|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **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»

**ОТЧЕТ**

*к лабораторной работе №1*

*По курсу: «Математическая статистика»*

*На тему: «Гистограмма и эмпирическая функция   
распределения»*

Студентка ИУ7-65Б

Оберган Т.М

14 вариант

Преподаватели

Власов П.А.

Волков И.К.

*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
   1. вычисление максимального значения 𝑀max и минимального значения 𝑀min;
   2. размаха 𝑅 выборки;
   3. вычисление оценок и 𝑆2 математического ожидания M𝑋 и дисперсии D𝑋;
   4. группировку значений выборки в 𝑚 = [log2 𝑛] + 2 интервала;
   5. построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием и дисперсией 𝑆2;
   6. построение на другой координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием и дисперсией 𝑆2.
2. Провести вычисления и построить графики для выборки из индивидуального варианта

1. Формулы для вычисления величин

**Минимальное и максимальное значение выборки:**

(1)

(2)

где — реализация случайной выборки.

**Размах выборки:**

(3)

где — максимальное значение выборки, — минимальное значение выборки.

**Оценка математического ожидания:**

(4)

**Выборочная дисперсия:**

(5)

**Несмещённая оценка дисперсии:**

(6)

2. Эмпирическая плотность и гистограмма

*Определение.* Эмпирической плотностью распределения случайной выборки называют функцию:

(7)

где — полуинтервал из

— количество полуинтервалов интервала ,

— длина полуинтервала равная

— количество элементов выборки в полуинтервале ,

— количество элементов в выборке.

*Определение.* График функции называют гистограммой.

3. Эмпирическая функция распределения

*Определение.* Пусть

* 1. — случайная выборка,
  2. — реализация случайной выборки,
  3. — количество элементов выборки , которые меньше ,

тогда эмпирической функцией распределения называют функцию

. (8)

*Замечание.*

1. обладает всеми свойствами функции распределения;
2. кусочно-постоянна;
3. если все элементы вектора различны, то

(9)

1. Эмпирическая функция распределения позволяет интерпретировать выборку как реализацию дискретной случайной величины , ряд распределения которой имеет вид:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| P |  | … |  |

Это позволяет рассматривать числовые характеристики случайной величины как приближенные значения числовых характеристик случайной величины .

4. Листинг программы

function lab1

clear

X = [3.38,1.21,1.85,2.24,4.17,2.99,4.81,2.71,2.70,4.41,3.21,3.15,2.77,4.05,3.89,1.56,2.78,2.04,2.82,3.28,2.63,1.89,3.57,3.15,3.80,5.40,3.25,2.04,2.61,5.06,2.87,2.66,4.80,3.86,0.09,2.45,2.40,2.14,1.69,2.36,5.44,2.77,1.94,2.55,3.97,1.88,3.01,4.21,4.74,2.02,2.38,2.46,3.51,2.89,1.57,3.53,0.77,3.31,3.58,2.77,3.61,3.71,2.38,3.06,4.29,4.76,1.69,1.59,3.21,2.74,3.99,3.53,3.52,2.84,1.21,2.82,4.34,3.65,2.22,2.87,3.14,3.58,1.96,3.41,3.85,1.96,3.02,4.22,3.10,2.68,3.67,1.70,5.47,5.02,2.52,3.09,2.19,4.44,2.33,2.27,3.34,3.05,4.35,3.58,3.43,4.49,3.57,3.20,1.53,3.53,3.53,1.27,3.40,4.53,2.21,3.28,3.50,2.01,3.30,1.86];

Mmin = min(X)

Mmax = max(X)

R = Mmax - Mmin

mu = getmu(X)

Ssqr = getS(X)

m = getNIntervals(size(X, 2))

% graph1

findIntervals(X, m)

hold on;

f(X, mu, Ssqr, mu, R);

legend('bar chart', 'density function');

hold off;

% graph2

figure;

empericalF(sort(X));

hold on;

F(sort(X), mu, Ssqr, m, R);

grid on;

legend('empirical distribution function', 'distribution function');

legend('Location', 'southeast')

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 m = getNIntervals(size)

m = floor(log2(size)) + 2;

end

function findIntervals(X, m)

sortX = sort(X);

n = size(sortX,2);

delta = (sortX(end) - sortX(1)) / m;

J = sortX(1):delta:sortX(end)

nEl = zeros(1, m);

for i = 1:n

for j = 1:(size(J,2) - 1)

if (sortX(i) >= J(j) && sortX(i) < J(j+1))

nEl(j) = nEl(j) + 1;

break;

end

end

end

nEl(end) = nEl(end) + 1

for i = 1:size(nEl,2)

nEl(i) = nEl(i)/(n \* delta);

end

J = [J(1) J];

nEl = [0 nEl 0];

stairs(J, nEl), grid;

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, '-.');

end

function empericalF(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

5. Результаты расчетов

a)

b)

c)

d)



