*Государственное образовательное учреждение высшего профессионального образования*

*«Московский государственный технический университет имени Н. Э. Баумана»*

ОТЧЕТ

По лабораторной работе №1

По курсу «Математическая статистика»

Тема: «Гистограмма и эмпирическая функция распределения»

Вариант 8

|  |  |
| --- | --- |
| Студент: | Иванов И.В. |
| Группа: | ИУ7-63 |
| Преподаватель: | Власов П. А. |

Москва, 2015

**1. Формулы вычисления**

**2. Определения**

**Эмпирическая плотность распределения,** соответствующая реализации случайной выборки из генеральной совокупности – функция , где  
 - размер интервала ,  
,  
,  
 – число элементов выборки, принадлежащих интервалу .

**Гистограмма** – график функции (которая представляет собой кусочно-постоянную функцию на промежутке ).

**Эмпирическая функция распределения** = скалярная функция где – количество элементов выборки , меньших , а – объём этой выборки.  
Если все элементы выборки различны, то функцию можно задать формулой

**3. Листинг программы**

function Lab1()

sample = importdata('data.txt');

sample = sort(sample);

[Min, Max, span] = **CalcMinMaxSpan**(sample);

fprintf('min: %.6f | max: %.6f\n', Min, Max);

fprintf('span (R): %.6f\n', span);

[exp, disp] = **CalcExpDisp**(sample);

fprintf('mat. expectation: %.6f\ndispersion\n: %.6f', exp, disp);

[m, delta, group] = **GroupSample**(sample, span);

for i = 1:m

if i == 1

fprintf('[%.6f', Min);

else

fprintf('[%.6f', group(i-1, 1));

end

fprintf(' .. %.6f', group(i, 1));

if i == m

fprintf(']');

else

fprintf(')');

end

fprintf(': %d elements\n', group(i, 2));

end

figure(1);

**PlotChartAndDensity**(group, delta, sample, exp, disp);

Figure(2);

**PlotEmpiricAndDistribution**(sample, exp, disp);

end

function [Min, Max, span] = CalcMinMaxSpan(sample)

%% вычисление минимального и максимального значений выборки,

%вычисление размаха выборки

Min = min(sample);

Max = max(sample);

span = Max - Min;

end

function [exp, disp] = CalcExpDisp(sample)

%% вычисление значений \mu\cap и S^2 математического ожидания и дисперсии

n = length(sample);

exp = sum(sample) / n;

disp = sum((sample - exp).^2) / (n-1);

end

function [m, delta, group] = GroupSample(sample, span)

%% группировка значений выборки в

% $x^2+m=[log\_{2}(n)]+2$

%интервала

n = length(sample);

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

delta = span / m;

group = zeros(m, 2);

%указываем границы интервалов

for j = 1:m

group(j,1) = sample(1)+delta\*j;

end

%раскидываем элементы по интервалам

j = 1;

i = 1;

border = sample(1)+delta;

while i < n

if sample(i) >= border && border < sample(n)

border = border + delta;

j = j + 1;

continue; %чтобы корректно обрабатывать ситуациии, когда в

%интервал не попадает ни один элемент выборки

end

group(j, 2) = group(j, 2) + 1;

i = i + 1;

end

group(m, 2) = group(m, 2) + 1; %последний (n-й) элемент всегда будет

%принадлежать последнему (m-му) интервалу

end

function PlotChartAndDensity(group, delta, sample, exp, disp)

%% построение на одной координатной плоскости гистограммы и графика функции

%плотности распределения вероятностей нормальной случайной величины с

%математическим ожиданием M и дисперсией S2

%модифицируем Х для гистограммы - чтобы значения отображались для середин

%интервалов

m = length(group);

gist = zeros(m,2);

gist(1,1) = (sample(1) + group(1,1)) ./ 2;

for i = 2:m

gist(i,1) = (group(i-1,1) + group(i,1)) ./ 2;

end

%модифицируем Y для гистограммы - количество\_попаданий / (n\*delta)

n = length(sample);

for i = 1:m

g = group(i,2);

g = g ./ (n\*delta);

gist(i,2) = g;

end

%вычисляем значения функции плотности распределения для всех Х из выборки

F = normpdf(sample, exp, disp);

%отображаем значения

stairs(gist(:,1), gist(:,2));

hold on;

plot(sample, F,'r'), grid;

hold off;

end

function PlotEmpiricAndDistribution(sample, exp, disp)

%% построение на другой координатной плоскости графика эмпирической функции

%распределения и функции распределения нормальной случайной величины с

%математическим ожиданием M и дисперсией S2

%получаем функцию распределения для нормальной случайной величины

F = normcdf(sample, exp, disp);

%строим эмпирическую функцию распределения

n = length(sample);

E = zeros(n);

for i = 1:n

E(i) = F\_Empiric(sample(i), sample);

end

%для сравнения, упрощенный вариант эмпирической

% ! при этом в выборке есть повторяющиеся значения

E2 = zeros(n);

for i = 1:n

E2(i) = (i-1)./n;

end

%строим графики

hold on;

plot(sample, F), grid;

plot(sample, E, 'r'), grid;

plot(sample, E2, 'k'), grid;

hold off;

end

function f = F\_Empiric(x, sample)

n = length(sample);

count = 0;

for i = 1:n

if sample(i) < x

count = count+1;

end

end

f = count / n;

end

**4. Результаты работы для индивидуальной выборки**

>> Lab1

min: 4.390000 | max: 9.890000

span (R): 5.500000

mat. expectation: 6.944500

dispersion: 1.171956

[4.390000 .. 5.077500): 5 elements

[5.077500 .. 5.765000): 12 elements

[5.765000 .. 6.452500): 19 elements

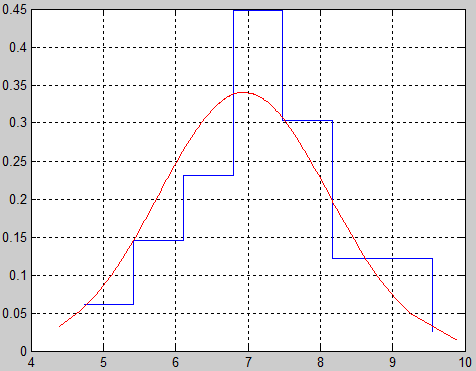
[6.452500 .. 7.140000): 37 elements

[7.140000 .. 7.827500): 25 elements

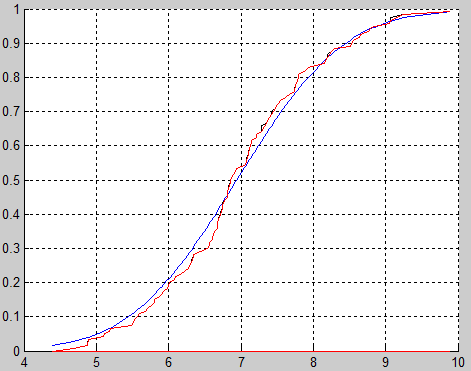
[7.827500 .. 8.515000): 10 elements

[8.515000 .. 9.202500): 10 elements

[9.202500 .. 9.890000]: 2 elements



**Рис. 1: Гистограмма и функция плотности**



**Рис.2: функция распределения (син.),**

**эмпирическая функция распределения (крас.)**