Московский государственный технический университет
имени Н. Э. Баумана

Факультет: Информатика и системы управления

Кафедра: Программное обеспечение ЭВМ и информационные технологии

Дисциплина: Математическая статистика

Лабораторная работа №1
Гистограмма и эмпирическая функция распределения

Выполнила: Покасова А.И.
Группа:ИУ7-61
Вариант: 16

Москва, 2019 г.

1 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
   - вычисление максимального значения $M_{\max}$ и минимального значения $M_{\min}$;
   - вычисление размаха R выборки;
   - вычисление оценок $\hat{\mu}$ и S^2 математического ожидания $\mathsf{M}X$ и дисперсии $\mathsf{D}X$;
   - группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
   - построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
   - 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.

2 Отчёт

2.1 Формулы для вычисления величин

Количество интервалов

$$m = [\log_2 n] + 2 \tag{1}$$

Минимальное значение выборки

$$M_{\min} = \min\{x_1, \ldots, x_n\}, \quad \text{где} \tag{2}$$

- $(x_1, \ldots, x_n)$ — реализация случайной выборки.

Максимальное значение выборки

$$M_{\max} = \max\{x_1, \ldots, x_n\}, \quad \text{где} \tag{3}$$

- $(x_1, \ldots, x_n)$ — реализация случайной выборки.

Размах выборки

$$R = M_{\max} - M_{\min}, \quad \text{где} \tag{4}$$

- $M_{\max}$ — максимальное значение выборки;
- $M_{\min}$ — минимальное значение выборки.

Оценка математического ожидания

$$\hat{\mu}(\vec{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{5}$$

Исправленная оценка дисперсии

$$S^2(\vec{X}) = \frac{n}{n-1}\hat{\sigma}^2(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\,. \tag{6}$$

2.2 Эмпирическая плотность и гистограмма

Определение 2.1. *Эмпирической плотностью распределения выборки* $\vec{x}$ называют функцию

$$f_n(x) = \begin{cases} \frac{n_i}{n\,\Delta}, & x \in J_i,\ i = \overline{1, m}; \\ 0, & \text{иначе.} \end{cases}, \quad \text{где} \tag{7}$$

- J_i, $i = \overline{1; m}$, — полуинтервал из $J = [x_{(1)}, x_{(n)}]$, где

  $$x_{(1)} = \min\{x_1, \ldots, x_n\}, \qquad x_{(n)} = \max\{x_1, \ldots, x_n\}; \tag{8}$$

  при этом все полуинтервалы, кроме последнего, не содержат правую границу т. е.

  $$J_i = [x_{(1)} + (i-1)\Delta, x_{(1)} + i\Delta), \quad i = \overline{1, m-1}; \tag{9}$$
  $$J_m = [x_{(1)} + (m-1)\Delta, x_{(1)} + m\Delta]; \tag{10}$$

- m — количество полуинтервалов интервала $J = [x_{(1)}, x_{(n)}]$;
- Δ — длина полуинтервала J_i, $i = \overline{1, m}$ равная

$$\Delta = \frac{x_{(n)} - x_{(1)}}{m} = \frac{|J|}{m}; \tag{11}$$

- n_i — количество элементов выборки в полуинтервале J_i, $i = \overline{1, m}$;
- n — количество элементов в выборке.

Определение 2.2. График функции $f_n(x)$ называют гистограммой.

2.3 Эмпирическая функция распределения

Определение 2.3. Эмпирической функцией распределения, отвечающей выборке $\vec{x}$ называют функцию

$$F_n \colon \mathbb{R} \to \mathbb{R}, \qquad F_n(x) = \frac{n(x, \vec{x})}{n}, \tag{12}$$

где $n(x, \vec{x})$ — количество элементов выборки $\vec{x}$, которые меньше x.

3 Листинг программы

```
function lab1()
    clear all;
    X = csvread("data.csv");

    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('R = %s\n', num2str(R));

    mu = sExp(X);
    fprintf('mu = %s\n', num2str(mu));

    sSqr = unDeviation(X);
    fprintf('S^2: %s\n', num2str(sSqr));

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

    intervals(X, m);
    hold on;
    f(X, mu, sSqr, m);

```

```
    figure;
    empiricF(X);
    hold on;
    F(X, mu, sSqr, m);

end

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

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

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

    mu = sum / n;
end

function sSqr = unDeviation(X)
    n   = length(X);
    mu  = sExp(X);
    sum = 0;

    for i = 1:n
        sum = sum + (X(i) - mu)^2;
    end

    sigmaSqr = sum / n;
    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);
    J(length(J)+1) = 18;

    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) (%.2f;)\t', J(j-1), J(j), count(j-1));
```

```
                end
            end
            count(j) = count(j) + 1;
        end
        fprintf('[%2.2f;%2.2f; %2.2f]\n', J(m), J(m + 1), count(m));

        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) = 18;
        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)
        X = [8, X];
        [yy, xx] = ecdf(X);
        yy(2) = 0;
        yy(length(yy)+1) = 1;
        xx(length(xx)+1) = 18;
        stairs(xx, yy);
end
```

4 Результаты расчётов

$$M_{\min} = 9.54;$$
$$M_{\max} = 15.21;$$
$$R = 5.67;$$
$$\hat{\mu}(\vec{X}_n) = 11.9532;$$
$$S^2(\vec{X}_n) = 1.1818.$$

Интервальная группировка значений выборки при $m = 8$:

$$[9.54; 10.25), 5, [10.25; 10.96), 17, [10.96; 11.67), 28, [11.67; 12.38), 24, [12.38; 13.08), 28,$$

$$[13.08; 13.79), 12, [13.79; 14.50), 5, [14.50; 15.21], 1.$$

5 Графики

5.1 Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

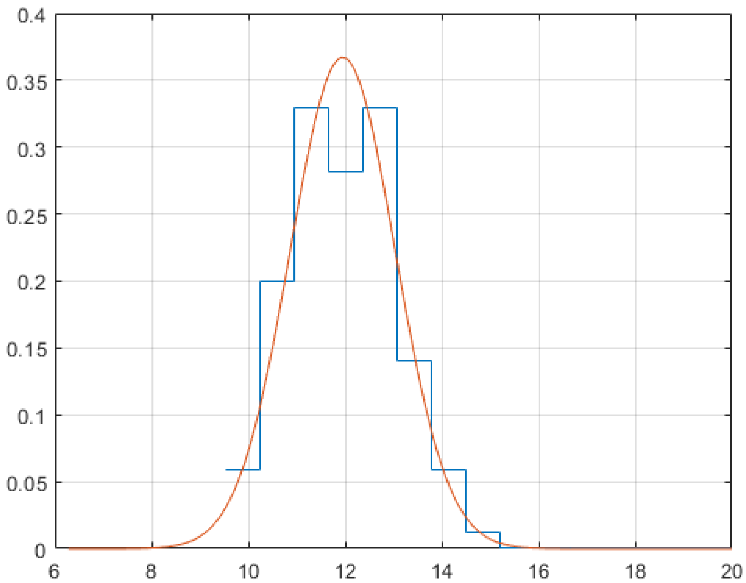

Рис. 1: Гистограмма и график функции плотности распределения нормальной случайной величины.

5.2 График эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

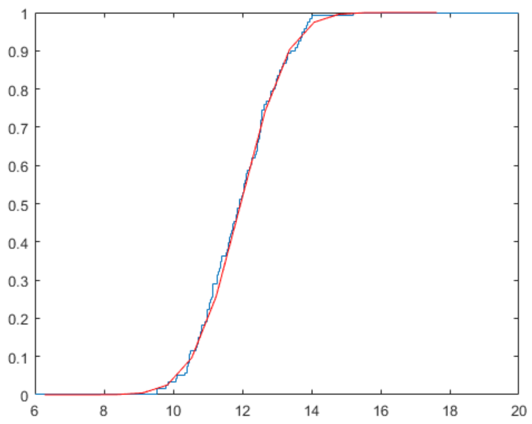

Рис. 2: График эмпирической функции распределения и функции распределения нормальной случайной величины.