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 распределения

Дисциплина: Математическая статистика

Студент ИУ7-61Б _______________ Н. А. Котляров
Группа — Подпись, дата — И. О. Фамилия

Преподаватель _______________ П. А. Власов
Подпись, дата — И. О. Фамилия

Москва, 2023 г.

Оглавление

1 Содержание

Цель работы: построение гистограммы и эмпирической функции распределения.

- Для выборки объема n из генеральной совокупности X реализовать в виде программы на ЭВМ:

  1. вычисление максимального значения M_{max} и минимального значения M_{min};
  2. размаха R выборки;
  3. вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;
  4. группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
  5. построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
  6. построение на другой координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.

- Провести вычисления и построить графики для выборки из индивидуального варианта.

2 Теория

Пусть $\vec{x} = (x_1, \dots x_n)$ – выборка из генеральной совокупности X объема n.

1. Вариационный ряд — отсортированная выборка их генеральной совокупности X;

2. Максимальное значение выборки: $M_{min} = x_{(1)}$, являющийся первым элементом вариационного ряда;

3. Минимальное значение выборки: $M_{max} = x_{(n)}$, являющийся последним элементом вариационного ряда;

4. Размах выборки: $R = M_{max} - M_{min}$,

5. Оценка математического ожидания: $\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$,

6. Оценка дисперсии: $S^2(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

$n(t, \vec{x})$ – число компонент вектора $\vec{x}$, которые меньше чем t. *Эмпирической функцией распределения*, построенной по выборке $\vec{x}$ называется функция $F_n : \mathbb{R} \to \mathbb{R}$, определенная правилом 2.1.

$$F_n(t) = \frac{n(t, \vec{x})}{n} \tag{2.1}$$

Интервальный статистический ряд — это ряд $J = [x_{(i)}, x_{(n)}]$, который разбивают на m промежутков, ширина которых определяется согласно 2.2:

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(i)}}{m}, \tag{2.2}$$

$$\begin{aligned} &J_i = \left[x_{(1)} + (i-1)\,\Delta;\ x_{(i)} + i\Delta\right),\ i = \overline{1,\ m-1}, \\ &J_m = \left[x_{(1)} + (m - 1\Delta),\ x_{(n)}\right). \end{aligned} \tag{2.3}$$

Эмпирической плотностью распределения соответствующей выборке $\vec{x}$ называется функция 2.4:

$$f_n(x) = \begin{cases} \dfrac{n_i}{n \cdot \Delta} & , x \in J_i, \\ 0 & , \text{иначе}. \end{cases} \tag{2.4}$$

3 Практика

```
function main()
        pkg load statistics

        X = dlmread("1.txt", ",");

        X = sort(X);
        % disp(X)

        % (а) минимальное и максимальное значение
        m_max = max(X);
        m_min = min(X);

        fprintf("(а) Максимальное значение выборки (M_max) = %f\n", m_max)
        fprintf("    Минимальное значение выборки  (M_min) = %f\n", m_min)
        fprintf("————————————————————————————————————————————————————————\n")

        % (б) размах выборки
        r = m_max - m_min;
        fprintf("(б) Размах выборки (R) = %f\n", r)
        fprintf("——————————————————————————————————\n")

        % (в) вычисление оценок MX DX
        n = length(X);
        mu = sum(X) / n;
        s_2 = sum((X - mu).^2) / (n - 1);
        sigma = sqrt(s_2);

        fprintf("(в) Оценка математического ожидания (mu) = %f\n", mu)
        fprintf("    Оценка дисперсии (s_2) = %f\n", s_2)
        fprintf("——————————————————————————————————————————————\n")

        % (г) группировка значений выборки в m = [log_2 n] + 2 интервала
        m = floor(log2(n)) + 2;

        bins = [];
        cur = m_min;

        for i = 1:(m + 1)
                bins(i) = cur;
                cur = cur + r / m;
        end

        eps = 1e-6;
        counts = [];
        j = 1;

```

```
    for i = 1:(m - 1)
        cur_count = 0;
        for j = 1:n
            if (bins(i) < X(j) || abs(bins(i)-X(j)) < eps) && X(j) < bins(i+1)
                cur_count = cur_count + 1;
            endif
        endfor
        counts(i) = cur_count;
    endfor

    cur_count = 0;

    for j = 1:n
        if (bins(m) < X(j) || abs(bins(m) - X(j)) < eps) && (X(j) < bins(m + 1) || abs(bins(m + 1) - X(j)) < eps)
            cur_count = cur_count + 1;
        endif
    endfor

    counts(m) = cur_count;

    fprintf("(г) группировка значений выборки в m = [log_2 n] + 2 интервала:\n");

    for i = 1:(m - 1)
        fprintf("    [%f : %f) - %d вхожд.\n", bins(i), bins(i + 1), counts(i));
    end

    fprintf("    [%f : %f] - %d вхожд.\n", bins(m), bins(m + 1), counts(m));
    fprintf("--------------------------------------------------\n");

    % (д) построение гистограммы и графика функции плотност

    fprintf("(д) построение гистограммы и графика функции плотности\n");
    fprintf("    распределения вероятностей нормальной случайной величины\n");

    figure;
    hold on;
    grid on;
    n = length(X);
    delta = r / m;
    middles = zeros(1, m);
    xx = zeros(1, m);

```

```
    for i = 1:m
        xx(i) = counts(i) / (n * delta);
    endfor

    for i = 1:m
        middles(i) = bins(i + 1) - (delta / 2);
    endfor

    fprintf("    высоты столбцов гистограммы:\n");

    for i = 1:m
        fprintf("    [%d] : %f\n", i, xx(i));
    endfor

    fprintf("[проверка] площадь гистограммы s = %f\n", sum(xx) * delta);

    set(gca, "xtick", bins);
    set(gca, "ytick", xx);
    set(gca, "xlim", [min(bins) - 1, max(bins) + 1]);
    bar(middles, xx, 1, "facecolor", "g", "edgecolor", "w");

    X_n = m_min:(sigma / 100):m_max;
    X_pdf = normpdf(X_n, mu, sigma);
    plot(X_n, X_pdf, "r");
    xlabel('X')
    ylabel('P')
    print -djpg hist.jpg
    hold off;

    fprintf("------------------------------------------------------------\n");

    % (e) построение графика эмпирической функции распределения

    fprintf("(e)построение графика эмпирической функции распределения\n");
    fprintf("   и функции распределения нормальной случайной величины\n");

    figure;
    hold on;
    grid on;
    n = length(X);
    xx = zeros(1, length(X));
    curss = 1;

    bins = zeros(1, length(X));
    bins(1) = X(1);
    for i = 2:length(X)
        if (bins(curss) != X(i))
            curss+= 1;
```

```
                bins(curss) = X(i);
                xx(curss) = xx(curss-1);
            endif
            if (bins(curss) == X(i))
                xx(curss) += 1;
            endif
    end

    xx = xx ./ length(X);

    for i = curss:length(X)
        xx(i) = 1;
    end

    X_n = (min(X) - 0.5):(sigma / 100):(max(X) + 1.5);
    X_cdf = normcdf(X_n, mu, sigma);
    plot(X_n, X_cdf, "r");

    stairs(bins, xx);
    xlabel('X')
    ylabel('F')
    print -djpg cdf.jpg
    hold off;
end
```

4 Результаты

```
(а) Максимальное значение выборки (M_max) = 0.010000
    Минимальное значение выборки  (M_min) = -4.330000
-----------------------------------------------
(б) Размах выборки (R) = 4.340000
-----------------------------
(в) Оценка математического ожидания (mu) = -2.058500
    Оценка дисперсии (s_2) = 0.943988
---------------------------------------
(г) группировка значений выборки в m = [log_2 n] + 2 интервала:
    [-4.330000 : -3.787500) - 5 вхожд.
    [-3.787500 : -3.245000) - 8 вхожд.
    [-3.245000 : -2.702500) - 22 вхожд.
    [-2.702500 : -2.160000) - 15 вхожд.
    [-2.160000 : -1.617500) - 24 вхожд.
    [-1.617500 : -1.075000) - 25 вхожд.
    [-1.075000 : -0.532500) - 17 вхожд.
    [-0.532500 : 0.010000] - 4 вхожд.
---------------------------------------
(д) построение гистограммы и графика функции плотности
    распределения вероятностей нормальной случайной величины
    высоты столбцов гистограммы:
    [1] : 0.076805
    [2] : 0.122888
    [3] : 0.337942
    [4] : 0.230415
    [5] : 0.368664
    [6] : 0.384025
    [7] : 0.261137
    [8] : 0.061444
[проверка] площадь гистограммы s = 1.000000
---------------------------------------
```

Рисунок 4.1 – Результат работы программы

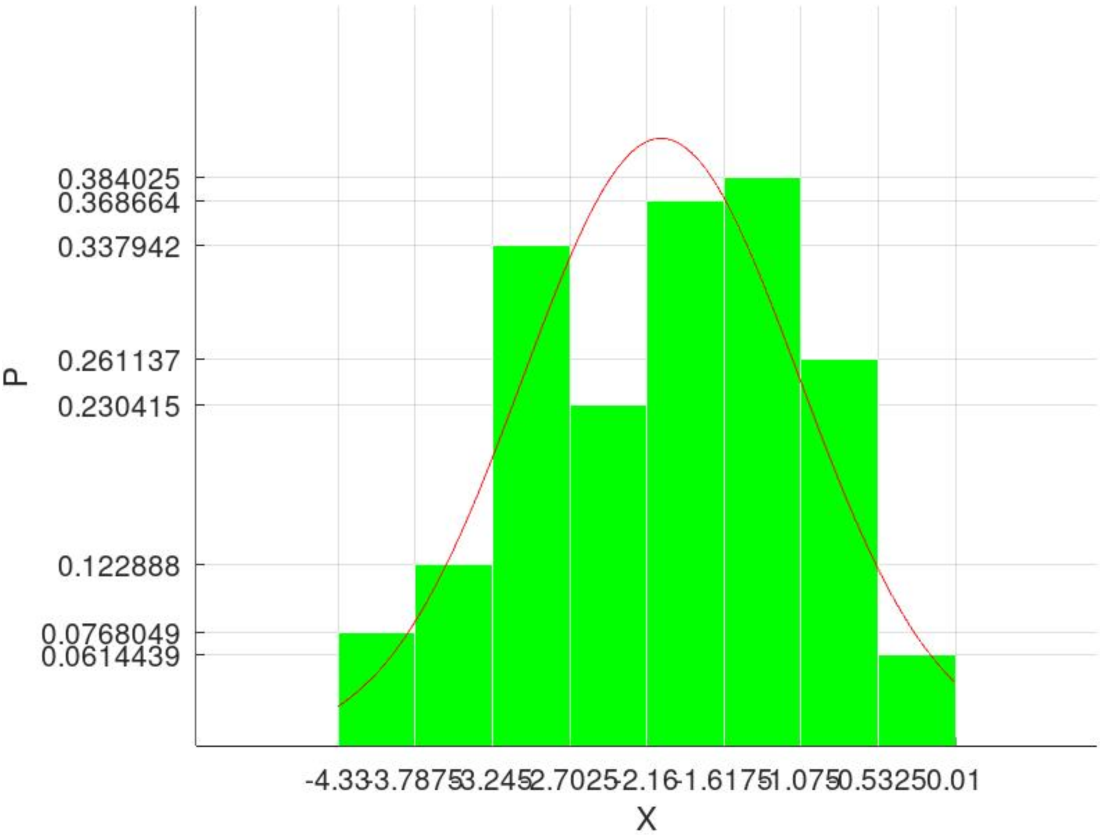

Рисунок 4.2 – Гистограмма и график функции плотности распределения вероятностей нормальной случайной величины

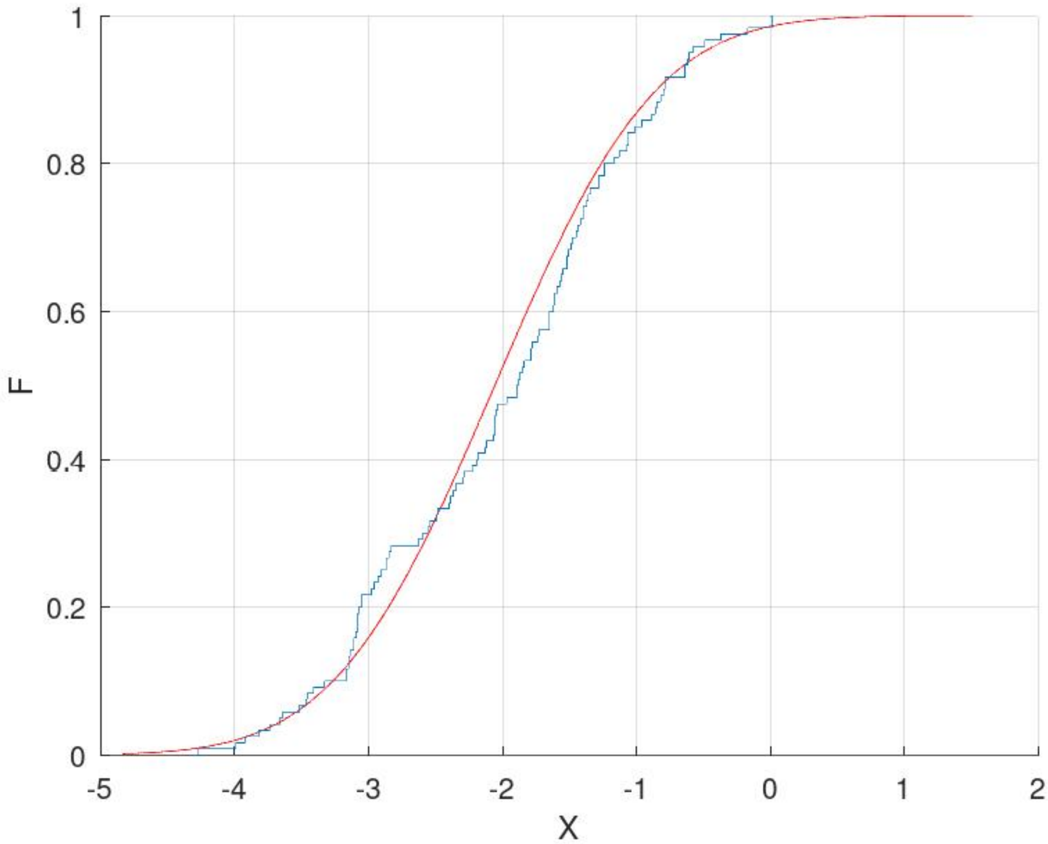

Рисунок 4.3 – График эмпирической функции распределения и функции распределения нормальной случайной величины