Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Московский государственный технический
университет имени Н.Э.Баумана (национальный
исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

Дисциплина:
«Математическая статистика»

Лабораторная работа №1.
Гистограмма и эмпирическая функция распределения
Вариант 15

Студент: Рязанов М.С.
Группа: ИУ7-62Б
Преподаватель: Власов П. А.

Москва, 2020 г.

Содержание

1 Задание

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
  - а)вычисление максимального значения M_{max} и минимального значения M_{min};
  - б)размаха R выборки;
  - в)вычисление оценок $\hat{\mu}$ и S_2 математического ожидания MX и дисперсии DX;
  - г)группировку значений выборки в $m = [log_2 n] + 2$ интервала;
  - д)построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
  - е)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 Формулы для вычисления

Далее приведены формулы для вычисления статистик случайной выборки $\overrightarrow{X_n}$ объема n из генеральной совокупности X.

Максимальное значение:

$$M_{max} = max\{X_1, X_2, .., X_n\} \tag{1}$$

Минимальное значение:

$$M_{min} = min\{X_1, X_2, .., X_n\} \tag{2}$$

Размах выборки:

$$R = M_{max} - M_{min} \tag{3}$$

Оценка математического ожидания:

$$\hat{\mu}(\overrightarrow{X_n}) = \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{4}$$

Оценка дисперсии(несмещенная выборочная дисперсия)

$$S^2(\overrightarrow{X_n}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 \tag{5}$$

3 Определение эмпирической плотности и гистограммы

Пусть $\overrightarrow{x}$ - выборка из генеральной совокупности X, значения x_i богут быть сгруппированы в **интервальный статистический ряд**. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ делят на m равновеликих частей(в данной лабораторной работе $m = [log_2 n] + 2$):

$$J_i = [a_{i-1}, a_i), i = \overline{1; m-1}$$

$$J_m = [a_{m-1}, a_m]$$

где $a_i = x_{(1)} + i\Delta, i = \overline{0; m}$, $\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}$

Интервальным статистическим рядом называется таблица:

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

где n_i - количество элементов выборки $\overrightarrow{x}$, которые $\in J_i$

Пусть для выборки $\overrightarrow{x}$ построен интервальный статистический ряд

$$(J_i, n_i), i = \overline{1, m}$$

Тогда эмпирической плотностью(отвечающей выборке $\overrightarrow{x}$) называется функция

$$\hat{f}_n = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1, m} \\ 0, \textit{иначе} \end{cases}$$

Гистограммой называется график эмпирической плотности.

4 Определение эмпирической функции распределения

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

Эмпирической функцией распределения называют функцию

$$F_n : \mathbb{R} \to \mathbb{R}$$

определенную условием

$$F_n(x) = \frac{n(x, \overrightarrow{x})}{n}$$

5 Текст программы

```
function lab1()
    % Выборка
    X = [-2.79,-3.01,-4.07,-2.85,-2.43,-3.20,-3.72,
    -4.27,-5.48,-2.38,-4.69,-4.34,-5.08,-5.01,-4.08,-4.20,-4.74,-1.88,-3.25,
    -2.78,-3.56,-3.54,-3.79,-3.18,-5.08,-4.30,-2.86,
    -2.45,-3.08,-3.22,-2.76,-3.20,-3.33,-4.91,-4.06,-3.81,-3.96,-3.65,-3.77,
    -4.60,-5.21,-2.67,-1.95,-2.43,-1.73,-2.50,-3.96,
    -3.75,-2.70,-4.26,-3.42,-4.07,-4.74,-3.00,-4.37,-5.42,-5.00,-4.08,-2.46,
    -4.33,-4.08,-3.72,-4.09,-2.96,-3.71,-1.51,-3.70,
    -6.48,-4.26,-4.39,-3.16,-4.63,-2.66,-2.22,-4.79,-2.46,-3.69,-3.35,-2.32,
    -4.17,-3.85,-4.93,-2.05,-3.15,-3.49,-5.70,-2.53,
    -3.85,-4.32,-3.37,-3.98,-3.74,-5.28,-2.56,-3.21,-3.10,-3.78,-3.36,-3.32,
    -2.59,-2.45,-3.34,-3.20,-4.14,-4.00,-4.79,-4.02,
    -4.58,-4.45,-3.69,-4.53,-3.98,-4.51,-4.44,-3.78,
    -4.24,-4.00,-2.46,-2.58,-4.04];

    n = length(X); % Объем выборки

    % Подсчет статистик
    M_max = max(X)
    M_min = min(X)

    R = M_max - M_min % Размах выборки

    mu = mean(X) % Оценка матожидания
    s = std(X, 0)
    s_sqr = s^2 % Исправленная выборочная дисперсия

    var_series = sort(X); % Вариационный ряд
    stat_series = [var_series(1); 1]; % Стат. ряд
    % Подсчет статистического ряда
    % 1 - строка матрицы - уникальные значения
    % 2 - строка матрицы - число вхождений
    uniqs = 1; % Число уникальных

    for i = 2:n
        if var_series(i) == var_series(i-1)
            stat_series(2, uniqs) = stat_series(2, uniqs) + 1;
        else
            uniqs = uniqs + 1;
            stat_series(:, uniqs) = [var_series(i); 1];
        end
    end

    % Группировка значений выборки
    m = floor(log2(n)) + 2

    % Построение интервального статистического ряда
    delta = (M_max - M_min) / m;
    intervals = M_min:delta:M_max % Интервалы
    counts = histc(var_series, intervals); % Число элементов в интервале

    % Обработка случая J_m = [x_p, x_p+1]
```

```
    counts(end-1) = sum(counts(end-1:end));
    counts(end) = 0

    % Гистограмма и график функции распределения
    counts = counts / (delta * n); % Значение эмпирической функци на интервалах;
    [sx, sy] = stairs(intervals, counts);

    % Функция плотности распределения N(mu,s_sqr)
    x = linspace(M_min, M_max);
    y = normpdf(x, mu, s);

    figure(1);
    plot(sx, sy, x,y);
    xlabel('x');
    ylabel('f');
    legend({'Эмпирическая функция плотности', 'Функция плотности нормальной
   слвеличины.'}, 'Location','northwest');

    % Эмпирическая функция распределения и функция распределения N(mu, s_sqr)
    % Вычисление эмпирической функции распределения
    F = [];
    for i=1:uniqs
        F(i) = sum(stat_series(2, 1:i-1));
    end
    F = F / n;

    % Функция распределения нормальной случайной величины
    y = normcdf(x, mu, s);

    % Построение графиков
    [sx, sy] = stairs(stat_series(1, :), F);

    figure(2);
    plot(sx, sy, x, y);
    xlabel('x');
    ylabel('F');
    legend({'Эмпирическая функция распределения','Функция распределения нормальной
   сл. величины'},'Location','northwest');
end
```

6 Результат расчетов для выборки из индивидуального варианта

В формулах 6, 7, 8, 9, 10 приведены результаты расчетов для выборки из индивидуального варианта.

$$M_{max} = -1.51 \tag{6}$$

$$M_{min} = -6.48 \tag{7}$$

$$R = 4.97 \tag{8}$$

$$\hat{\mu} = -3.6762 \tag{9}$$

$$S^2 = 0.8664 \tag{10}$$

Интервалы в которые были сгруппированы значения выборки

- $J_1 - [-6.48; -5.86)$
- $J_2 - [-5.86; -5.24)$
- $J_3 - [-5.24; -4.62)$
- $J_4 - [-4.62; -3.99)$
- $J_5 - [-3.99; -3.37)$
- $J_6 - [-3.37; -2.75)$
- $J_7 - [-2.75; -2.13)$
- $J_8 - [-2.13; -1.51]$

Интервально статический ряд:

J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8
1	4	13	30	24	25	18	5

На рисунке 1 изображена гистограмма и график функции плотности распределения для нормальной случайной величины с параметрами $\hat{\mu}$ и S^2.

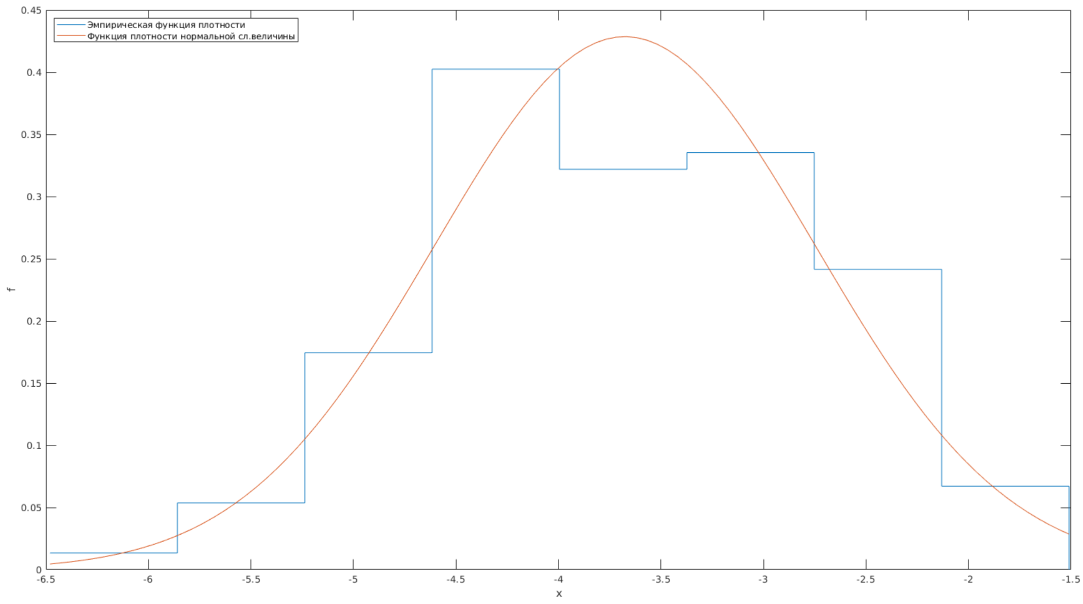

Рисунок 1: Гистограмма и график функции плотности

На рисунке 2 изображены графики эмпирической функции распределения и функции распределения нормальной случайной величины с параметрами $\hat{\mu}$ и S^2

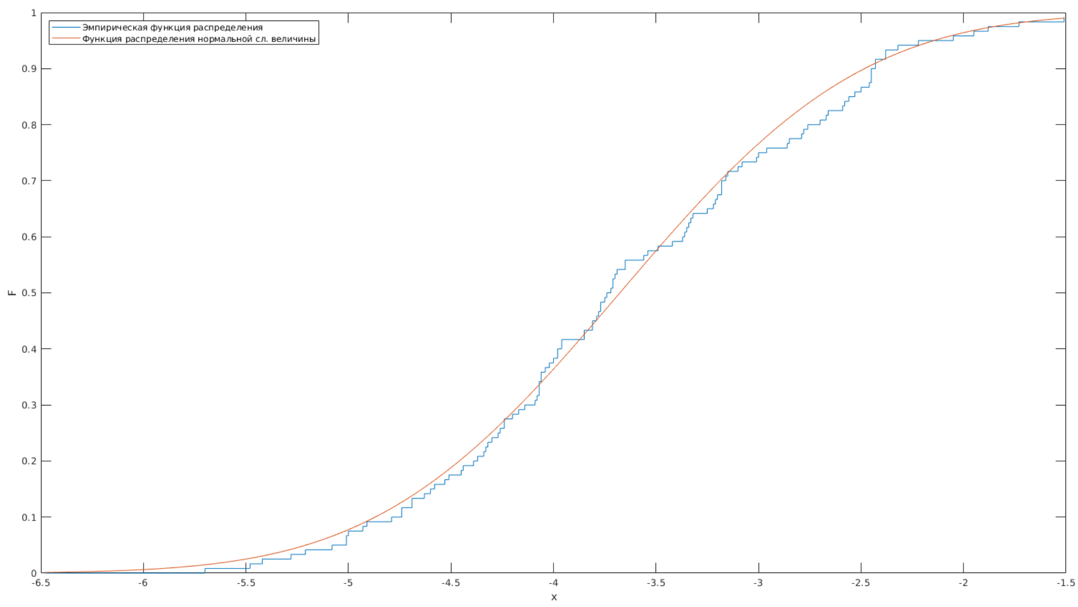

Рисунок 2: Функции распределения