Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э.Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э.Баумана)

Отчёт по лабораторной работе №1
По дисциплине «Математическая статистика»
Тема: «Гистограмма и эмпирическая функция распределения»
Вариант: 1

Студент: ____________________ Барсуков Н.М.

Группа: ____________________ ИУ7-66

Преподаватель: ____________________ Саркисян П.С.

Москва, 2020

Содержание

Введение

Цель работы: построение гистограммы и эмпирической функции распределения.

Содержание работы:

Для выборки объёма n из генеральной совокупности X реализовать в виде программы на ЭВМ:

1) вычисление максимального значения M_{max} и минимального значения M_{min};

2) вычисление размаха R выборки;

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

4) группировку значений выборки в $m = [log_2 n] + 2$ интервала;

5) построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;

6) построение на одной координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.

Провести вычисления и построить графики для выборки из индивидуального варианта.

Теоретическая часть

Формулы для вычисления величин:

1) максимальное значение

$$M_{max} = max(x_1, x_2, ..., x_n); \tag{1}$$

2) минимальное значение

$$M_{min} = min(x_1, x_2, ..., x_n); \tag{2}$$

3) размах выборки

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

4) выборочное среднее

$$\hat{\mu} = \frac{1}{n} * \sum_{i=1}^{n} X_i; \tag{4}$$

5) несмещённая выборочная дисперсия

$$S^2 = \frac{1}{(n-1)} * \sum_{i=1}^{n} (X_i - \hat{\mu})^2. \tag{5}$$

Предположим, что для данной выборки $\overrightarrow{x_n}$ построили интервальный статистический ряд. Выберем некоторое m и разобьём отрезок $[x_{(1)}, x_{(n)}]$.

Определение: эмпирической плотностью распределения, соответствующей реализации $\overrightarrow{x_n}$, называют функцию

$$f_n(x) \quad = \quad \begin{cases} \frac{n_i}{n * \Delta}, & x \in J_i \\ 0, & x \notin J_i \end{cases}. \tag{6}$$

Интервалы:

$$J = [x_{(1)}, x_{(n)}]. \tag{7}$$

Ширина интервалов

$$\Delta = \frac{|J|}{m}, \tag{8}$$

где число интервалов

$$m = [log_2 n] + 2. \tag{9}$$

$$J_i = [x_{(1)} + (i-1) * \Delta, x_{(1)} + i * \Delta), \tag{10}$$

где $i = \overline{1 : m-1}$.

$$J_m = [x_{(1)} + (m-1) * \Delta, x_{(1)} + m * \Delta]. \tag{11}$$

n_i - число элементов выборки, принадлежащих интервалу J_i, $i = \overline{1:m-1}$.

Определение: Гистограммой называется график этой вышеуказанной функции эмпирической плотности распределения.

Пусть $n(x, \overrightarrow{x_n})$ - количество элементов выборки $\overrightarrow{x_n}$, меньших x.

Определение: Эмипирической функцией распределения, построенной по выборке $\overrightarrow{x_n}$, называется отображение $F_n : R \to R$ по правилу $F_n = \frac{n(x, \overrightarrow{x_n})}{n}$, где $n(x, \overrightarrow{x_n})$ - количество элементов выборки $\overrightarrow{x_n}$, которые имеют значение, меньшее x.

Если все элементы выборки $\overrightarrow{x_n}$ попарно различны, то

$$F_n = \begin{cases} 0, & x \le x_{(1)} \\ \frac{i}{n}, & x \in (x_{(i)}, x_{(i+1)}], \quad i = \overline{1:n-1}. \\ 1, & x > x_{(n)} \end{cases} \tag{12}$$

Листинг программы

Текст программы (*labwork*1.*m*):

```
function lab1()
X = [-0.23,-1.03,-4.11,-0.65,-2.58,-0.79,-1.53,-0.18,-2.79,-1.97,-2.21,-1.59,-0.22,-3
-1.87,-2.30,-0.94,-0.74,-2.45,-1.40,-2.09,-0.68,0.02,-1.80,-2.25,-1.19,-2.17,-1.89,-1
-1.51,-2.11,-2.24,-0.72,0.94,-0.67,-2.44,-2.27,-1.33,-3.03,-0.42,-2.86,-2.00,-1.37,-1
-2.79,-0.21,-1.29,-2.81,-0.29,-1.55,-0.45,-1.16,-3.96,-3.77,-3.36,-1.81,0.13,-2.61,-3
-1.49,-1.89,-1.24,-0.00,-2.72,-1.69,-1.25,-1.59,0.20,-1.08,-2.42,-3.14,-2.54,-2.09,-2
-2.33,-1.97,-0.54,-1.13,-2.04,0.77,-1.03,-1.55,-1.47,-0.09,-2.11,-2.08,-1.79,-1.36,-1
X = sort(X);

Mmax = max(X);
Mmin = min(X);

fprintf('Mmin = %s\n', num2str(Mmin));
fprintf('Mmax = %s\n', num2str(Mmax));

R = Mmax - Mmin;
fprintf('R = %s\n', num2str(R));

MU = getMU(X);
fprintf('MU = %s\n', num2str(MU));

Ssqr = getSsqr(X);
fprintf('S^2 = %s\n', num2str(Ssqr));

m = getNumberOfIntervals(X);
fprintf('m = %s\n', num2str(m))

createGroup(X);
hold on;
distributionDensity(X, MU, Ssqr, m);

figure;
empiricF(X);
hold on;
distribution(X, MU, Ssqr, m);
end

function mu = getMU(X)
n = length(X);
mu = sum(X)/n;
```

```
end

function Ssqr = getSsqr(X)
n = length(X);
MX = getMU(X);
Ssqr = sum((X - MX).^2) / (n-1);
end

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

function createGroup(X)
n = length(X);
m = getNumberOfIntervals(X);

intervals = zeros(1, m+1);
numCount = zeros(1, m+1);
Delta = (max(X) - min(X)) / m;

for i = 0: m
intervals(i+1) = X(1) + Delta * i;
end

j = 1;
count = 0;
for i = 1:n
if (X(i) >= intervals(j+1))
j = j + 1;
end
numCount(j) = numCount(j) + 1;
count = count + 1;
end

graphBuf = numCount(1:m+1);
for i = 1:m+1
graphBuf(i) = numCount(i) / (n*Delta);
end

stairs(intervals, graphBuf),grid;
end

function distributionDensity(X, MX, DX, m)
R = X(end) - X(1);
```

```
delta = R/m;
Sigma = sqrt(DX);

Xn = (MX - R): delta/50 :(MX + R);
Y = normpdf(Xn, MX, Sigma);
plot(Xn, Y), grid;
end

function distribution(X, MX, DX, m)
R = X(end) - X(1);
delta = R/m;

Xn = (MX - R): delta :(MX + R);
Y = 1/2 * (1 + erf((Xn - MX) / sqrt(2*DX)));
plot(Xn, Y, 'r'), grid;
end

function empiricF(X)
[yy, xx] = ecdf(X);

stairs(xx, yy), grid;
end
```

Результаты расчётов для выборки из индивидуального варианта

Для выборки согласно варианту были получены

1) минимальное значение $M_{min} = -4.11$;
2) максимальное значение $M_{max} = 1.400$;
3) размах выборки $R = 5.510$;
4) выборочное среднее $\hat{\mu} = -1.604583$;
5) несмещённая выборочная дисперсия $S^2 = 1.034091$.

Таблица 1. Интервальный ряд для индивидуального варианта

Интервал	Число
[-4.1100; -3.4213)	4
[-3.4213; -2.7325)	11
[-2.7325; -2.0438)	26
[-2.0438; -1.3550)	33
[-1.3550; -0.6663)	26
[-0.6663; 0.0225)	15
[0.0225; 0.7112)	2
[0.7112; 1.4000]	3

В результате проведённых вычислений были получены два графика. На графике 2 представлены гистограмма и функция плотности распределения нормальной величины, на графике 1 - эмпирическая функция распределения и функция распределения нормальной случайной величины.

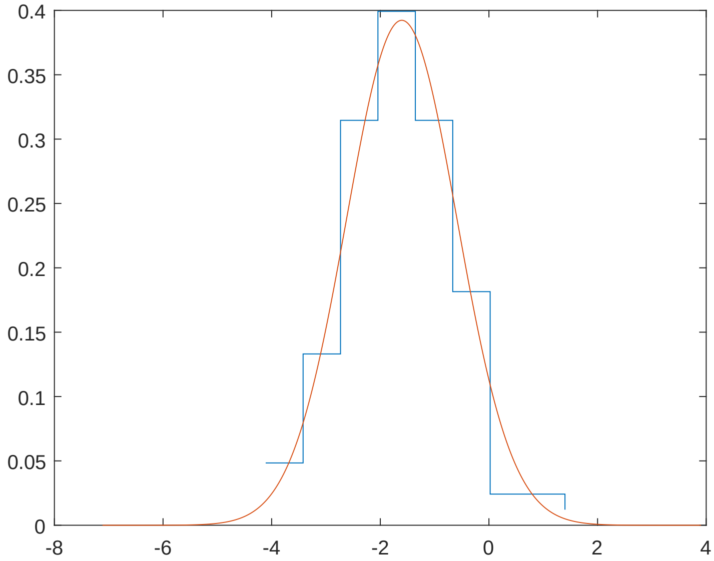

Рис. 1. Гистограмма и график функции плотности распределения нормальной случайной величины

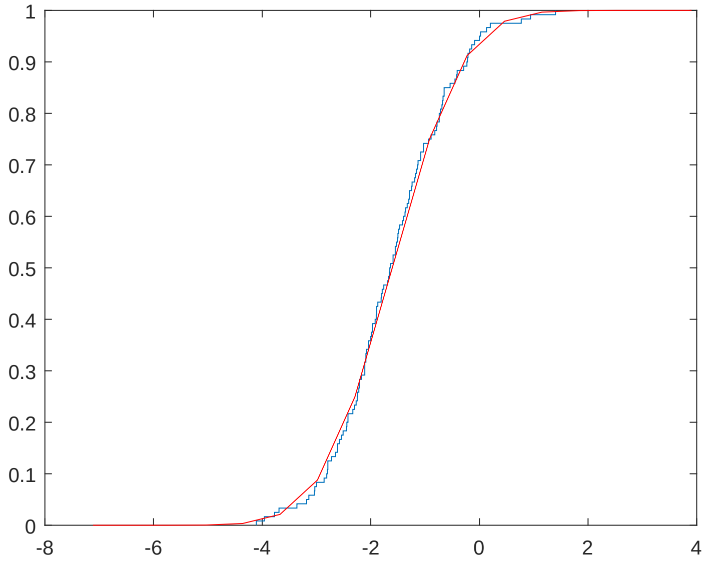

Рис. 2. График эмпирической функции распределения и функции распределения нормальной случайной величины

Заключение

В результате выполнения лабораторной работы для заданной согласно варианту выборки были получены гистограмма и эмпирическая функция распределения. Для выполнения вычислений был написан код MatLab, позволяющий вычислить интервальный ряд распределения, минимальное и максимальное значения, размах выборки, выборочное среднее, несмещённую выборочную дисперсию.