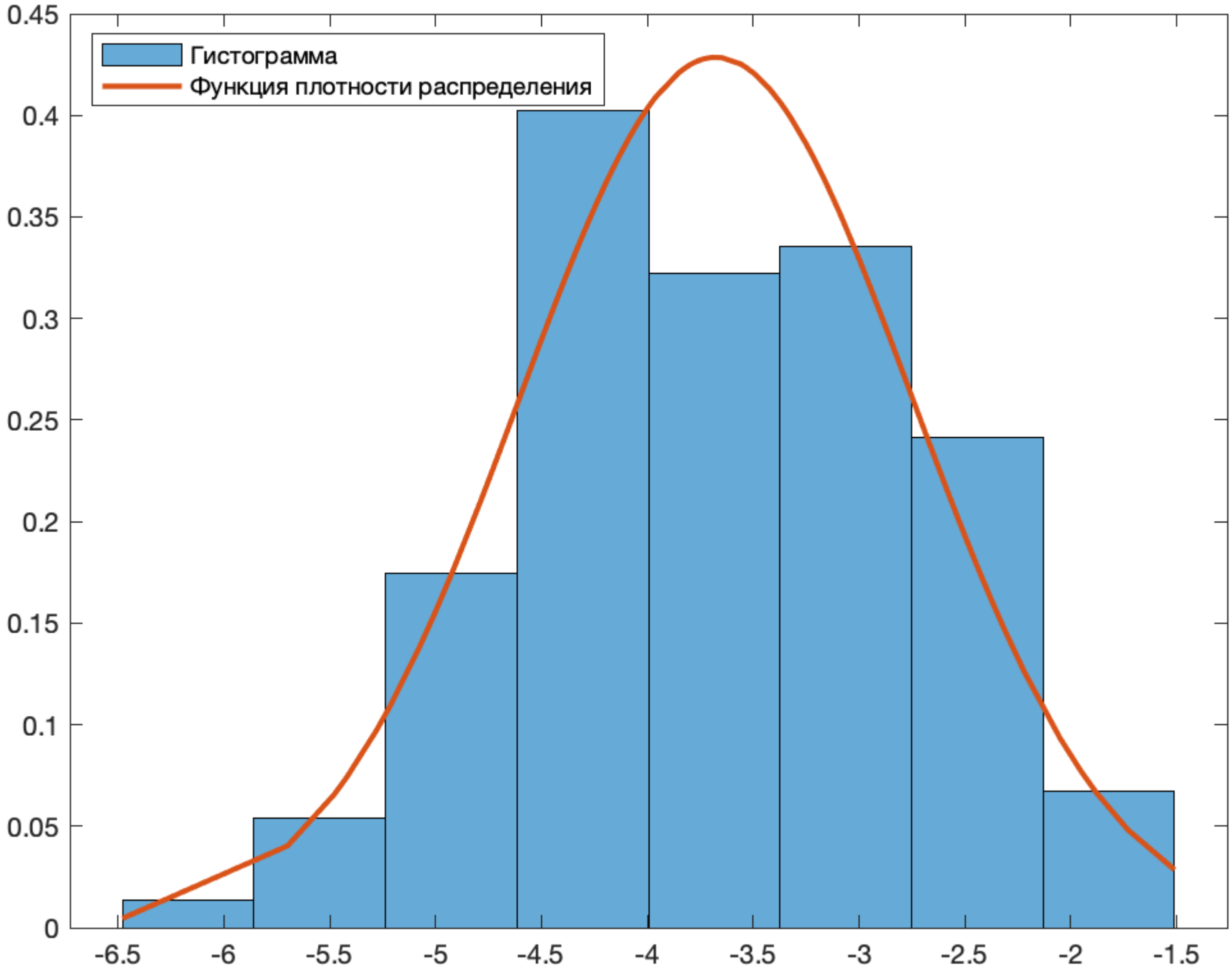
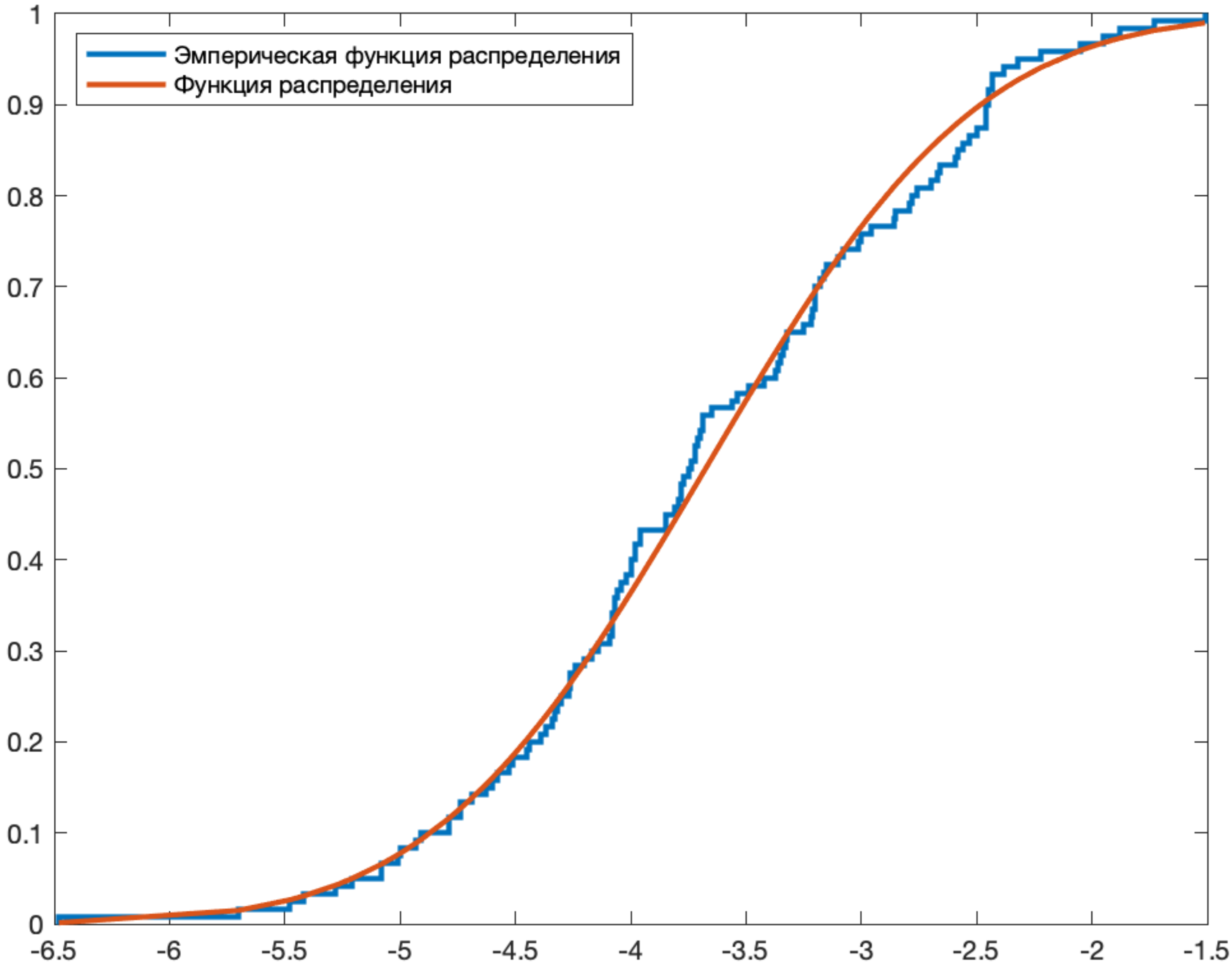


Рис. 2.1

На рисунке 2.2 представлены графики эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2


Эмперическая функция распределения
Функция распределения
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6.5
-6
-5.5
-5
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
Рис. 2.2