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»

НАПРАВЛЕНИЕ ПОДГОТОВКИ «09.03.04 Программная инженерия»

# ОТЧЕТ
# ПО ЛАБОРАТОРНОЙ РАБОТЕ №1

Название: Гистограмма и эмпирическая функция плотности

Дисциплина: Математическая статистика

| | | | |
|---|---|---|---|
| Студент | ИУ7-64Б | | С. Д. Параскун |
| | Группа | Подпись, дата | И. О. Фамилия |
| Преподаватель | | | М. А. Велищанский |
| | | Подпись, дата | И. О. Фамилия |

Москва, 2022 г.

# 1. Задание на лабораторную

***Цель работы:*** построение гистограммы и эмпирической функции распределения.

***Содержание отчета:***

1. Для выборки объема n из генеральной совокупности X реализовать в виде программы на ЭВМ:
   - вычисление максимального значения $M_{max}$ и минимального значения $M_{min}$;
   - размаха $R$ выборки;
   - вычисление оценок $\hat{\mu}$ и $S^2$ математического ожидания MX и дисперсии DX;
   - 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.

***В работе использовалась выборка по 13 варианту.***

# 2. Вычисление параметров

Минимальное и максимальное значение выборки:

$$\begin{aligned} M_{min} &= X_{(1)}, \\ M_{max} &= X_{(n)}. \end{aligned} \tag{2.1}$$

Размах значений выборки:

$$R = M_{max} - M_{min}. \tag{2.2}$$

Оценка мат. ожидания:

$$\hat{\mu}(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2.3}$$

Оценка исправленной дисперсии:

$$S^2(\vec{X}_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2. \tag{2.4}$$

# 3. Эмпирическая плотность и гистограмма

Пусть $\vec{x}$ – выборка из генеральной совокупности $X$. Если объем этой выборки $n$ велик, то ее значения $x_i$ группируют в интервальный статистический ряд. Для определения интервалов отрезок $J = [x_{(1)}, x_{(n)}]$ делят на $m$ равных частей, причем количество интервалов определяется как:

$$m = [\log_2 n] + 2. \tag{3.1}$$

Сами интервалы определяются следующими выражениями:

$$J_i = [x_{(1)} + (i-1) \cdot \Delta, x_{(1)} + i \cdot \delta), i = \overline{1, m-1}, \tag{3.2}$$

$$J_m = [x_{(1)} + (m-1) \cdot \Delta, x_{(n)}], \tag{3.3}$$

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}. \tag{3.4}$$

Интервальным статистическим рядом называется таблица:

| | | | | |
|---|---|---|---|---|
| $J_1$ | ... | $J_i$ | ... | $J_m$ |
| $n_1$ | ... | $n_i$ | ... | $n_m$ |

где $J_i$ – $i$-ый полуинтервал статистического ряда, $n_i$ – количество элементов выборки $\vec{x}$, попавших в $J_i$.

Эмпирическая плотность, отвечающая выборке $\vec{x}$ – функция вида:

$$\hat{f}(x) = \begin{cases} \dfrac{n_i}{n\Delta}, & x \in J_i, i = \overline{1, m}, \\ 0, & \text{иначе}, \end{cases} \tag{3.5}$$

где $J_i$ – $i$-ый полуинтервал статистического ряда, $n_i$ – количество элементов выборки $\vec{x}$, попавших в $J_i$, $n$ – количество элементов выборки.

Гистограммой называется график эмпирической плотности.

# 4. Эмпирическая функция распределения

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

Эмпирической функцией распределения называют функцию $F_n : \mathbb{R} \to \mathbb{R}$, определенную как:

$$F_n(x) = \frac{n(x, \vec{x})}{n}. \tag{4.1}$$

Замечание:

1. Обладает всеми свойствами функции распределения;
2. Кусочно-постоянна;
3. Если все элементы вектора различны, то

$$F_n(x) = \begin{cases} 0, & x \le x_{(1)}, \\ \dfrac{i}{n}, & x_{(i)} < x \le x_{(i+1)}, \\ 1, & x_{(n)} < x. \end{cases} \tag{4.2}$$

# 5. Код программы

```
function main()
    f = fopen("X.txt", "r");
    buf = textscan(f, "%f", 'Delimiter', ' ');
    X = buf{1,1};
    X = X';
    fclose(f);
    X = sort(X);

    M_min = X(1);
    M_max = X(end);
    fprintf("M_min = %d, M_max = %d\n", M_min, M_max);

    R = M_max - M_min;
    fprintf("R = %d\n", R);

    MU = MX(X);
    S2 = DX(X);
    fprintf('MU = %d, S^2 = %d\n', MU, S2);

    m = floor(log2(length(X)) + 2);
    groupOnIntervals(X, m);
    hold on;

    sigma = sqrt(S2);
    x_values = (M_min - 1.5):(R / 100):(M_max + 1.5);
    Y = zeros(1, length(x_values));
    for i = 1:length(x_values)
        Y(i) = (1 / (sqrt(2 * pi) * sigma)) * exp(-(x_values(i) - MU) *
            (x_values(i) - MU) / (2 * sigma * sigma));
    endfor
    plot(x_values, Y), grid;

    figure;

    Y = zeros(1, length(X)+10);
    Xd = zeros(1, length(X)+10);
    for i = 1 : 5
        Xd(i) = X(1) - (6-i)*0.1;
        Xd(i+length(X)+5) = X(end) + 0.1 * i;
    endfor
    for i = 1 : length(X)
        Xd(i+5) = X(i);
    endfor

    for i = 1 : length(Xd)
```

```
        if Xd(i) < X(1)
            Y(i) = 0;
        elseif Xd(i) > X(length(X))
            Y(i) = 1;
        else
            Y(i) = (i - 5) / length(X);
        endif
    endfor
    stairs(Xd, Y), grid;

    hold on;

    Y = zeros(1, length(x_values));
    func = @(x) exp(-(x - MU).* (x - MU) / (2 * S2));
    for i = 1:length(x_values)
        Y(i) = 1 / (sigma * sqrt(2 * pi)) *
            integral(func, -Inf, x_values(i));
    endfor
    plot(x_values, Y), grid;
end

function mx = MX(X)
    n = length(X);
    sum = sum(X);
    mx = sum / n;
endfunction

function dx = DX(X)
    n = length(X);
    MX = MX(X);
    dx = sum((X - MX).^2) / (n - 1);
endfunction

function groupOnIntervals(X, m)
    n = length(X);
    delta = (X(end) - X(1)) / m;
    intervals = zeros(2, m + 1);
    intervals(1, 1) = X(1);
    for i = 1:m
        intervals(1, i + 1) = intervals(1, i) + delta;
    endfor

    interval_ind = 1;
    for i = 1:n
        if (X(i) < intervals(1, interval_ind + 1))
            intervals(2, interval_ind) += 1;
        elseif i != n
            interval_ind += 1;
```

```
                intervals(2, interval_ind) += 1;
            endif
        endfor

        fprintf("m = %d\n", m);
        for i = 1:m-1
            fprintf("[%6.2f; %6.2f) - %d\n", intervals(1,i),
                intervals(1, i + 1), intervals(2, i));
        endfor
        fprintf("[%6.2f; %6.2f] - %d\n", intervals(1, m),
            intervals(1, m + 1), intervals(2, m));

        for i = 1:m
            intervals(2, i) /= (n * delta);
        endfor

        stairs(intervals(1, :), intervals(2, :)), grid;
endfunction
```

# 6. Результаты расчетов для выборки по варианту

| $M_{min}$ | $M_{max}$ | $R$ | $\hat{\mu}(\vec{x}_n)$ | $S^2(\vec{x}_n)$ | $m$ |
|---|---|---|---|---|---|
| -12.2 | -7.77 | 4.43 | -10.1318 | 0.846041 | 8 |

| $i$ | $J_i$ | $n_i$ |
|---|---|---|
| 1 | $[-12.20; -11.65)$ | 4 |
| 2 | $[-11.65; -11.09)$ | 15 |
| 3 | $[-11.09; -10.54)$ | 23 |
| 4 | $[-10.54; -9.98)$ | 27 |
| 5 | $[-9.98; -9.43)$ | 17 |
| 6 | $[-9.43; -8.88)$ | 25 |
| 7 | $[-8.88; -8.32)$ | 6 |
| 8 | $[-8.32; -7.77]$ | 3 |

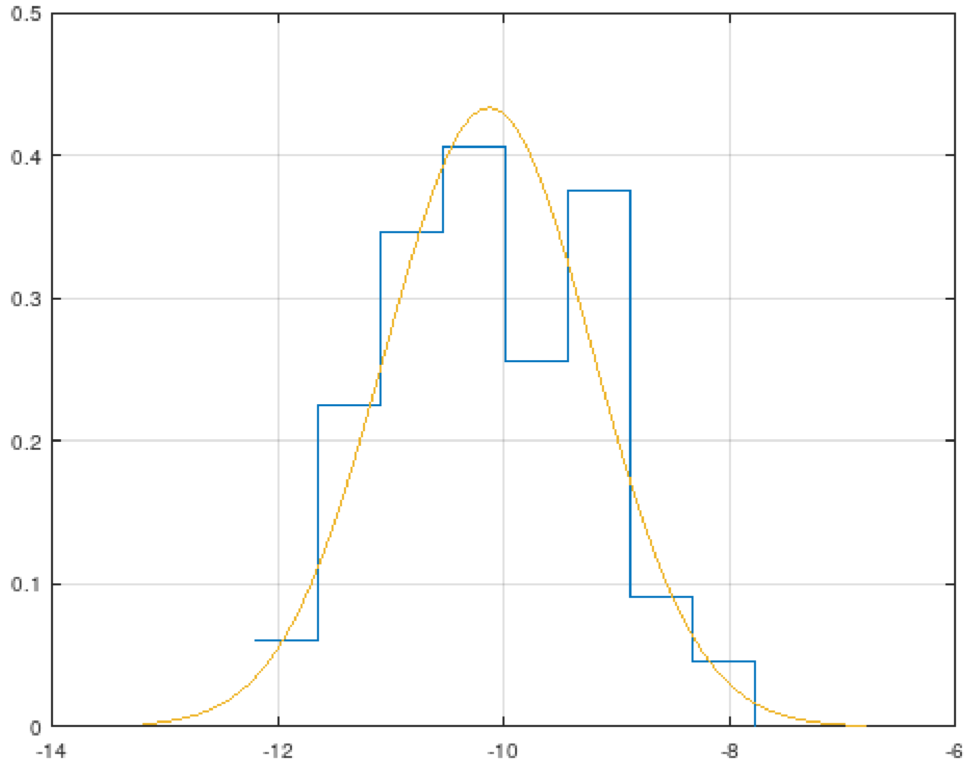


Рисунок 6.1 – Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочными мат. ожиданием и дисперсией

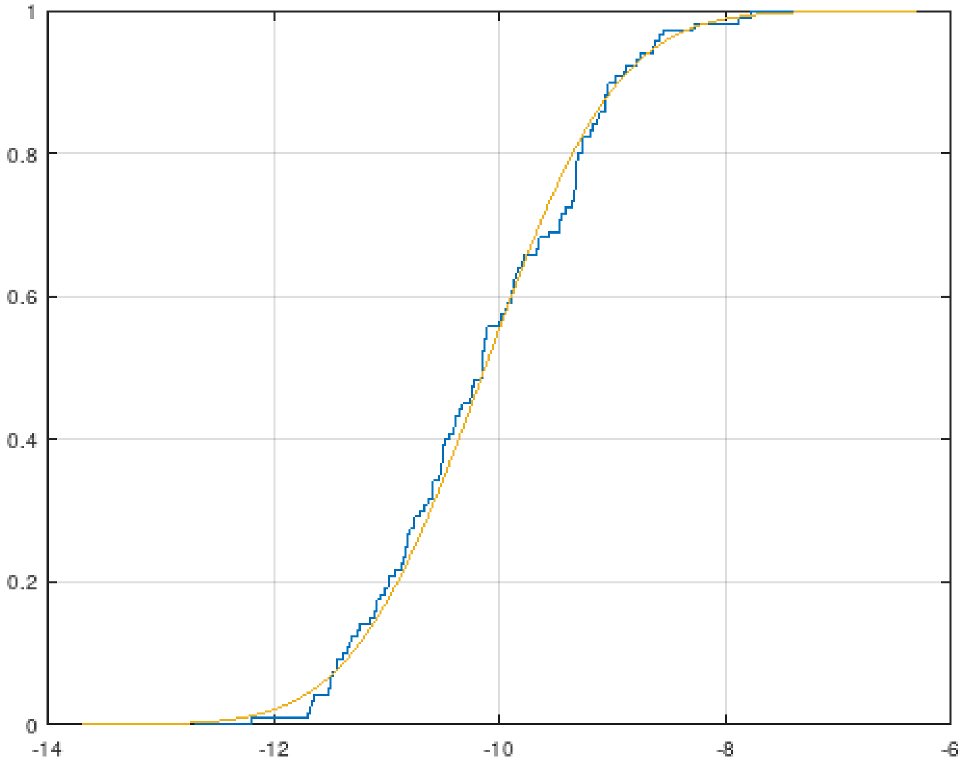


Рисунок 6.2 – График эмпирической функции распределения и функции распределения нормаль- ной случайной величины с выборочными мат. ожиданием и дисперсией