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

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

Листинг программы

Текст программы (*labwork1.m*):

```
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,-3.18,-1.18,-1.42,-1.29,-2.22,-0.82,-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,-1.14,-1.50,-1.76,-0.69,-2.21,-1.65,-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,-1.90,
-2.80,-0.89,-2.04,-1.66,-0.14,-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,-3.69,-3.00,-2.61,
-0.74,-0.41,-0.78,-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,-2.51,-2.65,-2.42,-1.30,-0.65,
1.40,-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,-1.92,-3.04,-1.08,-1.67,-2.11,-1.99,-1.64];

var_series = sort(X);
N = length(var_series);

% минимальное и максимальное значения выборки, размах
Mmin = min(var_series);
Mmax = max(var_series);
R = Mmax - Mmin;

% Мат.ожидание, дисперсия, S^2

mX = sum(var_series) / N;

S2 = 0;
for i = 1:N
temp = var_series(i) - mX;
S2 = S2 + temp^2;
end
S2 = S2 / (N - 1);

% Интервалы
m = floor(log2(N)) + 2;
Delta = R / m;

intervals_x = zeros(1, m);
intervals_count = zeros(1, m);

for i = 0:m
intervals_x(i+1) = (var_series(1)+ i * Delta);
```

```
end

for i = 1:N
for j = 1:m
if var_series(i) >= intervals_x(j) && var_series(i) < intervals_x(j+1)
intervals_count(j) = intervals_count(j) + 1;
break
end
if var_series(i) == var_series(end)
intervals_count(m) = intervals_count(m) + 1;
break
end
end
end

% вывод границ интервалов и соответствующих чисел
intervals_x
intervals_count

% работа с графиками

step = (max(var_series) - min(var_series)) / N;
% гистограмма
graph_x = min(var_series) : Delta : max(var_series);
graph_y = intervals_count / N;
graph_y(end + 1) = graph_y(end);
stairs(graph_x, graph_y);
hold on;
% плотность нормального распределения
graph_x = min(var_series) : step : max(var_series);
graph_y = normpdf(graph_x, mX, sqrt(S2));
plot(graph_x, graph_y), grid;
figure;
% эмпирическая функция распределения
[graph_y, graph_x] = ecdf(var_series);
graph_x = [[min(var_series) - R / 2], graph_x', [max(var_series) + R / 2]];
graph_y = [[0], graph_y', [1]];
stairs(graph_x, graph_y);
hold on;
% нормальная функция распределения
graph_x = min(var_series) - R / 2 : step : max(var_series) + R / 2;
graph_y = (1 + erf((graph_x - mX) / sqrt(2 * S2))) / 2;
plot(graph_x, graph_y), grid;
```

```
% вывод вычисленных значений
fprintf('Mmin = %f \n', Mmin)
fprintf('Mmax = %f \n', Mmax)
fprintf('R = %f \n', R)
fprintf('mX = %f \n', mX)
fprintf('S2 = %f \n', S2)
```

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.

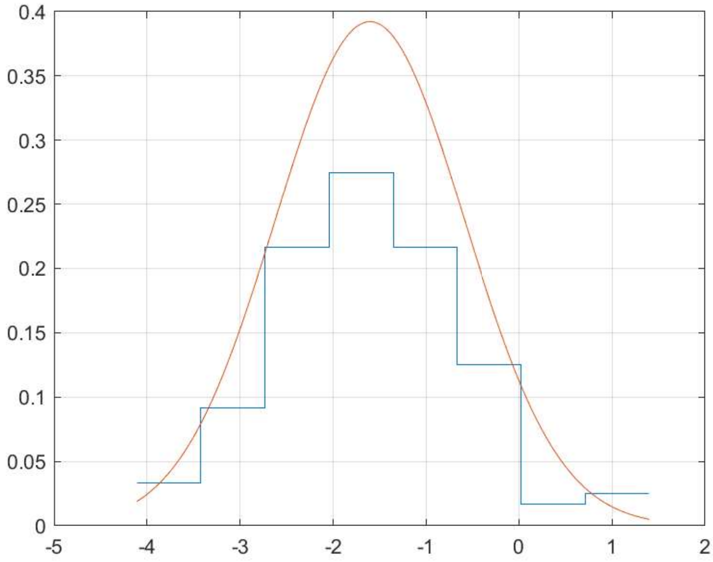

Рис. 1. Гистограмма и график функции плотности распределения нормальной случайной величины

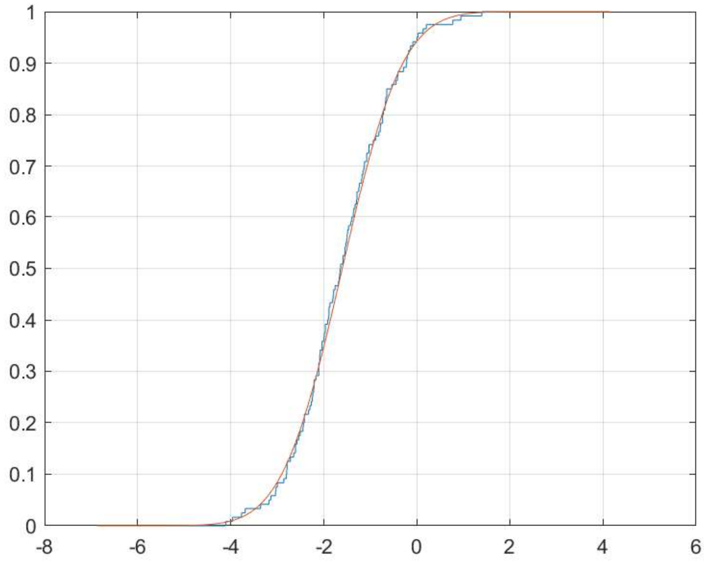

Рис. 2. График эмпирической функции распределения и функции распределения нормальной случайной величины

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.