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 предмету: «Математическая статистика»

Тема: Гистограмма и эмпирическая функция распределения

Вариант 25

Студент: Юмаев Артур Русланович
Группа: ИУ7-65Б

Москва, 2020 г.

Цель и содержание работы

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
   a. вычисление максимального значения M_{max} и минимального значения M_{min};
   b. размаха R выборки;
   c. вычисление μ 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 μ и дисперсией S^2;
   f. построение на другой коордитной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием μ и дисперсией S^2;
2. Провести вычисления и построить графики для выборки из индивидуального варианта.

Формулы для вычисления величин

Минимальное значение выборки:

$M_{min} = \min\{x_1, \dots, x_n\}$, где $(x_1, \dots, x_n)$ – реализация случайной выборки.

Максимальное значение выборки:

$M_{max} = \max\{x_1, \dots, x_n\}$, где $(x_1, \dots, x_n)$ – реализация случайной выборки.

Размах выборки:

$R = M_{max} - M_{min}$

Выборочным средним (выборочным математическим ожиданием) называется статистика:

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

Несмещенная оценка дисперсии:

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

Эмпирическая плотность и гистограмма

$m = [\log_2 n] + 2,$

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

$x_{(n)} = \max\{x_1, \dots, x_n\}$

При больших объемах выборки n $(n > 50)$ обычно производят группирование исходных данных следующим образом. Промежуток $J = [x_{(1)}, x_{(n)}]$, содержащий все выборочные значения, разбивают на m полуинтервалов $J_1, \dots, J_m$, как правило, одинаковой длины Δ и таких, что каждый из них, кроме последнего, содержит левую границу, а последний содержит обе границы, и подсчитывают число n_i элементов выборки, попавших в i-ый промежуток $J_i, i = \overline{1, m}, n = n_1 + \cdots + n_m$, а результаты представляют в виде слудующей таблицы 1, которую называют интервальным статистическим рядом.

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

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

$m -$ количество полуинтервалов отрезка $J = [x_{(1)}, x_{(n)}]$

$\Delta -$ длина полуинтервала $J_i, i = \overline{1,m}$ равная

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

$n_i -$ количество элементов выборки в полуинтервале $J_i, i = \overline{1,m}$,

$n -$ количество элементов в выборке.

Таблица 1. Интервальный статистический ряд

J_1	J_2	…	J_m	
n_1	n_2	…	n_m	$\sum_{i=1}^{m} n_i$

Определение. График функции $p_n(x) = \begin{cases} \frac{n_i}{n\Delta}, & x \in J_i \\ 0, & x \notin J \end{cases}$, представляющий собой кусочно постоянную функцию называют ***гистограммой*** (рис. 1).

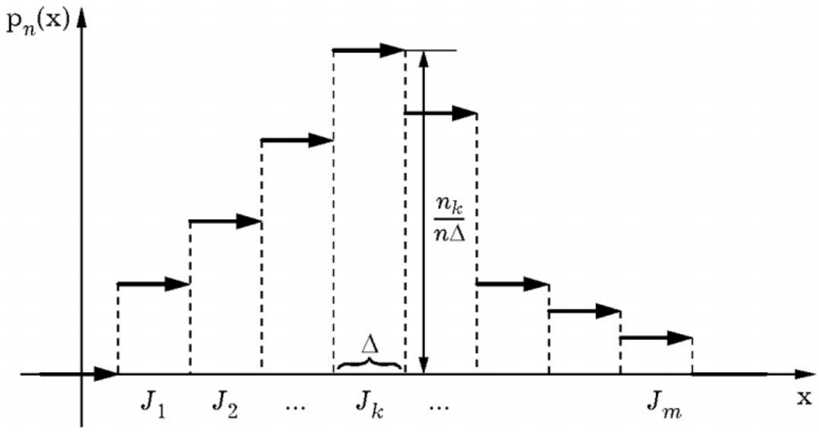

Рисунок 1. Гистограмма

Часто гистограммой называют диаграмму, составленную из прямоугольников с основанием Δ и высотами $n_i/(n\Delta)$, $i = \overline{1,m}$. Нетрудно увидеть, что суммарная площадь всех прямоугольников, образующих такую диаграмму, равна 1, так как:

$$\sum_{i=1}^{m} \frac{n_i}{n\Delta} \Delta = \frac{1}{n} \sum_{i=1}^{m} n_i = 1.$$

Площадь каждого прямоугольника n_i/n есть частота попадания элементов выборки в соответсвующий интервал J_i статистического ряда.

Пусть

1. $\overrightarrow{X_n} = (X_1, \dots, X_n) -$ случайная выборка,
2. $\overrightarrow{x_n} = (x_1, \dots, x_n) -$ реализация случайной выборки,
3. $n(x, \overrightarrow{x_n}) -$ количество элементов выборки $\overrightarrow{x_n}$, которые имеют значения меньше *x*.

Определение. Эмпирической функцией распределения называют функцию

$F_n: R \to R$, определенную условием $F_n(x) = \frac{n(x,\overrightarrow{x_n})}{n}$.

Листинг

Листинг 1. Программа на Matlab к лабораторной работе №1

```
function lab1
    clear
    X=[-17.04,-18.29,-17.38,-18.11,-18.96,-17.65,-17.02,-17.22,-16.25,-17.44,-17.69,-
17.61,-17.09,-17.19,-16.02,-17.56,-16.94,-17.29,-16.93,-16.61,-19.38,-17.53,-16.39,-
17.89,-17.98,-17.04,-16.22,-19.09,-18.91,-17.77,-18.30,-17.44,-18.84,-16.39,-16.13,-
18.37,-16.37,-16.70,-17.78,-17.03,-17.76,-17.87,-17.20,-18.44,-17.19,-17.75,-16.81,-
17.97,-18.03,-16.87,-16.10,-19.16,-16.51,-18.39,-16.48,-18.08,-17.49,-18.89,-19.09,-
17.96,-18.40,-16.96,-18.15,-18.71,-17.81,-17.86,-19.47,-17.86,-17.60,-17.30,-17.60,-
17.71,-18.42,-16.88,-16.76,-18.00,-17.97,-16.83,-18.00,-18.08,-17.61,-17.02,-16.73,-
17.64,-18.76,-17.68,-18.04,-16.45,-18.79,-18.03,-17.38,-15.27,-15.97,-17.41,-18.61,-
18.00,-17.42,-17.77,-19.05,-16.16,-16.27,-18.00,-18.90,-17.05,-17.46,-17.49,-18.20,-
17.59,-15.78,-18.88,-18.53,-17.39,-17.83,-18.17,-16.15,-17.66,-17.76,-18.32,-17.70,-
17.56];

    % Пункт а)
    Mmax = max(X);
    Mmin = min(X);

    % Пункт б)
    R = Mmax - Mmin;

    % Пункт в)
    mu = find_mu(X);
    S2 = find_S2(X);

    % Пункт г)
    % Гистограмма
    m = find_m(X);
    [J, count] = intervalization(X, m);
    stairs([J(1) J], [0 count 0]);
    % График 1
    hold on;
    Y1 = f(X, mu, S2);
    plot(X, Y1, '.');
    grid on;
    legend('Гистограмма','Функция плотности распределения нормальной СВ');
    hold off;
    % График 2
    figure;
    empF(sort(X));
    hold on;
    F(sort(X), mu, getS(X), m, R);
    grid on;
    legend('Эмпирическая функция распределения','Функция распределения нормальной СВ');
    hold off;

    function [mu] = find_mu(X)
        mu = sum(X)/size(X,2);
    end

    function [S2] = find_S2(X)
        S2 = sum((X - find_mu(X)) .* (X - find_mu(X)))/ (size(X,2) - 1);
    end

    function sigma = getSigmaSqr(X)
        tempMu = find_mu(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] = find_m(X)
        m = floor(log2(size(X, 2))) + 2;
    end

    function [J, count] = intervalization(X, m)
        sortX = sort(X);
        n = size(sortX,2);
        delta = (sortX(end) - sortX(1)) / m;
        J = sortX(1):delta:sortX(end);
        count = zeros(1,m);

        for i = 1:(size(J,2)-1)
            for j = 1:n
                if (sortX(j) >= J(i) && sortX(j) < J(i+1))
                    count(i) = count(i) + 1;
                end
            end
        end

        count(end) = count(end) + sum(sortX==(sortX(end)));

        for i = 1:size(count,2)
            count(i) = count(i)/(n * delta);
        end

    end

    function [Y] = f(X, MX, DX)
        Y = normpdf(X, MX, sqrt(DX));
    end

    function empF(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
```

Результат работы программы

$M_{max} = -15.27$

$M_{min} = -19.47$

$R = 4.2$

$\hat{\mu} = -17.5894$

$S^2 = 0.7286$

Таблица 1. Интервальная группировка значений выборки при m = 8

[-19.470, -18.945)	[-18.945, -18.419)	[-18.410, -17.895)	[-17.895, -17.369)	[17.369, -16.845)	[-16.845, -16.320)	[-16.320, -15.795)	[15.795, -15.270]
7	12	23	37	18	12	9	2

Графики

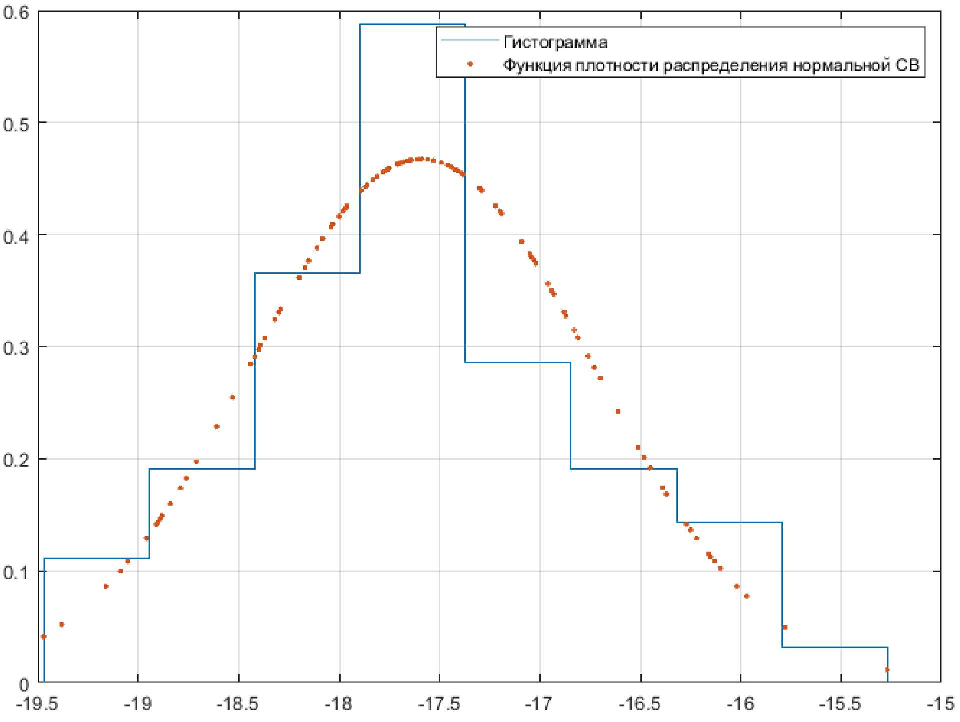

Рисунок 2. Гистограмма и график функции плотности распределения нормальной случайной

Рисунок 3. График эмпирической функции распределения и функции распределения нормального распределения