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»

Отчёт по лабораторной работе №1 по курсу «Математическая статистика»

Тема Гистограмма и эмпирическая функция распределения

Студент Богаченко А. Е.

Группа ИУ7-65Б

Оценка (баллы) ____________________

Преподаватели Андреева Т. В.

Москва — 2022 г.

Оглавление

1 Формулы для вычисления величин

$$\vec{x} = (x_1, ..., x_n)$$

1. **Максимальное значение выборки**

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

2. **Минимальное значение выборки**

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

3. **Размах выборки**

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

4. **Выборочное среднее (математическое ожидание)**

$$\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

5. **Состоятельная оценка дисперсии**

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

где $\overline{x} = \hat{\mu}$

2 Определение эмпирической плотности и гистограммы

Эмпирической плотностью (отвечающей выборке $\vec{x}$) называют функцию

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

где (J_i, n_i) – интервальный статистический ряд

Пусть $\vec{x}$ – выборка из генеральной совокупности X. Если объем n этой выборки велик, то значения x_i группируют не только в статистический ряд, но и в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ (где $x_{(1)} = \min\{x_1, .., x_n\}$, $x_{(n)} = \max\{x_1, .., x_n\}$) делят на p равновеликих частей:

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

$$J_p = [a_p, a_{p+1}]$$

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

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

Интервальным статистическим рядом называют таблицу

J_1	...	J_i	...	J_p
n_1	...	n_i	...	n_p

Здесь n_i – количество элементов выборки $\vec{x}$, которые $\in J_i$

В нашем случае $p = m = [\log_2 n] + 2$

Гистограммой называют график эмпирической плотности.

3 Определение эмпирической функции распределения

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

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную условием $F_n(x) = \frac{n(x,\vec{x})}{n}$.

4 Текст программы

```
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
];

% а) вычисление Mmax и Mmin
% sort row
sort_row = sort(X);  % вариац. ряд
% size of vector
n = length(sort_row);

% calculate Mmax, Mmin, R, MX, DX
M_max = max(X);
M_min = min(X);

Mmax = sort_row(n);
Mmin = sort_row(1);

fprintf('a) (stand) M_max = %.2f, M_min = %.2f\n', M_max, M_min);
fprintf('a) Mmax = %.2f, Mmin = %.2f\n\n', Mmax, Mmin);

% б) размаха R выборки
R = M_max - M_min;
fprintf('б) R = %.2f\n\n', R);

% в) вычисление оценок μ, S^2 математического ожидания MX и дисперсии DX
MX = mean(X); % среднее значение
DX = var(X); % Дисперсия

MX1 = sum(X) / n;
S2 = sum((X - MX).^2) / (n - 1);
fprintf('в) (stand) mu = %.2f, S^2 = %.2f\n', MX, DX);
fprintf('в) mu = %.2f, S^2 = %.2f\n\n', MX1, S2);

% г) Группировка значений выборки в m = [log2 n] + 2 интервала;
```

```
% Построить интервальный ряд
[count, edges, m] = groupInterval(X);

% д) построение на одной координатной плоскости гистограммы и графика функции
% плотности распределения вероятностей нормальной случайной величины с математическим
% ожиданием μ и дисперсией S

plotHistogram(X, count, edges, m);
% Построение на одной координатной плоскости
hold on;
% График функции плотности распределения вероятностей нормальной
% случайной величины
fn = @(X, MX, S2) normpdf(X, MX, S2);
plotGraph(fn, MX, S2, Mmin, Mmax, 0.1);

% е) построение на другой координатной плоскости графика эмпирической функции
% распределения и функции распределения нормальной случайной величины
% с математическим ожиданием μ и дисперсией S

% Новая координатная плоскость
figure;
% график эмпирической функции распределения
plotEmpiricalF(X);
% Построение на одной координатной плоскости
hold on;
% График функции плотности распределения вероятностей нормальной случайной величины
Fn = @(X, MX, S2) normcdf(X, MX, S2);
plotGraph(Fn, MX, S2, Mmin, Mmax, 0.1);

 % Функция для группировки значений выборки
function [count, edges, m] = groupInterval(X)
    % Нахождение количества интервалов
    m = floor(log2(length(X))) + 2;
    % С помощью функции histcounts разбиваем выборку на m интервалов от
    % минимума до максимума . Возвращаем интервалы и количество элементов
    % в каждом из них
    [count, edges] = histcounts(X, m, 'BinLimits', [min(X), max(X)]);

    lenC = length(count);

    %  Вывод интервалов и количества элементов
    fprintf('\nИнтервальный ряд для m = %d \n', m);
    for i = 1 : (lenC - 1)
        fprintf('[%f,%f) - %d\n', edges(i), edges(i + 1), count(i));
    end
    fprintf('[%f,%f] - %d\n', edges(lenC), edges(lenC + 1), count(lenC));
end

% функция для отрисовки гистограммы
function plotHistogram(X, count, edges, m)
```

```
    % построение гистограммы
    h = histogram();
    % задаем интервалы
    h.BinEdges = edges;
    % задаем значения в каждом интервале (эмпирическую плотность)
    h.BinCounts = count / length(X) / ((max(X) - min(X)) / m);
    h.LineWidth = 2;
    h.DisplayStyle = 'stairs';
end

% Функция для отрисовки графиков func, с математическим ожиданием mu
% и дисперсией s2, от min до max с шагом step
function plotGraph(func, mu, s2, min, max, step)
    x = min : step : max;
        % нормальная ф-я плотности вероятности, возвращает PDF (Normal probability
        ↪  density function)
    % со средним μ, и стандартным отклонением s2.
        y = func(x, mu, s2);
    plot(x, y, 'LineWidth', 2);
end

% график эмпирической функции распределения
function plotEmpiricalF(X)
    % поиск уникальных элементов
    u = unique(X);
    % подсчет кол-ва каждого из уникальных эл-тов
    count = histcounts(X, u);
    % подсчет кол-ва эл-тов, меньших текущего уникального эл-та
    for i = 2 : (length(count))
        count(i) = count(i) + count(i - 1);
    end
    count = [0 count];
    % отрисовка графика
    stairs(u, count / length(X), 'LineWidth', 2);
end

```

5 Результаты расчетов для выборки (вариант 17)

1. Максимальное значение выборки
2. Минимальное значение выборки
3. Размах выборки
4. Выборочное среднее (математическое ожидание)
5. Состоятельная оценка дисперсии
6. Группировка значений выборки в $m = [log_2 n] + 2$ интервала

```
а) (stand) M_max = -10.25, M_min = -14.60
а) Mmax = -10.25, Mmin = -14.60
б) R = 4.35
в) (stand) mu = -12.61, S^2 = 0.87
в) mu = -12.61, S^2 = 0.87
Интервальный ряд для m = 8
[-14.600000,-14.056250) - 4
[-14.056250,-13.512500) - 18
[-13.512500,-12.968750) - 20
[-12.968750,-12.425000) - 36
[-12.425000,-11.881250) - 16
[-11.881250,-11.337500) - 14
[-11.337500,-10.793750) - 6
[-10.793750,-10.250000] - 6
```

Рисунок 5.1 – Группировка значений

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

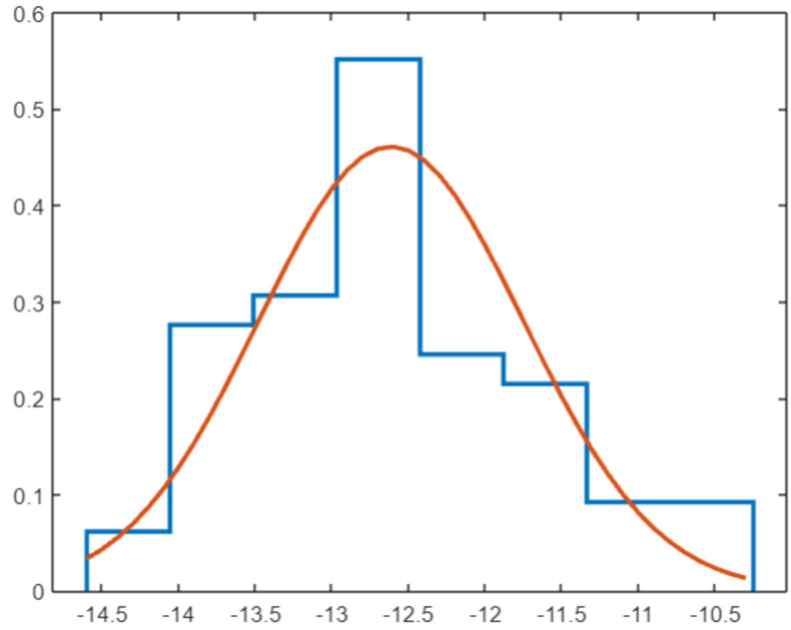

Рисунок 5.2 – Гистограмма

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

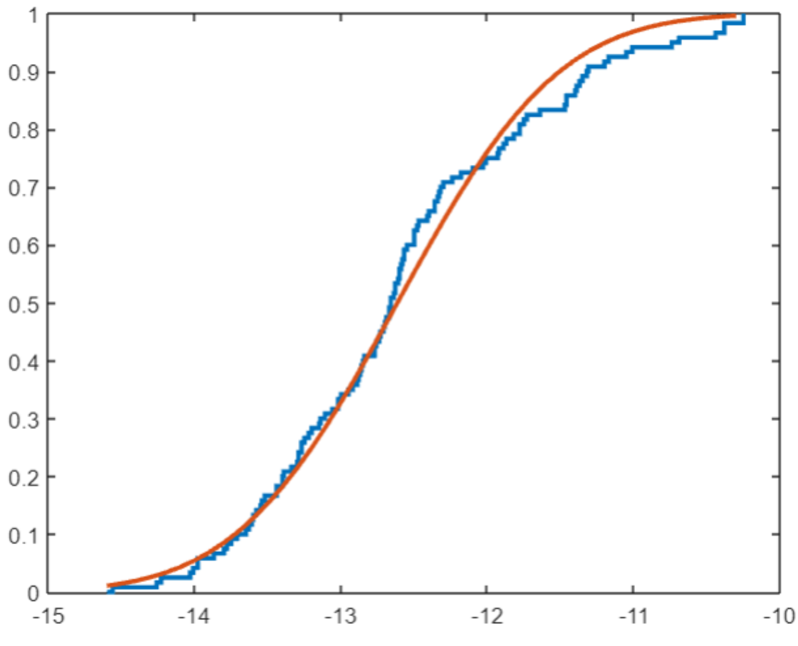

Рисунок 5.3 – Функции распределения