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 по лабораторной работе №1 по дисциплине «Математическая статистика»

Тема Гистрограмма и эмпирическая функция распределения

Студент Нгуен Фыок Санг

Группа ИУ7-66Б

Оценка (баллы) ____________________

Преподаватель Саркисян П. С.

Москва — 2021 г.

Задание

Цель работы

Построение гистограммы и эмпирической функции распределения.

Постановка задачи

1. Для выборки объёма 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 с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.
2. Провести вычисления и построить графики для выборки из индивидуального варианта.

Теоретические сведения

Формулы для вычисления величин

Минимальное и максимальное значения выборки

$$\begin{gathered} M_{\max} = X_{(n)} \\ M_{\min} = X_{(1)} \end{gathered} \tag{1}$$

Размах выборки

$$R = M_{\max} - M_{\min}. \tag{2}$$

Оценки математического ожидания и дисперсии

$$\begin{gathered} \hat{\mu}(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \\ S^2(\vec{X}_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \end{gathered} \tag{3}$$

Определение эмпирической плотности и гистограммы

Пусть $\vec{x}$ – выборка из генеральной совокупности X. Если объем n этой выборки велик, то значения x_i группируют в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ делят на m равновеликих частей:

$$J_i = [x_{(1)} + (i-1)\cdot\Delta, x_{(1)} + i\cdot\Delta), i = \overline{1; m-1}$$

$$J_m = [x_{(1)} + (m-1)\cdot\Delta, x_{(n)}]$$

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

Интервальным статистическим рядом называют таблицу:

где n_i – количество элементов выборки $\vec{x}$, которые $\in J_i$.

J_1	...	J_i	...	J_m
n_1	...	n_i	...	n_m

Обычно выборку разбивают на $m = [\log_2 n] + 2$ интервалов, где n – размер выборки.

Гистограмма – это график эмпирической плотности.

Эмпирической плотностью, отвечающей выборке $\vec{x}$, называют функцию:

$$\hat{f}(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1;m} \\ 0, \text{иначе} \end{cases} \tag{4}$$

где J_i – полуинтервал статистического ряда, n_i – количество элементов выборки, входящих в полуинтервал, n – количество элементов выборки.

Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, ..., x_n)$ – выборка из генеральной совокупности X. Обозначим $n(x, \vec{x})$ – число элементов вектора $\vec{x}$, которые имеют значения меньше x.

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную как:

$$F_n(x) = \frac{n(x, \vec{x})}{n} \tag{5}$$

Результаты работы программы

Код программы

```
function lab1()
    clear all;
   X = [10.06,8.32,8.50,8.82,6.02,6.44,7.90,7.85,5.90,7.62,8.66,...
    6.38,7.24,8.21,6.82,7.43,6.06,8.21,9.07,5.85,6.72,8.17,...
    8.53,8.68,7.21,8.43,8.77,7.27,5.79,9.78,6.44,7.24,6.83,...
    6.61,7.58,10.15,8.82,7.87,7.35,9.60,5.82,6.65,10.15,6.92,...
    6.77,9.35,6.92,7.76,6.45,7.47,6.99,9.95,7.22,7.38,7.87,...
    6.24,8.00,8.47,7.25,7.03,7.45,6.75,7.37,7.98,9.58,8.91,...
    6.14,8.19,5.07,7.47,7.29,8.78,7.86,7.82,10.09,8.54,7.21,...
    8.57,6.67,9.82,9.26,9.69,8.39,8.26,7.44,6.58,8.45,7.49,...
    7.16,9.17,8.16,8.38,7.60,8.53,6.10,7.39,7.70,8.45,7.73,...
    9.21,8.02,7.62,6.90,9.55,5.73,7.21,6.14,7.54,9.87,8.14,...
    8.16,7.50,7.60,6.25,7.03,7.07,6.61,9.68,7.65,8.32];
   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));

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

   S2 = getS2(X);
    fprintf('S^2 = %s\n', num2str(S2));

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


    createGroup(X);
    hold on;
```

```
  distributionDensity(X, MU, S2, m);

  figure;
  empiricF(X);
  hold on;
  distribution(X, MU, S2, m);
end

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

function S2 = getS2(X)
  n = length(X);
  MX = getMU(X);
  S2 = 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);

  MinX = min(X);
  Delta = (max(X) - min(X)) / m;
  fprintf('Delta = %s\n', num2str(Delta));

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

  j = 1;
  count = 0;
  for i = 1:n
    while (and( j < m, X(i) >= intervals(j+1)))
      j = j + 1;
    end
    numCount(j) = numCount(j) + 1;
    count = count + 1;
  end

  for i = 1:m-1
    fprintf('[%5.2f; %5.2f) ', intervals (i), intervals(i+1));
```

```
  end
  fprintf('[%5.2f, %5.2f]\n', intervals(m), intervals(m+1));

  for i = 1:m
    fprintf('%8d          ', numCount(i));
  end
  fprintf('\n\n');

  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 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)
  [y, x] = ecdf(X);
  stairs(x, y), grid;
end
```

Результаты расчётов

$$M_{\min} = 5.07$$
$$M_{\max} = 10.15$$
$$R = 5.08$$
$$\hat{\mu}(\vec{x}_n) = 7.7596$$
$$S^2(\vec{x}_n) = 1.2979$$
$$m = 8$$

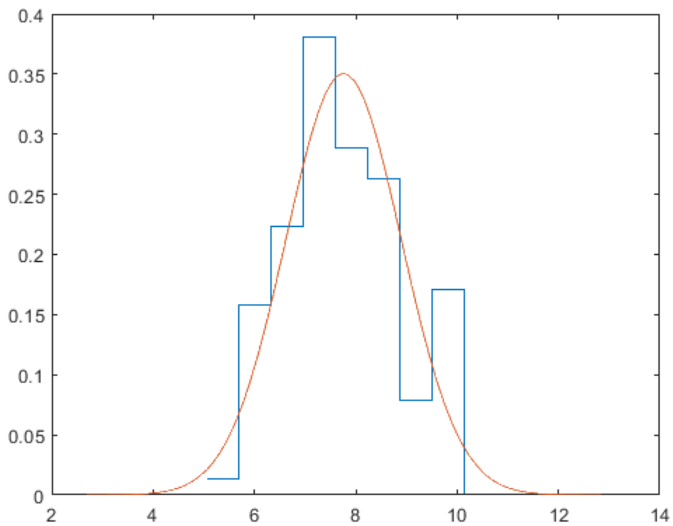

Рис. 1: Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочными мат. ожиданием и дисперсией

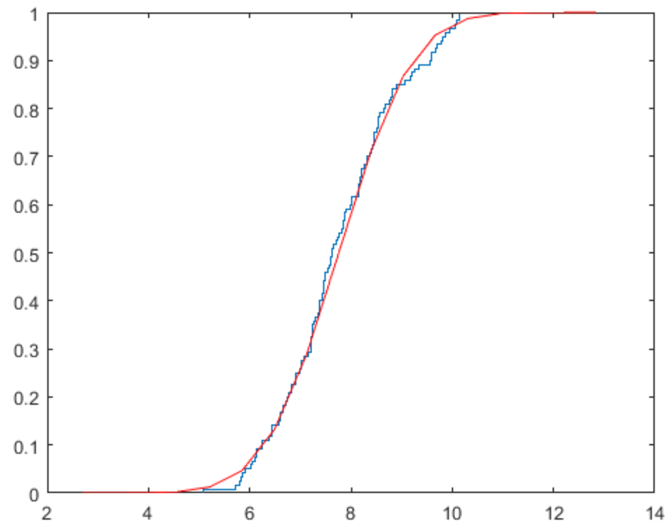

Рис. 2: График эмпирической функции распределения и функции распределения нормальной случайной величины с выборочными мат. ожиданием и дисперсией