Московский государственный технический университет
имени Н. Э. Баумана

Факультет: Информатика и системы управления
Кафедра: Программное обеспечение ЭВМ и информационные технологии
Дисциплина: Математическая статистика

Лабораторная работа №1
Гистограмма и эмпирическая функция распределения

Выполнила: Якубаускайте М.
Группа:ИУ7-65Б
Вариант: 14

Москва, 2019г.

Содержание

1 Постановка задачи

Цель работы: построение гистограммы и эмпирической функции распределения.

Содержание работы

1. Для выборки объёма 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 $\mathsf{M}X$ и дисперсии $\mathsf{D}X$;
   - 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.

2 Отчёт

2.1 Формулы для вычисления величин

Количество интервалов

$$m = [\log_2 n] + 2 \tag{1}$$

Минимальное значение выборки

$$M_{\min} = \min\{x_1, \ldots, x_n\}, \quad \text{где} \tag{2}$$

- $(x_1, \ldots, x_n)$ — реализация случайной выборки.

Максимальное значение выборки

$$M_{\max} = \max\{x_1, \ldots, x_n\}, \quad \text{где} \tag{3}$$

- $(x_1, \ldots, x_n)$ — реализация случайной выборки.

Размах выборки

$$R = M_{\max} - M_{\min}, \quad \text{где} \tag{4}$$

- $M_{\max}$ — максимальное значение выборки;
- $M_{\min}$ — минимальное значение выборки.

Оценка математического ожидания

$$\hat{\mu}(\vec{X}_n) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{5}$$

Выборочная дисперсия

$$\hat{\sigma}^2(\vec{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2\,. \tag{6}$$

Несмещённая оценка дисперсии

$$S^2(\vec{X}_n) = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\,. \tag{7}$$

2.2 Эмпирическая плотность и гистограмма

Определение 2.1. *Эмпирической плотностью распределения случайной выборки* $\vec{X}_n$ называют функцию

$$f_n(x) = \begin{cases} \frac{n_i}{n\,\Delta}, & x \in J_i,\ i = \overline{1, m}; \\ 0, & \text{иначе.} \end{cases}, \quad \text{где} \tag{8}$$

- J_i, $i = \overline{1;m}$, — полуинтервал из $J = [x_{(1)}, x_{(n)}]$, где
$$x_{(1)} = \min\{x_1, \ldots, x_n\}, \qquad\qquad x_{(n)} = \max\{x_1, \ldots, x_n\}; \tag{9}$$
при этом все полуинтервалы, кроме последнего, не содержат правую границу т. е.
$$J_i = [x_{(1)} + (i-1)\Delta, x_{(i)} + i\Delta), \quad i = \overline{1, m-1}; \tag{10}$$
$$J_m = [x_{(1)} + (m-1)\Delta, x_{(1)} + m\Delta); \tag{11}$$
- m — количество полуинтервалов интервала $J = [x_{(1)}, x_{(n)}]$;
- Δ — длина полуинтервала J_i, $i = \overline{1, m}$ равная
$$\Delta = \frac{x_{(n)} - x_{(1)}}{m} = \frac{|J|}{m}; \tag{12}$$
- n_i — количество элементов выборки в полуинтервале J_i, $i = \overline{1, m}$;
- n — количество элементов в выборке.

Определение 2.2. График функции $f_n(x)$ называют гистограммой.

2.3 Эмпирическая функция распределения

- $\vec{X}_n = (X_1, \ldots, X_n)$ — случайная выборка;
- $\vec{x}_n = (x_1, \ldots, x_n)$ — реализация случайной выборки $\vec{X}_n$;
- $n(x, \vec{x}_n)$ — количество элементов выборки $\vec{x}_n$, которые меньше x.

Определение 2.3. *Эмпирической функцией распределения* называют функцию
$$F_n \colon \Re \to \Re, \quad F_n(x) = \frac{n(x, \vec{x}_n)}{n} \tag{13}$$

Замечание. $F_n(x)$ обладает всеми свойствами функции распределения. При этом она кусочно-постоянна и принимает значения
$$0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{(n-1)}{n}, 1$$

Замечание. Если все элементы вектора $\vec{x}_n$ различны, то
$$F_n(x) = \begin{cases} 0, & x \le x_{(1)}; \\ \frac{i}{n}, & x_{(i)} < x \le x_{(i+1)},\ i = \overline{1, n-1}; \\ 1, & x > x_{(n)}. \end{cases} \tag{14}$$

Замечание. Эмпирическая функция распределения позволяет интерпретировать выборку $\vec{x}_n$ как реализацию дискретной случайной величины $\widetilde{X}$ ряд распределения которой

$\widetilde{X}$	$x_{(1)}$	$\ldots$	$x_{(n)}$
P	$1/n$	$\ldots$	$1/n$

Это позволяет рассматривать числовые характеристики случайной величины $\widetilde{X}$ как приближённые значения числовых характеристик случайной величины X.

3 Листинг программы

```
close all;


variationalSeries = sort(csvread
    ('C:\Users\marianna\МатСтат\Lab1\selection.csv'));
Mmin = variationalSeries(1);
Mmax = variationalSeries(length(variationalSeries));
R = Mmax - Mmin;
mu = NthSelectiveStartingTorque(variationalSeries, 1);
s2 = NthSelectiveAvgTorque(variationalSeries, 2);
m = floor(log2(length(variationalSeries))) + 2;
s = sqrt(s2);

m = floor(log2(length(variationalSeries))) + 2;
delta = R / m;

edges = Mmin : delta : Mmax;
for_histogram = histogram
    (variationalSeries, edges, 'normalization', 'pdf');

edge_center = zeros(length(for_histogram.Values), 1);
for i = 1 : length(edge_center)
    edge_center(i) = (edges(i) + delta / 2);
end

x_n = (mu - R) : delta / 100 :(mu + R);

y_plotn = normpdf(x_n, mu, s);
figure
bar(edge_center, for_histogram.Values, 1);
hold on;
plot(x_n, y_plotn, 'r');
hold off;

y_normal = normcdf(x_n, mu, s);
figure
ecdf(variationalSeries);
hold on;
plot(x_n, y_normal, 'r');
hold off;

function y = NthSelectiveStartingTorque(x, n)
sum = 0;
len = length(x);
for i = 1:len
    sum = sum + x(i)^n;
end
y = sum/len;
```

```
end

function k = NthSelectiveAvgTorque(x, n)
mu1 = NthSelectiveStartingTorque(x, 1);
sum = 0;
len = length(x);
for i = 1:len
    sum = sum + (x(i) - mu1)^n;
end
k = sum/(len - 1);
end
```

4 Результаты расчётов

$$
\begin{aligned}
M_{\min} &= 0.0900; \\
M_{\max} &= 5.4700; \\
R &= 5.3800; \\
\hat{\mu}(\vec{X}_n) &= 3.0554; \\
S^2(\vec{X}_n) &= 1.2173.
\end{aligned}
$$

Интервальная групировка значений выборки при $m = 8$:

$$[0.0900; 0.4263), [0.4263; 1.0988), [1.0988; 1.7713), [1.7713; 2.4438),$$

$$[2.4438; 3.1163), [3.1163; 3.7888), [3.7888; 4.4613), [4.4613; 5.4700]$$

5 Графики

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

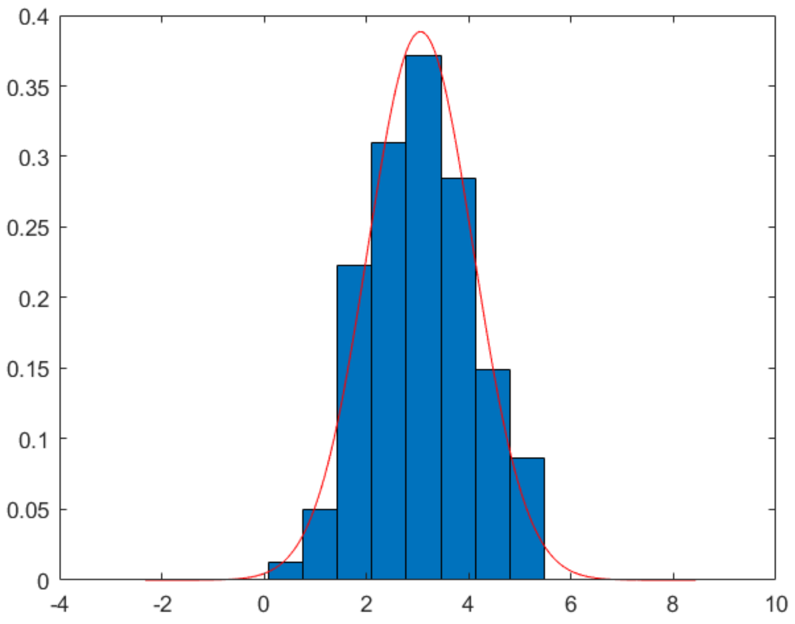

Рис. 1: Гистограмма и график функции плотности распределения нормальной случайной величины.

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

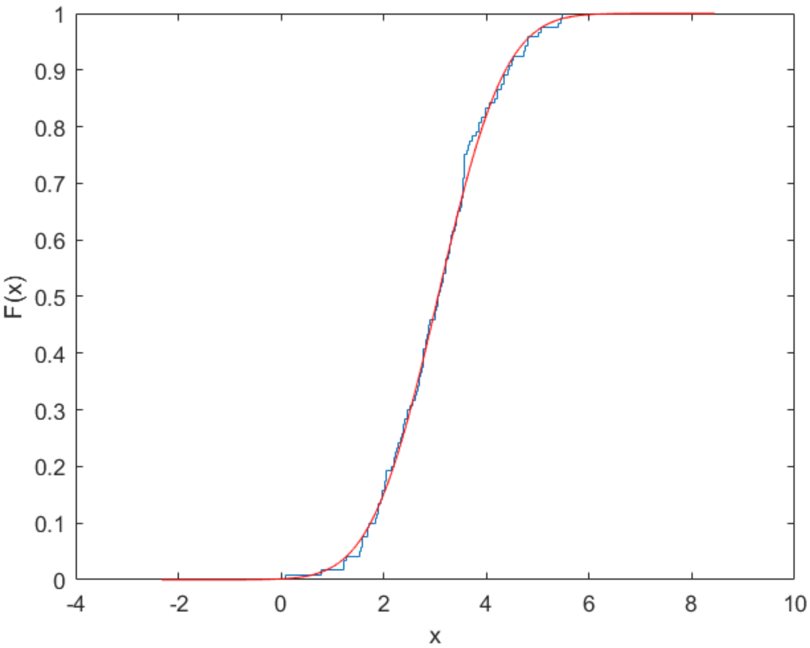

Рис. 2: График эмпирической функции распределения и функции распределения нормальной случайной величины.