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
По курсу «Математическая статистика»
Тема: Гистограмма и эмпирическая функция распределения

Студент: Кондрашова О.П.
Группа: ИУ7-65Б
Вариант: 7

Преподаватели: Власов П.А.
Волков И.К.

Москва, 2020г.

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
   (a) вычисление максимального значения $M_{\max}$ и минимального значения $M_{\min}$;
   (b) вычисление размаха R выборки;
   (c) вычисление оценок $\hat{\mu}$ и S^2 математического ожидания $\mathsf{M}X$ и дисперсии $\mathsf{D}X$;
   (d) группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
   (e) построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
   (f) 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 Отчет

2.1 Формулы для вычисления величин

Количество интервалов

$$m = [\log_2 n] + 2 \tag{1}$$

Минимальное значение выборки

$$M_{\min} = \min\{x_1, \ldots, x_n\}, \quad \text{где} \tag{2}$$

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

Максимальное значение выборки

$$M_{\max} = \max\{x_1, \ldots, x_n\}, \quad \text{где} \tag{3}$$

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

Размах выборки

$$R = M_{\max} - M_{\min}, \quad \text{где} \tag{4}$$

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

Оценка математического ожидания

$$\hat{\mu}(\vec{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{5}$$

Несмещенная оценка дисперсии

$$S^2(\vec{X}) = \frac{n}{n-1}\hat{\sigma}^2(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\,. \tag{6}$$

2.2 Эмпирическая плотность и гистограмма

Эмпирической плотностью распределения выборки $\vec{x}$ называют функцию

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

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

$$x_{(1)} = \min\{x_1, \ldots, x_n\}, \qquad x_{(n)} = \max\{x_1, \ldots, x_n\}; \tag{8}$$

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

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

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

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

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

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

График функции $f_n(x)$ называют *гистограммой*. Гистограмма представляет собой кусочно-постоянную функцию на промежутке J.

2.3 Эмпирическая функция распределения

Эмпирической функцией распределения, отвечающей выборке $\vec{x}$ называют функцию

$$F_n\colon \mathbb{R} \to \mathbb{R}, \qquad F_n(x) = \frac{n(x, \vec{x})}{n}, \tag{12}$$

где $n(x, \vec{x})$ — количество элементов выборки $\vec{x}$, которые меньше x.

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

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

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

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

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

$\widetilde{X}$	$x_{(1)}$	$\ldots$	$x_{(n)}$
P	$1/n$	$\ldots$	$1/n$

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

3 Листинг программы

```
function lab1()
    % Выборка объема n из генеральной совокупности X
    X = [-1.12,-1.06,0.46,-0.39,0.09,-1.44,-1.64,0.86,-0.24,-1.71, ...
        -0.84,-1.19,-0.84,-0.55,-1.11,-1.84,-0.60,-0.92,-0.69,0.23, ...
        0.51,-2.41,-0.53,-1.41,-0.23,-0.89,-0.13,-1.50,0.02,0.27, ...
        -0.75,-0.06,-0.48,0.14,0.20,-2.22,-1.42,-0.54,0.83,-1.77, ...
        -0.10,-0.07,-0.94,-0.13,-1.76,-0.77,-1.26,-0.29,-1.11, ...
        -0.56,1.19,-0.92,-2.02,-1.94,-0.36,-2.09,-2.51,-1.82,0.39, ...
        -2.08,-0.60,-1.38,-1.12,-0.34,0.77,-1.34,0.24,-0.30,-1.67, ...
        -1.50,-0.77,-0.10,-0.39,-0.35,-2.23,-0.84,-0.85,-0.44,-0.20, ...
        -1.76,-0.91,-1.30,-2.03,-2.50,1.08,0.19,0.03,1.17,-0.05, ...
        -2.88,-1.13,-0.05,-1.37,-0.22,0.88,-1.04,-0.52,-1.64,-0.43, ...
        -0.09,-2.44,-0.78,-2.48,-1.16,-0.44,-0.34,-0.60,-0.11,-0.41, ...
        -0.04,-1.09,-1.81,-0.74,-1.07,-1.07,-0.68,-0.36,-0.65,-1.72, ...
        -0.49];

    % Минимальное значение
    Mmin = min(X);
    fprintf('Mmin = %f\n', Mmin);

    % Максимальное значение
    Mmax = max(X);
    fprintf('Mmax = %f\n', Mmax);

    % Размах выборки
    R = Mmax - Mmin;
    fprintf('R = %f\n', R);

    % Выборочное среднее
    mu = mean(X);
    fprintf('mu = %f\n', mu);

    % Несмещенная оценка дисперсии
    s2 = var(X);
    fprintf('S2 = %f\n', s2);

    % Нахождение количества интервалов
    m = floor(log2(length(X))) + 2;

    % Разбиваем выборку на m интервалов от min до max
    [count, edges] = histcounts(X, m, 'BinLimits', [min(X), max(X)]);
    countLen = length(count);

    % Интервалы и количество элементов в них
    fprintf('\nИнтервальная группировка значений выборки при m = %d \n', m);
    for i = 1 : (countLen - 1)
        fprintf('[%f : %f) - %d\n', edges(i), edges(i + 1), count(i));
```

```
    end
    fprintf('[%f : %f] - %d\n', edges(countLen), edges(countLen + 1), count(countLen));

    % Гистограмма
    plotHistogram(X, count, edges, m);
    hold on;
    % График функции плотности распределения вероятностей нормальной случайной величины
    f(X, mu, s2, m, R);
    figure;
    % График эмпирической функции распределения
    plotEmpiricalF(X);
    hold on;
    % График функции распределения нормальной случайной величины
    F(sort(X), mu, s2, m, R);

function plotHistogram(X, count, edges, m)
    h = histogram();
    h.BinEdges = edges;
    h.BinCounts = count / length(X) / ((max(X) - min(X)) / m);
end

function f(X, MX, DX, m, R)
        delta = R/m;
        sigma = sqrt(DX);
        Xn = min(X):delta/20:max(X);
        Y = normpdf(Xn, MX, sigma);
        plot(Xn, Y, 'red');
end

function F(X, MX, DX, m, R)
        delta = R/m;
        Xn = min(X):delta/20:max(X);
        Y = 1/2 * (1 + erf((Xn - MX) / sqrt(2*DX)));
        plot(Xn, Y, 'red');
end

function plotEmpiricalF(X)
        [yy, xx] = ecdf(X);
        stairs(xx, yy);
end

end
```

4 Результаты расчётов

Mmin = -2.880000
Mmax = 1.190000
R = 4.070000
mu = -0.771000
S2 = 0.752555

```
Интервальная группировка значений выборки при m = 8
[-2.880000 : -2.371250) - 6
[-2.371250 : -1.862500) - 7
[-1.862500 : -1.353750) - 18
[-1.353750 : -0.845000) - 21
[-0.845000 : -0.336250) - 32
[-0.336250 : 0.172500) - 21
[0.172500 : 0.681250) - 8
[0.681250 : 1.190000] - 7
```

5 Графики

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

На рис. 1 представлена гистограмма и график функции плотности распределения вероятностей:

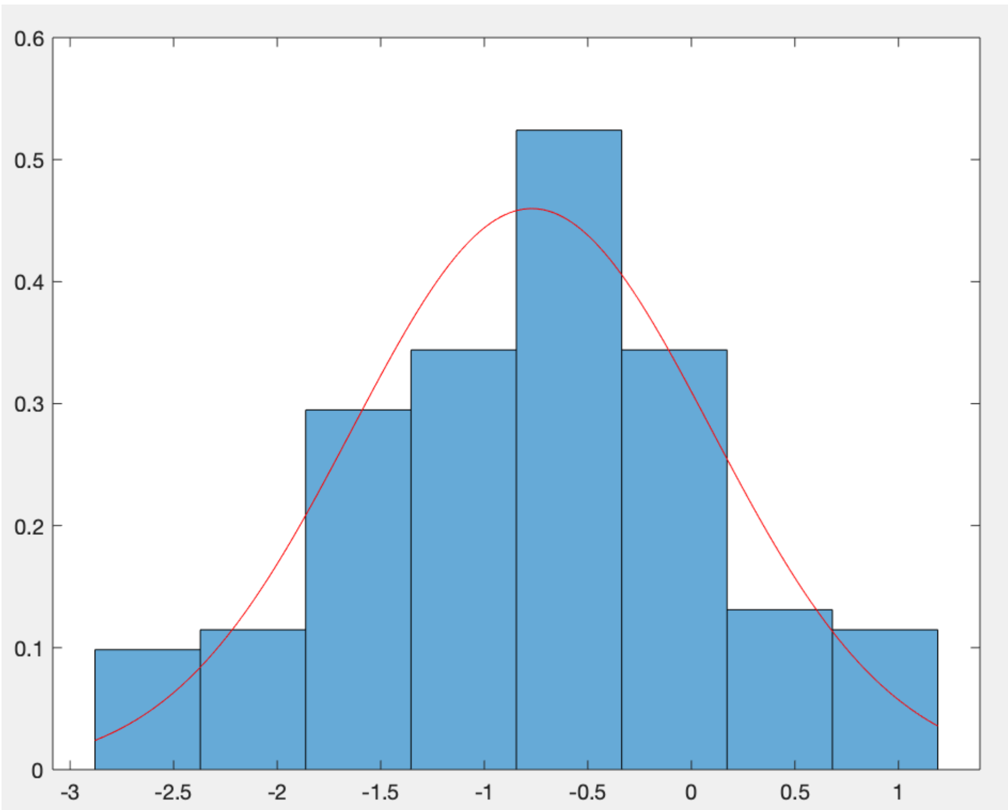

Рис. 1: Гистограмма и график функции плотности распределения вероятностей

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

На рис. 2 представлен график эмпирической функции распределения и функции распределения:

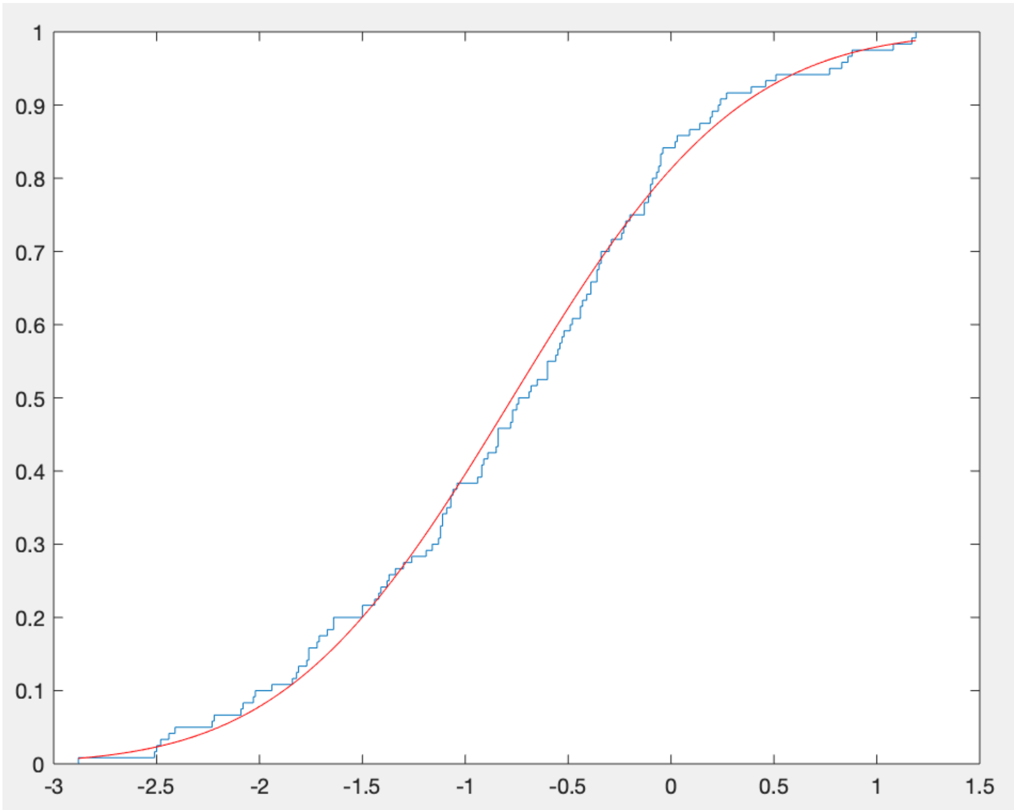

Рис. 2: График эмпирической функции распределения и функции распределения