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

Дисциплина Математическая статистика

Тема Гистограмма и эмпирическая функция распределения

Студент Набиев Ф.М.

Группа ИУ7–63Б

Преподаватель Власов П.А.

Москва, 2020 г.

1 Теоретическая часть

Пусть $\overrightarrow{X} = \big(x_1, \ldots, x_n\big)$ — выборка из генеральной совокупности X.

Тогда:

1° Максимальное значение выборки:

$$M_{\max} = \max\{x_1, \ldots, x_n\}$$

2° Минимальное значение выборки:

$$M_{\min} = \min\{x_1, \ldots, x_n\}$$

3° Размах выборки:

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

4° Выборочное среднее:

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

5° Состоятельная оценка дисперсии:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

■

Пусть $\overrightarrow{X}$ — выборка из генеральной совокупности X,

$$\overrightarrow{X} = (x_1, \ldots, x_n)$$

Расположим значения $x_1, \ldots, x_n$ в порядке неубывания

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

где $x_{(i)}$ — i-й элемент полученной последовательности. Такая последовательность называется вариационный ряд.

Если объем n статистической выборки $\vec{X}$ велик ($n \geq 50$), то можно сгруппировать выборку в интервальный статистический ряд. Для этого отрезок $J = \left[x_{(1)}, x_{(n)}\right]$ делят на $p = [\log_2 n] + 1$ (где $[a]$ — целая часть числа a) равновеликих частей:

$$J_i = \left[a_{i-1}, a_i\right), i = \overline{1, p-1}$$

$$J_p = \left[a_{p-1}, a_p\right]$$

где

$$a_i = x_{(1)} + i\Delta, i = \overline{0, p}$$

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

Интервальным статистическим рядом называют таблицу

J_1	$\dots$	J_i	$\dots$	J_p
n_1	$\dots$	n_i	$\dots$	n_p

где n_i — количество элементов $\overrightarrow{X}$, которые $\in J_i$.

Предположим, что для выборки $\overrightarrow{X}$ построен интервальный статистический ряд

$$\left(J_i, n_i\right), i = \overline{1; p}$$

Эмпирической плотностью (отвечающей выборке $\overrightarrow{X}$) называют функцию

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

Гистограммой называют график эмпирической плотности.

■

Для выборки $\overrightarrow{X}$ 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

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

■

2 Практическая часть

В листинге 2.1 приведён текст программы.

Листинг 2.1 – Текст программы

```
function lab1()
    X = [
        -13.40, -12.63, -13.65, -14.23, -13.39, -12.36, ...
        -13.52, -13.44, -13.87, -11.82, -12.01, -11.40, ...
        -13.02, -12.61, -13.06, -13.75, -13.55, -14.01, ...
        -11.75, -12.95, -12.59, -13.60, -12.76, -11.05, ...
        -13.15, -13.61, -11.73, -13.00, -12.66, -12.67, ...
        -12.60, -12.47, -13.52, -12.61, -11.93, -13.11, ...
        -13.22, -11.87, -13.44, -12.70, -11.78, -12.30, ...
        -12.89, -13.29, -12.48, -10.44, -12.55, -12.64, ...
        -12.03, -14.60, -14.56, -13.30, -11.32, -12.24, ...
        -11.17, -12.50, -13.25, -12.55, -12.85, -12.67, ...
        -12.41, -12.58, -12.10, -13.54, -12.69, -12.87, ...
        -12.71, -12.77, -13.30, -12.74, -12.73, -12.64, ...
        -12.18, -11.20, -12.40, -13.78, -13.71, -10.74, ...
        -11.89, -13.20, -11.31, -14.26, -10.38, -12.88, ...
        -11.39, -11.35, -12.55, -12.84, -10.25, -12.40, ...
        -14.01, -11.47, -13.14, -12.69, -11.92, -12.86, ...
        -13.06, -12.57, -13.63, -12.34, -12.84, -14.03, ...
        -13.34, -11.64, -13.58, -10.44, -11.37, -11.01, ...
        -13.80, -13.27, -12.32, -10.69, -12.92, -13.29, ...
        -12.58, -13.98, -11.46, -11.82, -12.33, -11.47, ...
    ];
    X = sort(X);

    % Максимальное значение выборки
    M_max = max(X);
    % Минимальное значение выборки
    M_min = min(X);

    % Разброс выборки
    R = M_max - M_min;

    % Оценка мат. ожидания
    M = mean(X);
    % Оценка дисперсии
    D = var(X);

    fprintf('M_max    = %9.5f\n', M_max);
    fprintf('M_min    = %9.5f\n', M_min);
    fprintf('R        = %9.5f\n', R);
    fprintf('mean     = %9.5f\n', M);
    fprintf('varience = %9.5f\n', D);
    fprintf('\n');
```

```

    m = floor(log2(length(X))) + 2;
    % Группировка значений в m интервалов
    [N, edges] = histcounts(X, m, 'BinLimits', [M_min, M_max]);

    % Вывод всех интервалов, кроме последнего,
    % потому что последний - это отрезок,
    % а все остальные - полуинтервалы
    fprintf('%d intervals:\n', m);
    for i = 1 : (length(N) - 1)
        fprintf('%3d values in [%f,%f)\n', ...
              N(i), edges(i), edges(i + 1));
    end
    % Вывод последнего отрезка
    fprintf('%3d values in [%f,%f]\n', ...
          N(end), edges(end - 1), edges(end));

    f = normpdf(X, M, sqrt(D));
    F = normcdf(X, M, sqrt(D));

    subplot(2, 1, 1);
    % Построение гистограммы
    histogram(X, m, 'Normalization', 'pdf', ...
            'BinLimits', [M_min, M_max]);
    hold on;
    % Построение графика функции плотности распределения
    plot(X, f, 'LineWidth', 2);
    hold off;

    subplot(2, 1, 2);
    % Построение графика эмпирической функции распределения
    [YY, XX] = ecdf(X);
    stairs(XX, YY, 'LineWidth', 2);
    hold on;
    % Построение графика функции распределения
    plot(X, F, 'LineWidth', 2);
    hold off;
end
```

В листинге 2.2 приведён результат выполнения описанной программы.

Листинг 2.2 – Результат программы

```
M_max    = -10.25000
M_min    = -14.60000
R        =   4.35000
mean     = -12.61483
```

```
varience =    0.86533

8 intervals:
  4 values in [-14.600000,-14.056250)
 18 values in [-14.056250,-13.512500)
 20 values in [-13.512500,-12.968750)
 36 values in [-12.968750,-12.425000)
 16 values in [-12.425000,-11.881250)
 14 values in [-11.881250,-11.337500)
  6 values in [-11.337500,-10.793750)
  6 values in [-10.793750,-10.250000]
```

Рис. 2.1 – Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2

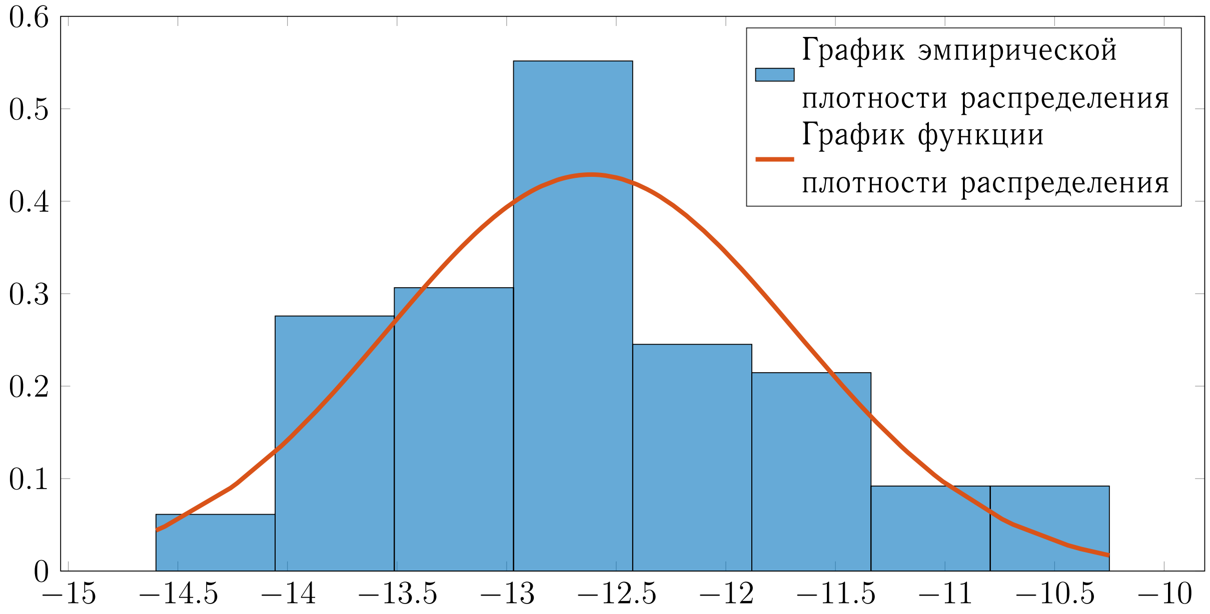

Рис. 2.2 – Графики эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2