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»

Преподаватели: Волков И. К.,
Власов П. А.,
Студент: Мирзоян С.А.,
Группа: ИУ7-65Б

Москва, 2020 г.

Введение

Цель работы: построение гистограммы и эмпирической функции распределения

Содержание работы:

1. Для выборки объема n из генеральной совокупности X реализовать в виде программы на ЭВМ:

    a. вычисление максимального значения M_{max} и минимального значения M_{min}

    b. размаха R выборки;

    c. вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;

    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 математическим ожиданием $\hat{\mu}$ и дисперсией S^2.

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

Теоретическая часть

Формулы для вычислений величин

Минимальное значение выборки: $M_{min} = min\{x_1, \dots, x_n\}$,

Максимальное значение выборки: $M_{max} = max\{x_1, \dots, x_n\}$,

Размах выборки: $R = M_{max} - M_{min}$

Оценка математического ожидания: $\hat{\mu}\left(\overrightarrow{X_n}\right) = \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Несмещенная оценка дисперсии: $S^2\left(\overrightarrow{X_n}\right) = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overrightarrow{X_n}\right)^2$

Выборочная дисперсия: $\hat{\sigma}^2\left(\overrightarrow{X_n}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overrightarrow{X_n}\right)^2$

Количество интервалов: $m = [log_2 n] + 2$

Эмпирическая плотность и гистограмма

Определение: *Эмпирической плотностью распределения случайной выборки называют функцию*

$$f_n(x) = \begin{cases} \dfrac{n_i}{n \triangle}, & x \in J_i, \ i = \overline{1:m} \\ 0, & \text{иначе} \end{cases}$$

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

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

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

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

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

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

$\triangle$ – длина полуинтервала J_i, $i = \overline{1:m}$, равная

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

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

Определение: *Гистограммой называется график функции* $f_n(x)$

Эмпирическая функция распределения

Определение: Пусть

- $\overrightarrow{X_n} = (X_1, \ \ldots, X_n)$ – случайная выборка
- $\overrightarrow{x_n} = (x_1, \ \ldots, x_n)$ – реализация случайной выборки
- $n(x, \vec{x}_n)$ – количество элементов выборки $\vec{x}_n$, которые меньше x,

Тогда эмпирической функцией распределения называют функцию

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

Замечание 1.

$F_n(x)$ - обладает всеми свойствами функции распределения.

Замечание 2.

$F_n(x)$ – кусочно-постоянна.

Замечание 3.

Если все элементы вектора различны, то

$$F_n(x) = \begin{cases} 0, & x \leq x_{(1)} \\ \frac{i}{n}, & x_{(i)} < x \, x_{(i+1)}, \qquad i = \overline{1 : n-1} \\ 1, & x \geq x_{(n)} \end{cases}$$

Замечание 4.

Эмпирическая функция распределения позволяет интерпретировать выборку $\overrightarrow{x_n}$ как реализацию дискретной случайной величины $\tilde{X}$, 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 $\tilde{X}$ как приближенные значения числовых характеристик случайной величины X.

Практическая часть.

Листинг.

```
function lab1 ()
    function Mu = MathExpect(MyMassive)
         Mu = sum(MyMassive)/size(MyMassive, 2);
    end

    function Sigma = Dispersion(MyMassive)
        tmp = MathExpect(MyMassive);
        Sigma = sum((MyMassive - tmp) .* (MyMassive - tmp))/size(MyMassive, 2);
    end

     function countinterv = subIntervals(size)
         countinterv = 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

    function f(X, M, D, m, R)
            delta = R/m;
            Sigma = sqrt(D);
            Xn = min(X):delta/20:max(X);
            Y = normpdf(Xn, M, Sigma);
            plot(Xn, Y, '-*');
        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
    function emprf(Mas)
        [eY, eX] = ecdf(sort(Mas));
        stairs(eX, eY)
    end

    clear all;

    MyMassive = [11.89,9.60,9.29,10.06,9.50,8.93,9.58,6.81,8.69,9.62,9.01,10.59,10.50,11.53,9.94,8.84,8.91,6.90,9.76,7.09,11.29,11.25,10.84,10.76,7.42,8.49,10.10,8.79,11.87,8.77,9.43,12.41,9.75,8.53,9.72,9.45,7.20,9.23,8.93,9.15,10.19,9.57,11.09,9.97,8.81,10.73,9.57,8.53,9.21,10.08,9.10,11.03,10.10,9.47,9.72,9.60,8.21,7.78,10.21,8.99,9.14,8.60,9.14,10.95,9.33,9.98,9.09,10.35,8.61,9.35,10.04,7.85,9.64,9.99,9.65,10.89,9.08,8.60,7.56,9.27,10.33,10.09,8.51,9.86,9.24,9.63,8.67,8.85,11.57,9.85,9.27,9.69,10.90,8.84,11.10,8.19,9.26,9.93,10.15,8.42,9.36,9.93,9.11,9.07,7.21,8.22,9.08,8.88,8.71,9.93,12.04,10.41,10.80,7.17,9.00,9.46,10.42,10.43,8.38,9.01];

    % Задание 1
    %% поиск Min и Max
    MinMas = min(MyMassive);
    MaxMas = max(MyMassive);

    %% размах R

```

```
    R = MaxMas - MinMas;

    %% оценки мат. ожидания и дисперсии

    Mu = MathExpect(MyMassive);
    Sigma = Dispersion(MyMassive);

    %% Интервалы

    countinterv = subIntervals(size(MyMassive, 2));

    %% график 1
    %%% f(X, MX, DX, m, R) - Нормальная функция плотности вероятности
    Intervals(MyMassive, countinterv);
    hold on;
    f(MyMassive, Mu, (size(MyMassive,2)/(size(MyMassive,2) * Sigma)), countinterv, R);
    legend('Гистограмма','Функция плотности распределения нормальной случайной величины');
    hold off;

    %% график 2
    %%% F(X, MX, DX, m, R) -
    figure;
    emprf(sort(MyMassive));
    hold on;
    F(sort(MyMassive), Mu, (size(MyMassive,2)/(size(MyMassive,2) * Sigma)), countinterv, R);
    grid on;
    legend('Эмпирическая функция распределения','Функция распределения нормальной случайной величины');
    hold off;
end
```

Результаты вычислений.

$$M_{min} = 6.81$$

$$M_{max} = 12.41$$

$$R = 5.6$$

$$\hat{\mu}\left(\overrightarrow{X_n}\right) = 9.4872$$

$$S^2\left(\overrightarrow{X_n}\right) = 1.2173$$

Интервальная группировка значений выборки при m = 8:

[6.81; 7.51),

[7.51; 8.21),

[8.21; 8.91),

[8.91; 9.61),

[9.61; 10.31),

[10.31; 11.01),

[11.01; 11.71),

[11.71; 12.41),

Графики

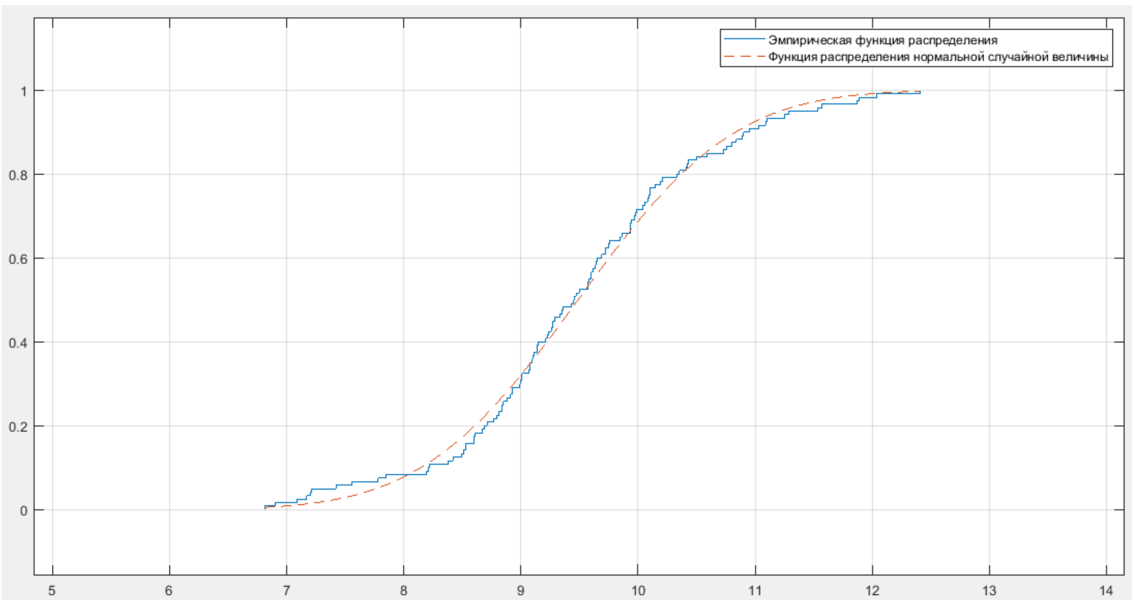

Рис.1 График эмпирической функции распределения и функции распределения нормальной случайной величины.

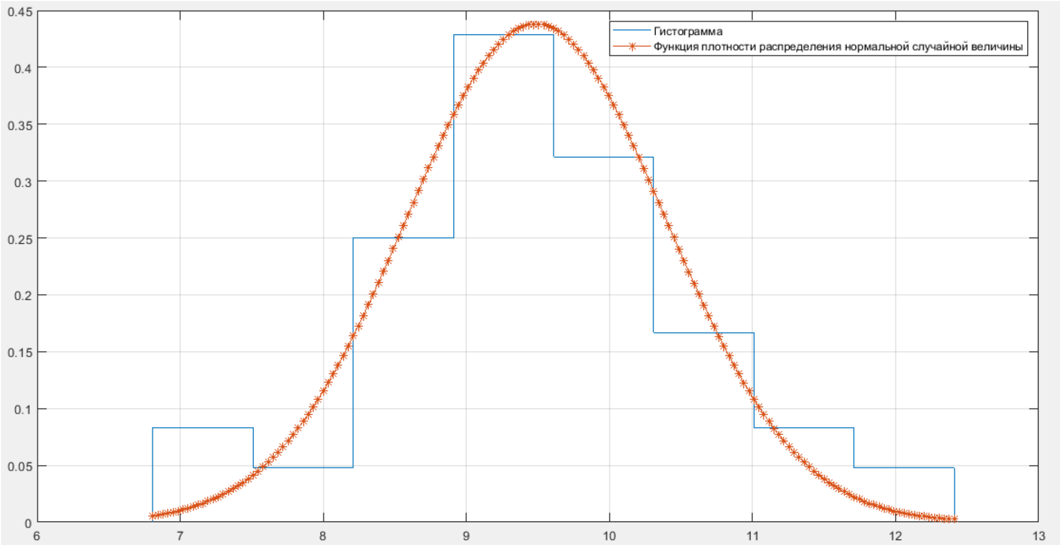

Рис.2 Гистограмма и график функции плотности распределения нормальной случайной величины.