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

Лабораторная работа №1 по курсу “Математическая статистика” по теме “ Гистограмма и эмпирическая функция распределения”

Студент:	Уласик Е.А.
Группа:	ИУ7-61
Вариант:	21

2020 г.

Постановка задачи

Цель работы: построение гистограммы и эмпирической функции распределения.

Содержание работы:

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.

1. Формулы для вычисления величин

a. Минимальное значение выборки

$M_{min} = \min\{x_1, \dots, x_n\},$ (1)

где $(x_1, \dots x_n)$ – реализация случайной выборки

b. Минимальное значение выборки

$M_{max} = \max\{x_1, \dots, x_n\},$ (2)

где $(x_1, \dots x_n)$ – реализация случайной выборки

c. Размах выборки

$R = M_{max} - M_{min},$ (3)

где M_{max} – максимальное значение выборки, M_{min} – минимальное значение выборки.

d. Оценка математического ожидания

$$\hat{\mu}(\overrightarrow{X_n}) = \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

e. Выборочная дисперсия

$$\hat{\sigma}^2(\overrightarrow{X_n}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2$$

f. Несмещённая оценка дисперсии

$$S^2(\vec{X}_n) = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

2. Эмпирическая плотность и гистограмма

Эмпирической плотностью распределения вероятностей называется функция, имеющая вид

$$f_n(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i \\ 0, x \notin [x_1; x_n] \end{cases}, \text{где}$$

J_i – i-тый интервал

x_i – наименьший элемент выборки

x_n – наибольший элемент выборки

n_i – число элементов $\bar{X}_n$ в J_i

Δ – нормированный коэффициент

Длина полуинтервала равна:

$$\Delta = \frac{x_n - x_1}{m}$$

Количество размахов определяется по формуле: $m = [log_2 n] + 2$

Участки разбиений – это полуинтервалы [;) за исключением последнего [:].

Эмпирическую плотность распределения вероятностей называют ещё гистограммой.

3. Эмпирическая функция распределения

Пусть $\vec{x}_n = (x_1, \dots, x_n)$ – выборка из генеральной совокупности

$n(x, \vec{x}_n)$ – количество элементов выборки $\vec{x}_n$, которые имеют значения меньше, чем x.

Эмпирической функцией распределения называют функцию:

$$F_n : \Re \to \Re, F_n(x) = \frac{n(x, \vec{x}_n)}{n}$$

Замечания:

1) F_n обладает всеми свойствами функции распределения

2) F_n кусочно постоянная и скачкообразная изменяет свои значения в точках статистического ряда выборки.

3) $F_n = \begin{cases} 0, x < x_1 \\ \frac{i}{n}, x_i < x \leq x_{i+1}, i = \overline{1, n} \\ 1, x > x_n \end{cases}$

4) Эмпирическая функция распределения $F_n(x)$ позволяет интерпретировать выборку $\vec{x}_n$ как реализацию дискретного случайного вектора $\tilde{X}$ как приближённые значения числовых характеристик случайной величины X.

$\tilde{X}$	$x_{(1)}$	…	$x_{(n)}$
P	*1/n*	…	*1/n*

В дальнейшем это позволяет рассмотреть числовые характеристики случайной величины $\tilde{X}$ как приближённые значения числовых характеристик случайной величины X.

4. Листинг программы

```
function lab1()
    X = [-14.34,-16.97,-14.09,-14.74,-16.69,-13.85,-15.55,-14.62,-13.30,-15.52,-
14.75,-16.51,-17.15,-16.87,-15.06,-13.60,-14.48,-14.71,-14.17,-13.88,-14.55,-
15.37,-14.81,-16.05,-17.06,-15.86,-15.12,-15.98,-14.16,-15.81,-15.06,-16.19,-
16.22,-16.19,-14.87,-15.62,-15.86,-15.25,-16.34,-14.44,-14.72,-15.17,-15.24,-
14.44,-15.93,-14.87,-16.53,-15.76,-15.12,-12.91,-16.06,-16.06,-14.89,-15.57,-
13.59,-16.84,-13.88,-14.33,-15.45,-16.58,-16.05,-14.34,-13.55,-16.78,-14.15,-
14.28,-14.40,-13.98,-16.23,-15.35,-14.77,-15.61,-15.59,-15.64,-14.76,-17.18,-
15.13,-15.01,-14.21,-13.91,-16.55,-15.44,-14.03,-16.44,-15.57,-15.07,-16.28,-
16.30,-15.74,-14.03,-14.85,-15.73,-15.81,-14.42,-14.14,-15.14,-15.49,-16.42,-
14.22,-14.20,-17.17,-15.82,-14.96,-14.75,-14.98,-13.64,-14.00,-17.29,-14.51,-
16.18,-15.70,-15.07,-14.28,-14.55,-13.85,-15.36,-15.74,-14.61,-16.32,-15.34];

    X = sort(X);

    minX = X(1);
    fprintf('Mmin = %s\n', num2str(minX));

    maxX = X(end);
    fprintf('Mmax = %s\n', num2str(maxX));

    R = maxX - minX;
    fprintf('Размах R = %s\n', num2str(R));

    mx = expectation(X);
    fprintf('Матожидание = %s\n', num2str(mx));

    sigma = dispersion(X);
    fprintf('Дисперсия = %s\n', num2str(sigma));

    s = unbiasedDispersion(X);
    fprintf('Несмещённая оценка: %s\n', num2str(s));

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

    intervals(X, m);
    hold on;
    f(X, mx, s, m);

    figure;
    empiricF(X);
    hold on;
    F(X, mx, s, m);

end

function X = readFromFile()
    X = csvread('data.csv');
end

function m = subintervals(size)
    m = floor(log2(size) + 2);
end

function mu = expectation(X)
    n   = length(X);
    sum = 0;

    for i = 1:n
        sum = sum + X(i);
    end

    mu = sum / n;
```

```
mu  = expectation(X);

    sigmaSqr = sum / n - mu^2;
end

function sSqr = unbiasedDispersion(X)
    sigmaSqr = dispersion(X);
    n        = length(X);

    sSqr = n / (n - 1) * sigmaSqr;
end

function intervals(X, m)
    count = zeros(1, m+2);
    delta = (X(end) - X(1)) / m;

    J = X(1):delta:X(end);
    fprintf('%d\n', X(end));
    J(length(J)+1) = 20;

    j = 1;
    n = length(X);

    for i = 1:n
        if (j ~= m)
            if ((not (X(i) >= J(j) && X(i) < J(j+1))))
                j = j + 1;
                fprintf('[%.2f;%.2f)\t', J(j-1), J(j));
            end
        end
        count(j) = count(j) + 1;
    end
    fprintf('[%2.2f;%2.2f]\n', J(m), J(m + 1));

    Xbuf = count(1:m+2);
          for i = 1:m+2
              Xbuf(i) = count(i) / (n*delta);
          end

          stairs(J, Xbuf), grid;
      end

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

          Xn = (MX - R): delta/50 :(MX + R);
          Xn(length(Xn)+1) = 20;
          Y = normpdf(Xn, MX, Sigma);

          plot(Xn, Y);
      end

      function F(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');
      end

      function empiricF(X)
          [yy, xx] = ecdf(X);
          yy(length(yy)+1) = 1;
          xx(length(xx)+1) = 20;

          stairs(xx, yy);
      end
```

4. Результат расчётов

Выборка:

X=(-14.34,-16.97,-14.09,-14.74,-16.69,-13.85,-15.55,-14.62,-13.30,-15.52,-14.75,-16.51,-17.15,-16.87,-15.06,-13.60,-14.48,-14.71,-14.17,-13.88,-14.55,-15.37,-14.81,-16.05,-17.06,-15.86,-15.12,-15.98,-14.16,-15.81,-15.06,-16.19,-16.22,-16.19,-14.87,-15.62,-15.86,-15.25,-16.34,-14.44,-14.72,-15.17,-15.24,-14.44,-15.93,-14.87,-16.53,-15.76,-15.12,-12.91,-16.06,-16.06,-14.89,-15.57,-13.59,-16.84,-13.88,-14.33,-15.45,-16.58,-16.05,-14.34,-13.55,-16.78,-14.15,-14.28,-14.40,-13.98,-16.23,-15.35,-14.77,-15.61,-15.59,-15.64,-14.76,-17.18,-15.13,-15.01,-14.21,-13.91,-16.55,-15.44,-14.03,-16.44,-15.57,-15.07,-16.28,-16.30,-15.74,-14.03,-14.85,-15.73,-15.81,-14.42,-14.14,-15.14,-15.49,-16.42,-14.22,-14.20,-17.17,-15.82,-14.96,-14.75,-14.98,-13.64,-14.00,-17.29,-14.51,-16.18,-15.70,-15.07,-14.28,-14.55,-13.85,-15.36,-15.74,-14.61,-16.32,-15.34)

Результат:

$M_{min} = -17.29$;

$M_{max} = -12.91$;

Размах R = 4.38;

$\hat{\mu}(X)$ = –15.2209;

$\hat{\sigma}^2(X)$= 0.95996;

$S^2(X)$= 0.96803;

m = 8

[-17.29;-16.74) [-16.74;-16.20) [-16.20;-15.65) [-15.65;-15.10) [-15.10;-14.55) [-14.55;-14.00) [-14.00;-13.46) [-13.46;-12.91]

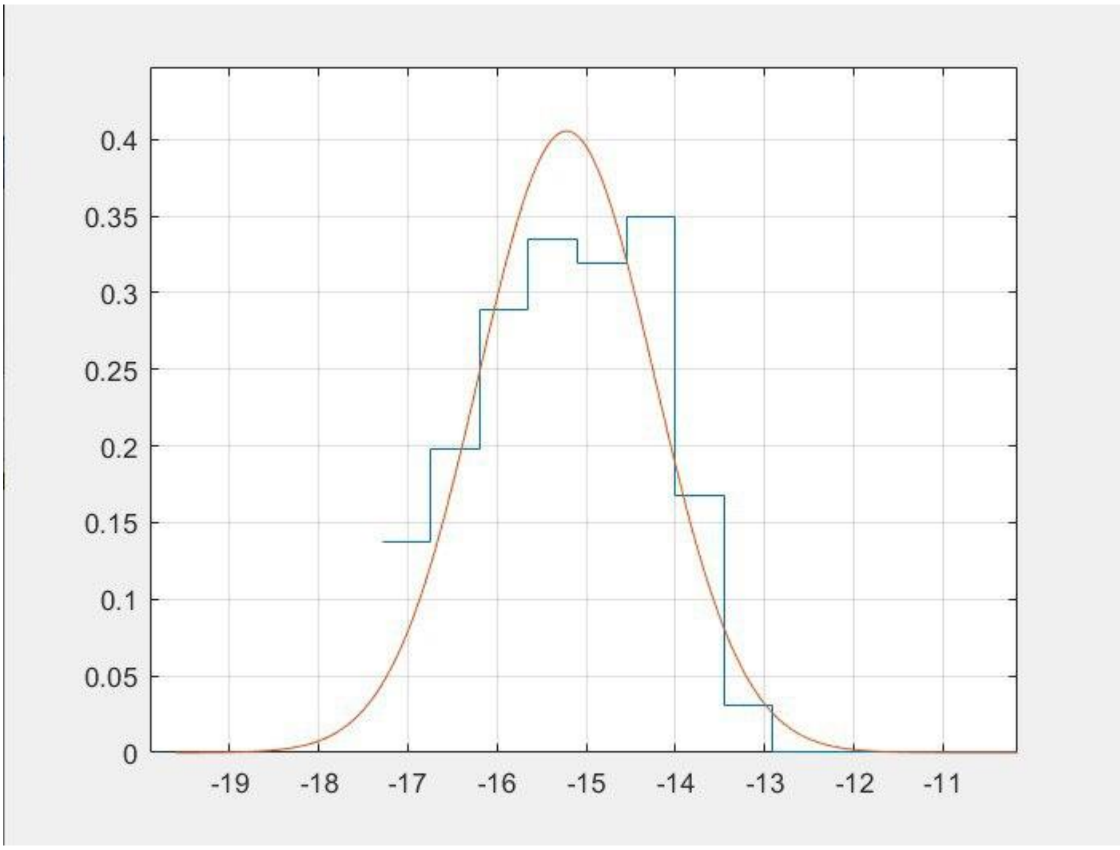

График 1. График эмпирической функции распределения и функции распределения нормальной случайной величины

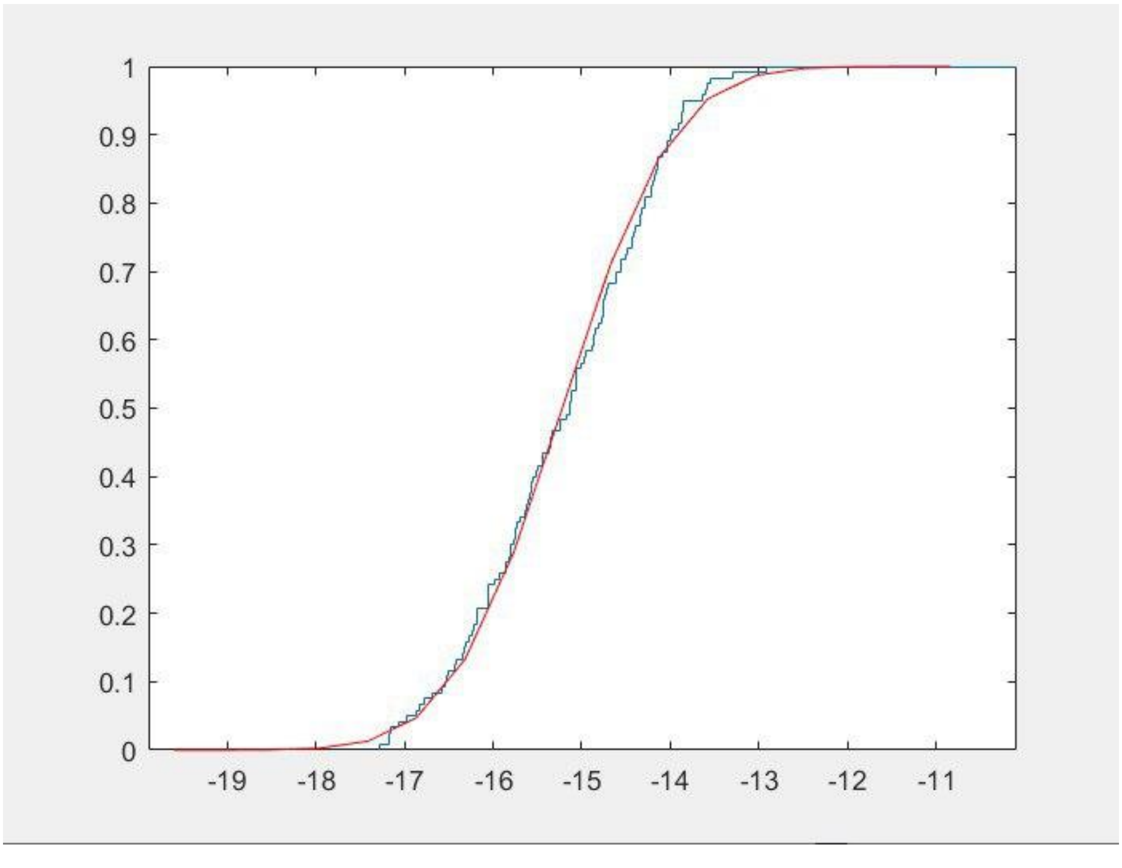

График 2. Гистограмма и график функции плотности распределения нормальной случайной величины