Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»
КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

Лабораторная работа № 1

Вариант 22

Дисциплина	Математическая статистика.
Тема	
Студент	Тимонин А. С.
Группа	ИУ7-62Б
Оценка (баллы)	
Преподаватель	Власов П.А.

Москва, 2020 г.

Теоретическая часть

Формулы для вычисления

$\vec{x} = (x_1, \ldots, x_n)$ **- выборка из генеральной совокупности** $\vec{X}$

Формула для вычисления максимального значения $M_{\max}$**:**

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

Формула для вычисления минимального значения $M_{\min}$**:**

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

Размах выборки R вычисляется по формуле:

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

Вычисление оценки математического ожидания MX:

$$\hat{\mu}(\hat{X}_n) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Вычисление оценки дисперсии DX:

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

Определение эмпирической плотности и гистограммы

Интервальный статистический ряд

Пусть $\vec{x}$ - выборка из генеральной совокупности X. Если объем n этой выборки велик ($n \geq 50$) отрезок $[x_{(1)}; x_{(n)}]$ разбивают на m равновеликих интервалов и для каждого интервала указывают количество принадлежащих элементов выборки. При этом, как правило, m определяют как: $m = [\log_2 n] + 1$ - количество интервалов, $\Delta = \frac{x_{(n)} - x_{(1)}}{m}$ - ширина одного интервала.

Интервалы:

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

$$J_m = [x_{(1)} + (m-i)\Delta; x_{(1)} + m\Delta]$$

где $x_{(n)} = x_{(1)} + m\Delta$

Опр Интервальным статистическим рядом называют таблицу

J_1	$\ldots$	J_i	$\ldots$	J_p
n_1	$\ldots$	n_i	$\ldots$	n_p

Здесь n_i- количество элементов выборки $\vec{x}$, которые $\in J_i$

Замечание

1. Очевидно, что $\sum_{i=1}^{p} n_i = n$

2. Для выборки p - числа интервалов можно пользоваться формулой $p = [\log_n n] + 1$

где [a] - целая часть числа a

Опр **Эмпирической плотностью** (отвечающей выборке $\vec{x}$) называют функцию:

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

Опр **Гистограммой** называют график эмпирической плотности

Определение эмперической функции распределения

Опр **Эмпирической функцией распределения** называют функцию

$$\mathcal{F}_n : \mathbb{R} \to \mathbb{R}$$

определенную условием

$$\hat{F}_n(x) = \frac{h(x, \vec{x})}{n}$$

где $h(x, \vec{x})$ - количество элементов выборки $\vec{x}$, которые имеют значение, меньшее x.

Практическая часть

Листинги программы

Листинг 1: Реализация программы

```
function lab1()
    X = [7.76,5.96,4.58,6.13,5.05,6.40,7.46,5.55,5.01,3.79, ...
    7.65,8.87,5.94,7.25,6.76,6.92,6.68,4.89,7.47,6.53, ...
    6.76,6.96,6.58,7.92,8.47,6.27,8.05,5.24,5.60,6.69, ...
    7.55,6.02,7.34,6.81,7.22,6.39,6.40,8.28,5.39,5.68, ...
    6.71,7.89,5.69,5.18,7.84,7.18,7.54,6.04,4.58,6.82, ...
    4.45,6.75,5.28,7.42,6.88,7.10,5.24,9.12,7.37,5.50, ...
    5.52,6.34,5.31,7.71,6.88,6.45,7.51,6.21,7.44,6.15, ...
    6.25,5.59,6.68,6.52,4.03,5.35,6.53,3.68,5.91,6.68, ...
    6.18,7.80,7.17,7.31,4.48,5.69,7.11,6.87,6.14,4.73, ...
    6.60,5.61,7.32,6.75,6.28,6.41,7.31,6.68,7.26,7.94, ...
    7.67,4.72,6.01,5.79,7.38,5.98,5.36,6.43,7.25,5.54, ...
    6.66,6.47,6.84,6.13,6.21,5.52,6.33,7.55,6.24,7.84];

    X = sort(X);

    % a)
    Xmin = X(1);
    fprintf('Xmin = %.3f\n', Xmin);
    Xmax = X(end);
    fprintf('Xmax = %.3f\n', Xmax);

    % b)
    R = Xmax - Xmin;
    fprintf('R = %.3f\n', R);

    % v)
    MX = expectation(X);
    fprintf('MX= %.3f\n', MX);
    DX = variance(X);
    fprintf('DX= %.3f\n', DX);

    % g)
    m = group(X);
    fprintf('m=%d\n', m);
    delta=R/m;
    fprintf('delta = %.3f\n', delta);
    J = Xmin : delta : Xmax;
    n = length(J);
    for i = 1:(n - 2)
      fprintf('[%.3f; %.3f)\n', J(i), J(i + 1));
    end
    fprintf('[%.3f; %.3f]\n', J(n - 1), J(n));

    sigma = sqrt(DX);
    Xn = (MX - R) : delta / 50 :(MX + R);

  % d)
    f(X, Xn, MX, sigma);
```

```

  % e)
    F(X, Xn, MX, sigma);

end

function MX = expectation(X)
    MX = mean(X);
end

function DX = variance(X)
    DX = var(X);
end

function m = group(X)
    m = floor(log2(length(X))) + 2;
end

function F(X, Xn, MX, sigma)
    figure;
    Ycdf = 1/2 * (1 + erf((Xn - MX) / sqrt(2) * sigma));
    ecdf (X);
    hold on;
    plot(Xn, Ycdf, 'r');
    hold off ;
end

function f(X, Xn, MX, sigma)
    Ypdf = normpdf(Xn, MX, sigma);
    histogram (X, 'normalization' , 'pdf' ) ;
    hold on;
    plt = plot(Xn, Ypdf, 'r');
    plt.LineWidth = 1;
end
```

Результат выполнения программы

$$M_{\max} = 9.12$$

$$M_{\min} = 3.68$$

$$R = 5.44$$

$$\mu = 6.4596$$

$$S^2 = 1.1013$$

$[3.68; 4.36)$	3
$[4.36; 5.04)$	8
$[5.04; 5.72)$	20
$[5.72; 6.40)$	22
$[6.40; 7.08)$	30
$[7.08; 7.76)$	25
$[7.76; 8.44)$	9
$[8.44; 9.12]$	3

Таблица 1: Результаты расчетов для выборки

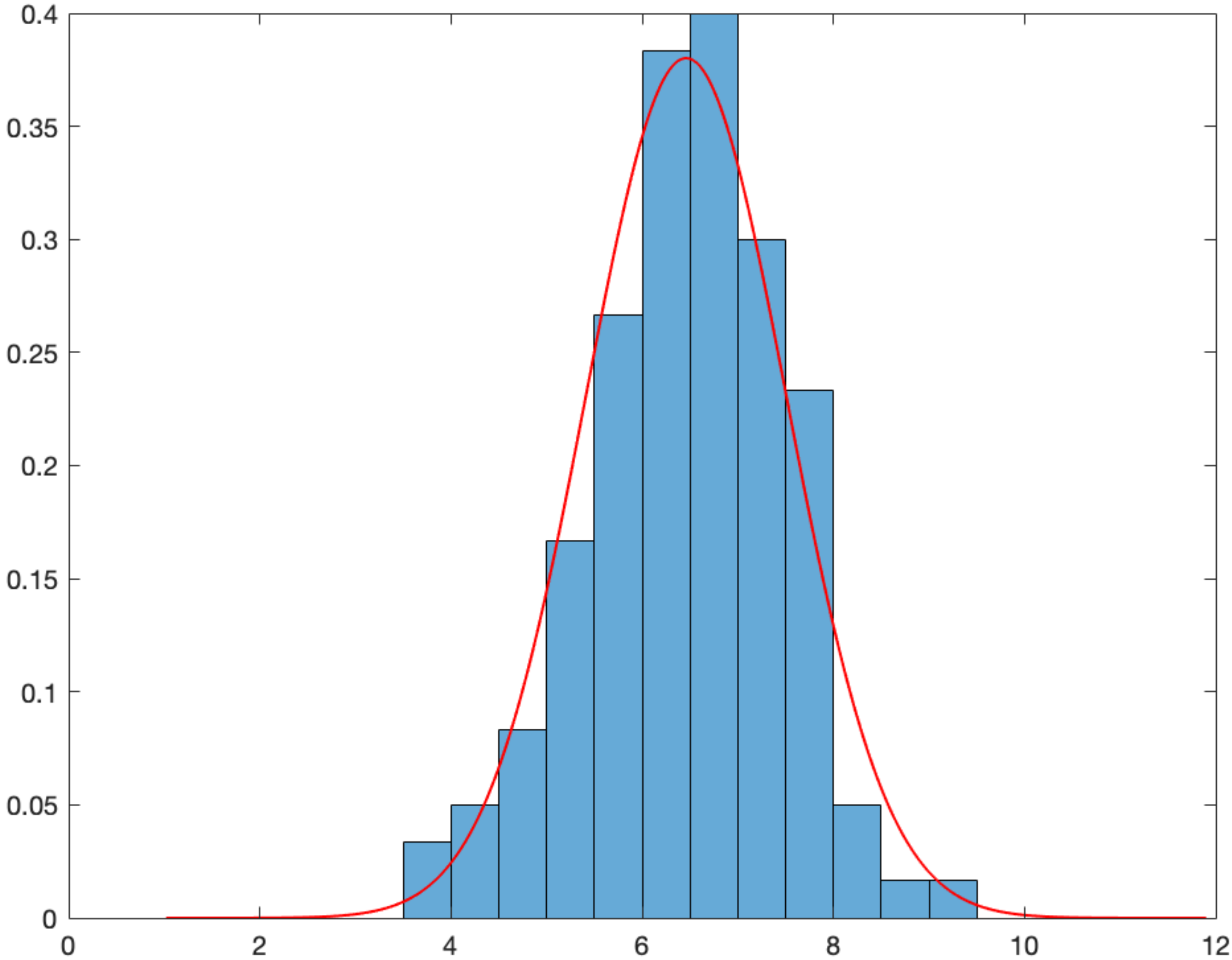

Рис. 1: Гистограмма

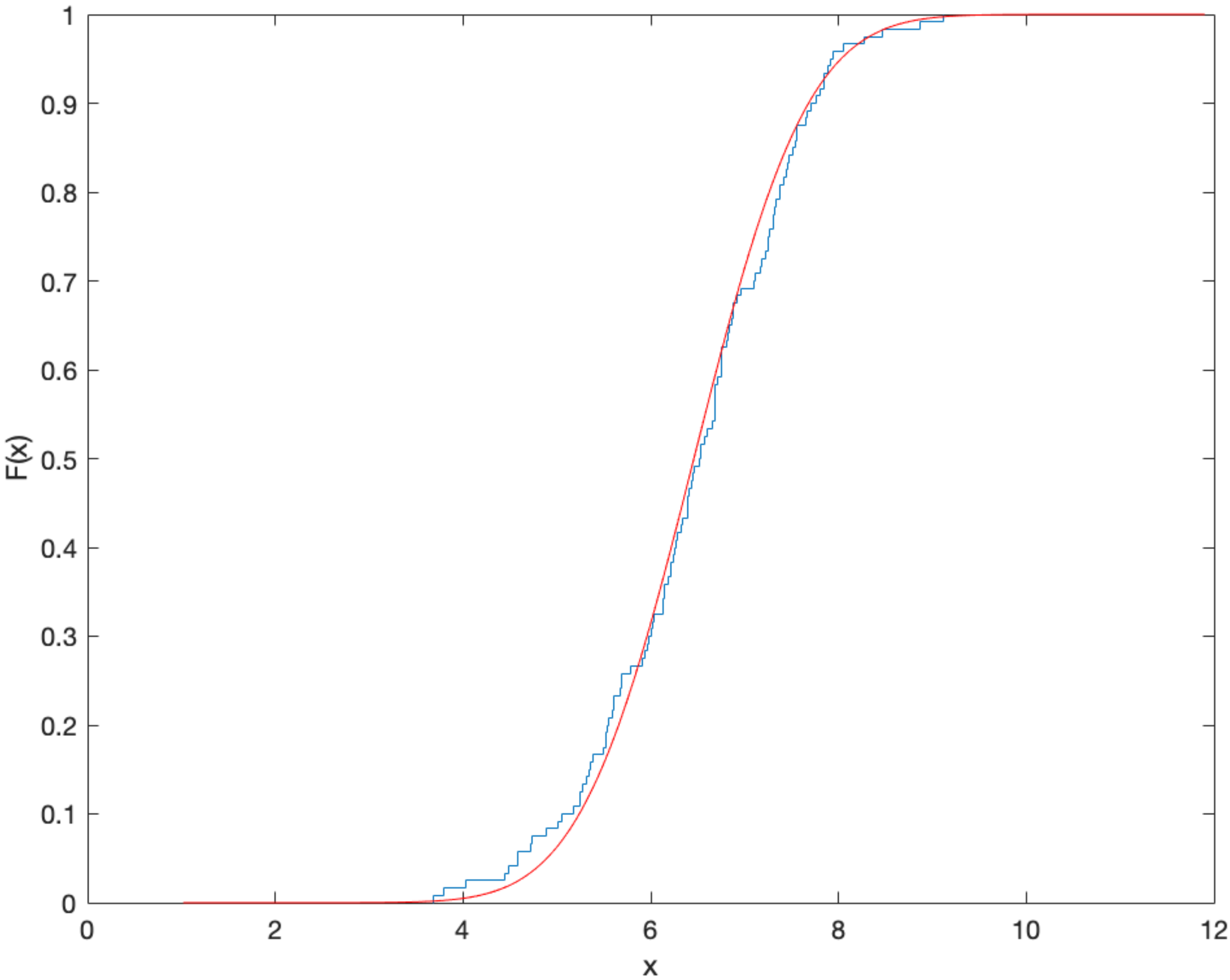

Рис. 2: Эмпирическая функция распределения