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 Фам Минь Хиеу

Группа ИУ7-62Б

Оценка (баллы) ______________________________

Преподаватель Власов П. А.

Москва — 2024 г.

Содержание работы

Цель работы: построение гистограммы и эмпирической функции распределения.

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

(a) вычисление максимального значения M_{max} и минимального значения M_{min};

(b) размаха R выборки;

(c) вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;

(d) группировку значений выборки в $m = [\log_2 n] + 2$ интервала;

(e) построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;

(f) 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.

Теоретические сведения

Множество возможных значений случайной величины X называют **генеральной совокупностью** случайной величины X.

Любое возможное значение $\vec{x} = (x_1, \ldots x_n)$ случайной выборки $\vec{X}_n$ будем называть **выборкой** из генеральной совокупности X. Число n характеризует объем выборки, а числа x_i представляют собой элементы выборки $\vec{X}_n$.

Пусть $\vec{x} = (x_1, \dots x_n)$ — выборка объема n из генеральной совокупности X. Ее можно упорядочить, расположив значения в неубывающем порядке:

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)} \quad (1)$$

Такую последовательность чисел из формулы 1 называют **вариационным рядом**.

Тогда $x_{(1)}$ является **минимальным** значением выборки, а $x_{(n)}$ - **максимальным** значением выборки.

Размахом выборки называют разность между максимальным и минимальным значениями.

Рассмотрим функцию $n(x, \vec{X_n})$, которая для каждого значения $x \in R$ и каждой реализации $\vec{x_n}$ случайной выборки $\vec{X_n}$ принимает значение $n(x, \vec{x_n})$, равное числу элементов в выборке $\vec{x_n}$, меньших х.

Тогда **эмпирической функцией распределения** называется функция:

$$F(x, \vec{X_n}) = \frac{n(x, \vec{x_n})}{n} \quad (2)$$

Оценку математического ожидания можно подсчитать по формуле:

$$\hat{\mu}(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad (3)$$

Смещённую выборочную дисперсию можно подсчитать по формуле:

$$S_n^2(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2. \quad (4)$$

Исправленную выборочную дисперсию можно подсчитать по формуле:

$$S^2(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \quad (5)$$

При больших объемах выборки n обычно производят группирование исходных данных следующим образом. Промежуток $J = [x_{(1)}, x_{(n)}]$, содержащий все выборочные значения, разбивают на m полуинтервалов $J_1, \dots, J_m$, как правило, одинаковой длины δ и таких, что каждый из них, кроме последнего, содержит левую границу, а последний содержит обе границы, и подсчитывают число n_i элементов выборки, попавших в i-ый промежуток J_i, а результаты представляют в виде таблицы, которую называют **интервальным статистическим рядом**.

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

$$f_n(x) = \begin{cases} \dfrac{n_i}{n \cdot \Delta} & , x \in J_i, \\ 0 & , \text{иначе.} \end{cases} \tag{6}$$

График эмпирической функции плотности называется **гистограммой**.

Код программы

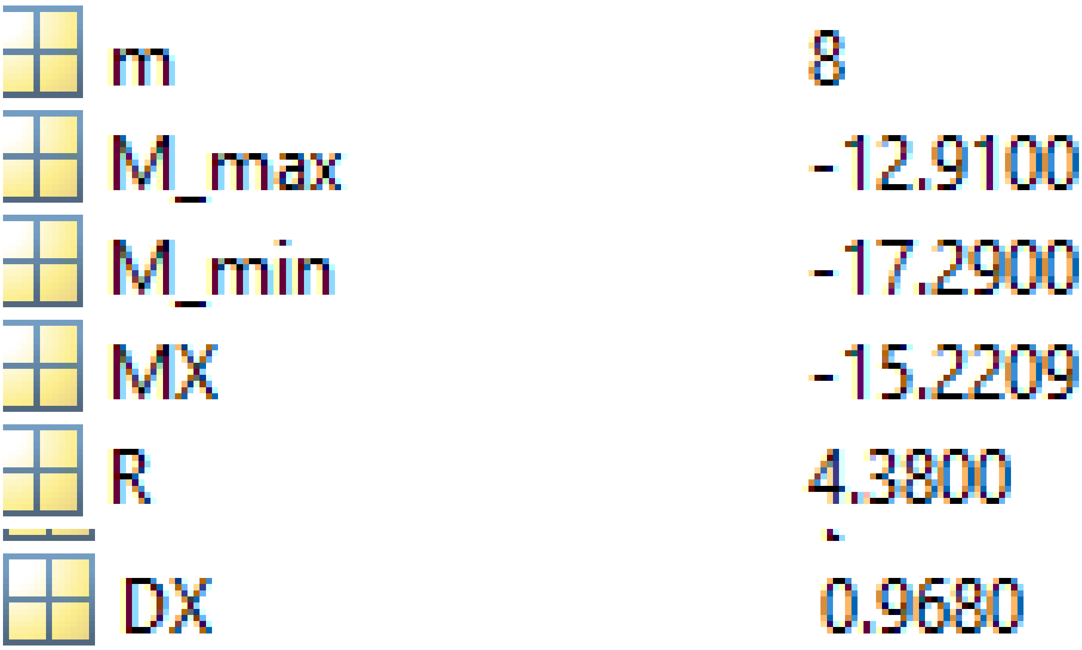
На листинге 1 представлен код программы:

```
X = [-14.34,-16.97,-14.09,-14.74,-16.69,-13.85,-15.55,-14.62,-13.30...
 ,-15.52,-14.75,-16.51,-17.15,-16.87,-15.06,-13.60,-14.48,-14.71,-14.17,...
 -13.88,-14.55,-15.37,-14.81,-16.05,-17.06,-15.86,-15.12,-15.98,-14.16,...
 -15.81,-15.06,-16.19,-16.22,-16.19,-14.87,-15.62,-15.86,-15.25,-16.34,...
 -14.44,-14.72,-15.17,-15.24,-14.44,-15.93,-14.87,-16.53,-15.76,-15.12,...
 -12.91,-16.06,-16.06,-14.89,-15.57,-13.59,-16.84,-13.88,-14.33,-15.45,...
 -16.58,-16.05,-14.34,-13.55,-16.78,-14.15,-14.28,-14.40,-13.98,-16.23,...
 -15.35,-14.77,-15.61,-15.59,-15.64,-14.76,-17.18,-15.13,-15.01,-14.21,...
 -13.91,-16.55,-15.44,-14.03,-16.44,-15.57,-15.07,-16.28,-16.30,-15.74,...
 -14.03,-14.85,-15.73,-15.81,-14.42,-14.14,-15.14,-15.49,-16.42,-14.22,...
 -14.20,-17.17,-15.82,-14.96,-14.75,-14.98,-13.64,-14.00,-17.29,-14.51,...
 -16.18,-15.70,-15.07,-14.28,-14.55,-13.85,-15.36,-15.74,-14.61,-16.32,-15.34];


M_max = max(X);
M_min = min(X);



R = M_max - M_min;


MX = mean(X);
DX = var(X);


m = floor(log2(length(X))) + 2;
h = histogram(X, m);



sigma = std(X);
x = (M_min - 1):(sigma / 100):(M_max + 1);
f = normpdf(x, MX, sigma); % normal probability distribution function
figure;
heights = h.Values / (sum(h.Values) * h.BinWidth);
centers = [];
for i = 1:(length(h.BinEdges) - 1)
    centers = [centers, (h.BinEdges(i + 1) + h.BinEdges(i)) / 2];
end


hold on;
```

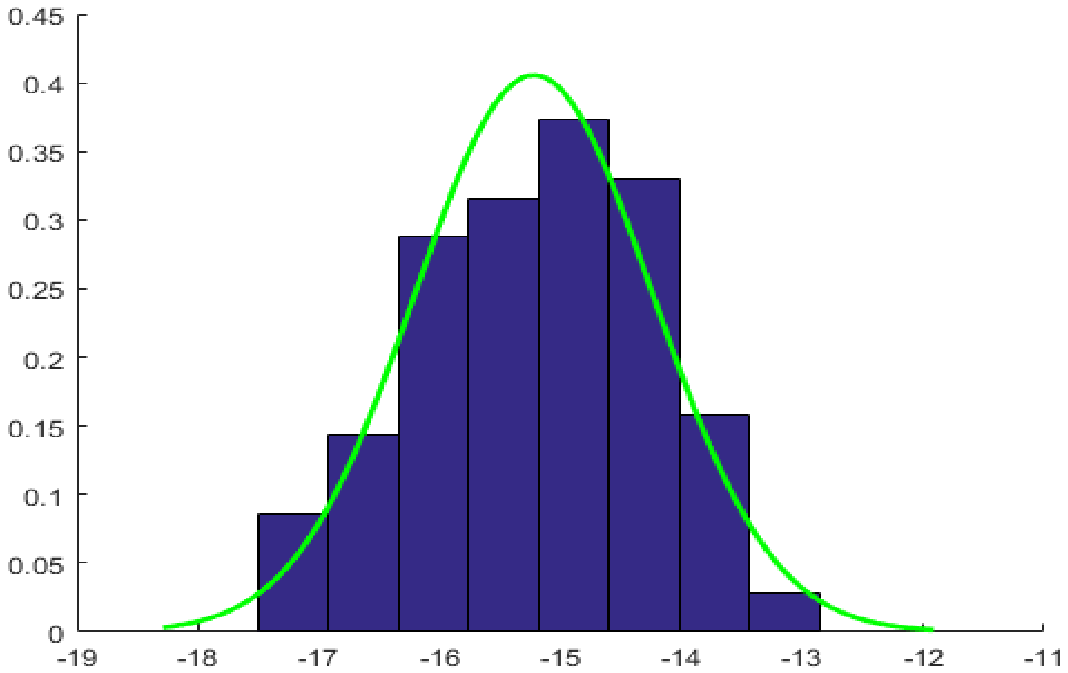
```
bar(centers, heights, 1);
plot(x, f, 'g', 'LineWidth', 2);

F = normcdf(x, MX, sigma);
figure;
hold on;
ecdf(X);
plot(x, F, 'r');
```

Результаты работы программы

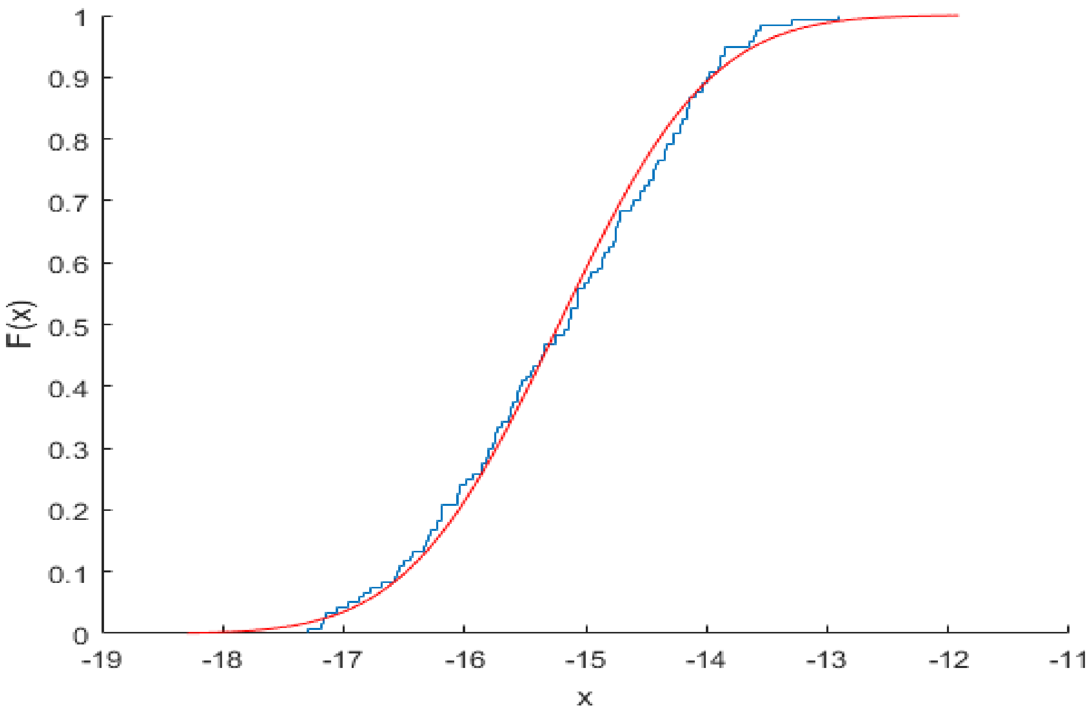
Результат работы программы представлен на рисунке 1:



На рисунке 2 представлен график гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины:



На рисунке 3 представлен графика эмпирической функции распределения и функции распределения нормальной случайной величины:



F(x)
x