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 предмету: «Математическая статистика»

Тема: Гистограмма и эмпирическая функция распределения

Студент:
Коротков Андрей Владимирович
Группа: ИУ7-65Б
Вариант: 8

Преподаватели:
Власов Павел Александрович
Волков Игорь Куприянович

Москва, 2020 г.

Введение

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 $\mathrm{M}X$ и дисперсии $\mathrm{D}X$;
- 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 вели- чины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.

2. Провести вычисления и построить графики для выборки из индивидуального варианта.

1. Теоретическая часть

В данной части рассмотрены формулы для вычисления величин, эмпирическая плотность, эмпирическая функция распределения и гистограмма.

1.1 Формулы для вычисления величин

Минимальное значение выборки

$$M_{\min} = \min\{x_1, \dots, x_n\}, \tag{1}$$

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

Максимальное значение выборки

$$M_{\max} = \max\{x_1, \dots, x_n\}, \tag{2}$$

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

Размах выборки

$$R = M_{\max} - M_{\min}, \tag{3}$$

где $M_{\max}$ — максимальное значение выборки,

$M_{\min}$ — минимальное значение выборки.

Оценка математического ожидания

$$\hat{\mu}(\overrightarrow{X}_n) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{4}$$

Несмещённая оценка дисперсии

$$S^2(\overrightarrow{X}_n) = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\,. \tag{5}$$

Выборочная дисперсия

$$\hat{\sigma}^2(\overrightarrow{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \tag{6}$$

Количество интервалов

$$m = [\log_2 n] + 2. \quad (7)$$

1.2 Эмпирическая плотность и гистограмма

Определение. Эмпирической плотностью распределения случайной выборки $\overrightarrow{X}_n$ называют функцию:

$$f_n(x) = \begin{cases} \frac{n_i}{n\,\Delta}, & x \in J_i,\ i = \overline{1,m}; \\ 0, & \text{иначе}\,. \end{cases}, \quad (8)$$

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

$x_{(1)} = \min\{x_1, \ldots, x_n\}$,

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

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

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

m — количество полуинтервалов интервала $J = [x_{(1)}, x_{(n)}]$,

Δ — длина полуинтервала J_i, $i = \overline{1,m}$ равная

$$\Delta = \frac{x_{(n)} - x_{(1)}}{m} = \frac{|\,J\,|}{m}, \quad (9)$$

n_i — количество элементов выборки в полуинтервале J_i, $i = \overline{1,m}$,

n — количество элементов в выборке.

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

1.3 Эмпирическая функция распределения

Определение. Пусть

1. $\overrightarrow{X}_n = (X_1, \ldots, X_n)$ — случайная выборка,
2. $\overrightarrow{x}_n = (x_1, \ldots, x_n)$ — реализация случайной выборки,
3. $n(x, \overrightarrow{x}_n)$ — количество элементов выборки $\overrightarrow{x}_n$, которые меньше x,

тогда *эмпирической функцией распределения* называют функцию

$$F_n\colon \mathfrak{R} \to \mathfrak{R}, \quad F_n(x) = \frac{n(x, \overrightarrow{x}_n)}{n}. \quad (10)$$

Замечание.

1. $Fn(x)$ обладает всеми свойствами функции распределения;

2. $Fn(x)$ кусочно-постоянна;

3. если все элементы вектора различны, то

$$F_n(x) = \begin{cases} 0, & x \le x_{(1)}; \\ \frac{i}{n}, & x_{(i)} < x \le x_{(i+1)}, \quad i = \overline{1,n-1}; \\ 1, & x > x_{(n)}. \end{cases} \tag{11}$$

4. Эмпирическая функция распределения позволяет интерпретировать выборку $\overrightarrow{x}_n$ как реализацию дискретной случайной величины $\widetilde{X}$, ряд распределения которой имеет вид:

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

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

2. Практическая часть

В данной части представлен листинг и результат работы программы.

2.1 Листинг программы

```
function lab1
    clear
    X = [7.76,6.34,5.11,7.62,8.84,4.68,8.65,6.90,8.79,6.61,6.62,7.13,6.75,7.28,7.74,7.08,5.57,8.20,7.78,7.92,6.00,4.88,6.75,6.56,7.48,8.51,9.06,6.94,6.93,7.79,5.71,5.93,6.81,5.76,5.88,7.05,7.22,6.67,5.59,6.57,7.28,6.22,6.31,5.51,6.69,7.12,7.40,6.86,7.28,6.82,7.08,7.52,6.81,7.55,4.89,5.48,7.74,5.10,8.17,7.67,7.07,5.80,6.10,7.15,7.88,9.06,6.85,4.88,6.74,8.76,8.53,6.72,7.21,7.42,8.29,8.56,9.25,6.63,7.49,6.67,6.79,5.19,8.20,7.97,8.64,7.36,6.72,5.90,5.53,6.44,7.35,5.18,8.25,5.68,6.29,6.69,6.08,7.42,7.10,7.14,7.10,6.60,6.35,5.99,6.17,9.05,6.01,7.77,6.27,5.81,7.80,9.89,4.39,6.83,6.53,8.15,6.68,6.87,6.31,6.83];

    % ----- task1 -----
    Mmin = min(X);
    Mmax = max(X);
    R = Mmax - Mmin;
    mu = getmu(X);
    sigmaSqr = getSigmaSqr(X);
    Ssqr = getS(X);

    % ----- task2 -----
    numSubIntervals = getNumSubIntervals(size(X, 2));

    % graph1
    intervals(X, numSubIntervals);
    hold on;
    f(X, mu, Ssqr, numSubIntervals, R);
    legend('гистограмма','ф-ция плотности распр-я нормальной СВ');
    hold off;

    % graph2
    figure;
    empiricF(sort(X));
    hold on;
    F(sort(X), mu, Ssqr, numSubIntervals, R);
    grid on;
    legend('эмпирич. ф-ция распр-я','ф-ция распр-я нормальной СВ');
    hold off;

    function mu = getmu(X)
        mu = sum(X)/size(X,2);
```

```
    end

    function sigma = getSigmaSqr(X)
        tempMu = getmu(X);
        sigma = sum((X - tempMu) .* (X - tempMu))/size(X,2);
    end

    function Ssqr = getS(X)
        n = size(X,2);
        Ssqr = n/ (n - 1) * getSigmaSqr(X);
    end

    function numSubIntervals = getNumSubIntervals(size)
        numSubIntervals = floor(log2(size)) + 2;
    end

    function intervals(X, m)
        sortX = sort(X);
        n = size(sortX,2);
        delta = (sortX(end) - sortX(1)) / m;
        J = sortX(1):delta:sortX(end);
        numElem = zeros(1, m);

        for i = 1:n
            for j = 1:(size(J,2) - 1)
                if (sortX(i) >= J(j) && sortX(i) < J(j+1))
                    numElem(j) = numElem(j) + 1;
                    break;
                end
            end
        end
        numElem(end) = numElem(end) + 1;

        for i = 1:size(numElem,2)
            numElem(i) = numElem(i)/(n * delta);
        end
        J = [J(1) J];
        numElem = [0 numElem 0];

        stairs(J, numElem), grid;
    end

    % Нормальная функция плотности вероятности
    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, '-.');
    end

    % Эмпирическая функция распределения
    function empiricF(X)
        [yy, xx] = ecdf(X);
```

```
        stairs(xx, yy);
    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, '--');
    end
end
```

2.2 Результат работы программы

$M_{\min} = 4.39;$

$M_{\max} = 9.89;$

$R = 5.5;$

$\hat{\mu}(\overrightarrow{X}_n) = 6.9445;$

$S^2(\overrightarrow{X}_n) = 1.1720.$

Интервальная группировка значений выборки при m = 8:

[4.3900, 5.0775)	[5.0775, 5.7650)	[5.7650, 6.4525)	[6.4525, 7.1400)
5	12	19	38
[7.1400, 7.8275)	[7.8275, 8.5150)	[8.5150, 9.2025)	[9.2025, 9.8900]
24	10	10	2

2.3 Графики

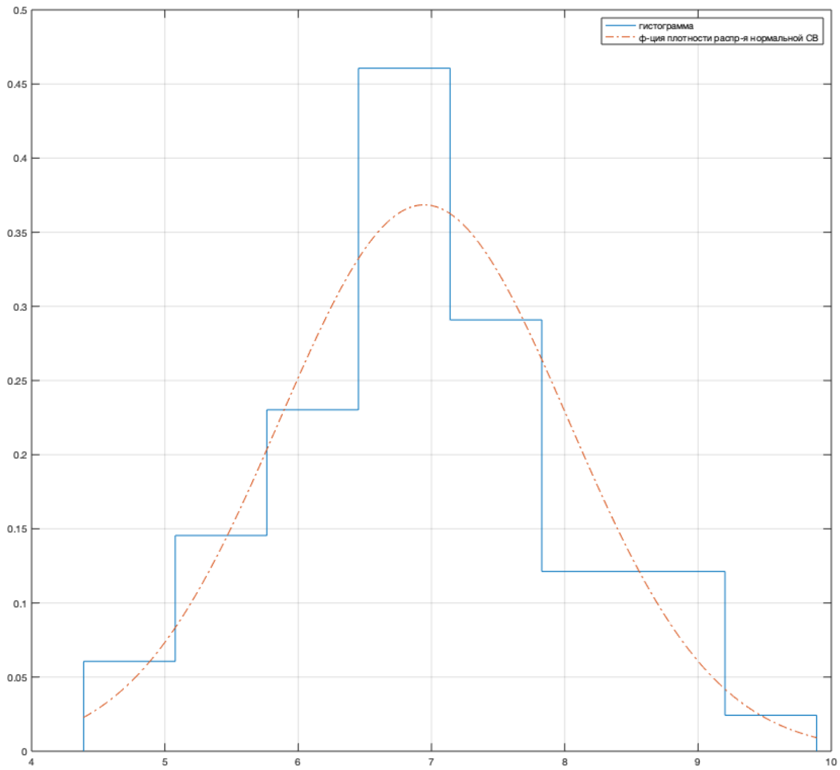

Рисунок 1. Гистограмма и график функции плотности распределения нормальной случайной величины.

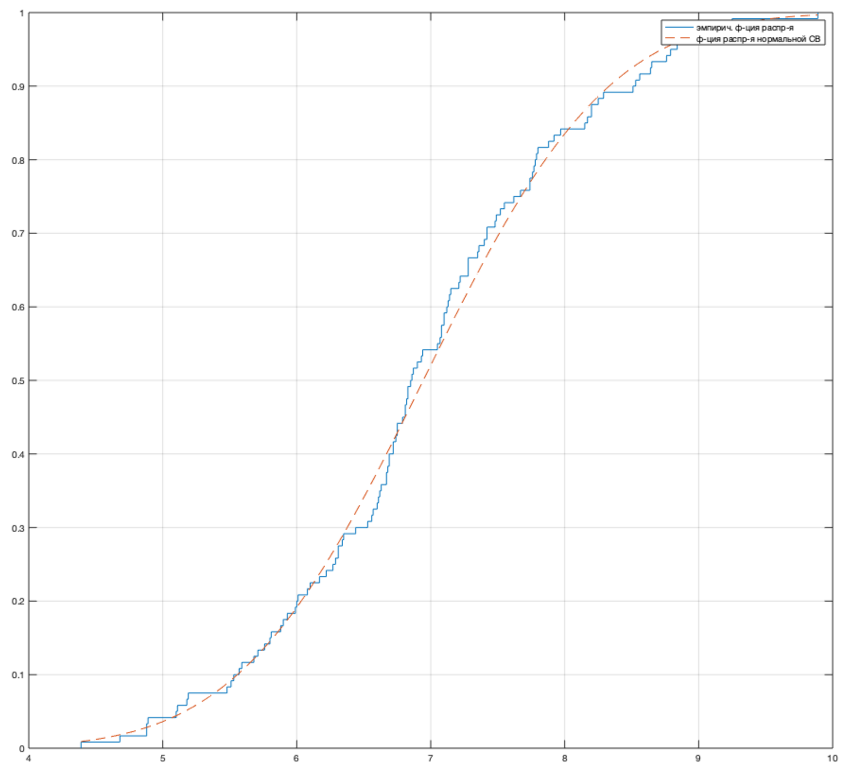

Рисунок 2. График эмпирической функции распределения и функции распределения нормальной случайной величины.