Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»
КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

Лабораторная работа № 1

Вариант 22

Дисциплина	Математическая статистика.
Тема	
Студент	Тимонин А. С.
Группа	ИУ7-62Б
Оценка (баллы)	
Преподаватель	Власов П.А.

Москва, 2020 г.

Теоретическая часть

Формулы для вычисления

Для генеральной совокупности $\vec{x} = (x_1, \ldots, x_n)$

Формула для вычисления максимального значения $M_{\max}$:

$$M_{\max} = \max(x_1, \ldots, x_n)$$

Формула для вычисления минимального значения $M_{\min}$:

$$M_{\min} = \min(x_1, \ldots, x_n)$$

Размах выборки R считается по формуле:

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

Вычисление оценки математического ожидания MX:

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

Вычисление оценки дисперсии DX:

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

Определение эмпирической плотности и гистограммы

Интервальный статистический ряд

Пусть $\vec{x}$ - выборка из генеральной совокупности X. Если объем n этой выборки велик ($n \geq 50$), то значения x_i группируют не только в статистический ряд, но и в так называемый интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n))}]$ делят на p равновеликих частей:

$$J_i = [a_i, a_{i+1}), i = \overline{0; p-i}$$

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

где $a_i = x_{(1)} + i\Delta,\ t = \overline{0;p}, \Delta = \frac{|J|}{p} = \frac{x_{(n)} - x_{(1)}}{p}$

Опр Интервальным статистическим рядом называют таблицу

J_1	$\ldots$	J_i	$\ldots$	J_p
n_1	$\ldots$	n_i	$\ldots$	n_p

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

Замечание

1. Очевидно, что $\sum_{i=1}^{p} n_i = n$

2. Для выборки р - числа интервалов можно пользоваться формулой $p = [\log_n n] + 1$

где [a] - целая часть числа a

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

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

Опр **Гистограммой** называют график эмпирической плотности

Определение эмперической функции распределения

Опр **Эмпирической функцией распределения** называют функцию

$$\mathcal{F}_n : \mathcal{R} \to \mathcal{R}$$

определенную условием

$$\hat{F}(x, \vec{x}) = \frac{n(x, \vec{x})}{n}$$

Практическая часть

Листинги программы

Листинг 1: Содержимое генеральной совокупности X

```
7.76, 5.96, 4.58, 6.13, 5.05,
6.40, 7.46, 5.55, 5.01, 3.79,
7.65, 8.87, 5.94, 7.25, 6.76,
6.92, 6.68, 4.89, 7.47, 6.53,
6.76, 6.96, 6.58, 7.92, 8.47,
6.27, 8.05, 5.24, 5.60, 6.69,
7.55, 6.02, 7.34, 6.81, 7.22,
6.39, 6.40, 8.28, 5.39, 5.68,
6.71, 7.89, 5.69, 5.18, 7.84,
7.18, 7.54, 6.04, 4.58, 6.82,
4.45, 6.75, 5.28, 7.42, 6.88,
7.10, 5.24, 9.12, 7.37, 5.50,
5.52, 6.34, 5.31, 7.71, 6.88,
6.45, 7.51, 6.21, 7.44, 6.15,
6.25, 5.59, 6.68, 6.52, 4.03,
5.35, 6.53, 3.68, 5.91, 6.68,
6.18, 7.80, 7.17, 7.31, 4.48,
5.69, 7.11, 6.87, 6.14, 4.73,
6.60, 5.61, 7.32, 6.75, 6.28,
6.41, 7.31, 6.68, 7.26, 7.94,
7.67, 4.72, 6.01, 5.79, 7.38,
5.98, 5.36, 6.43, 7.25, 5.54,
6.66, 6.47, 6.84, 6.13, 6.21,
5.52, 6.33, 7.55, 6.24, 7.84
```

Листинг 2: Точка входа программы

```
function lab01()
  X = csvread('/Users/antontimonin/Desktop/MatStat/lab_01/Data12.csv');
  X = sort(X);

  parametrs(X);
  intervals(X);
  graphs(X);
end
```

Листинг 3: Функции для вычисления параметров

```
function parametrs(X)
  X = sort(X);
  Xmin = X(1); fprintf('Mmin = %s\n', num2str(Xmin));
  Mmax = X(end); fprintf('Mmax = %s\n', num2str(Mmax));
  R = Mmax - Xmin; fprintf('R = %s\n', num2str(R));
  mu = expect(X); fprintf('mu = %s\n', num2str(mu));
  sigSqr = populVar(X); fprintf('sigma^2 = %s\n', num2str(sigSqr));
  sqr = unbisVariance(X); fprintf('S^2: %s\n', num2str(sqr));
  m = numSubintervals(length(X)); fprintf('m = %s\n ', num2str(m));
end

function m = expect(X)
```

```
  n = length(X);
  sum = 0;

  for i = 1:n
    sum = sum + X(i);
  end

  m = sum / n;
end

function sigSqr = populVar(X)
  n = length(X);
  sum = 0;

  for i = 1:n
    sum = sum + (X(i))^2;
  end

  mu = expect(X);
  sigSqr = sum / n - mu^2;
end

function sqr = unbisVariance(X)
  sigSqr = populVar(X);
  n = length(X);

  sqr = n / (n - 1) * sigSqr;
end

function m = numSubintervals(size)
  m = floor(log2(size) + 2);
end
```

Листинг 4: Функция вычисляющая интервалы

```
function intervals(X)
  m = numSubintervals(length(X));

  count = zeros(1, m+2);
  delta = (X(end) - X(1)) / m;

  J = X(1):delta:X(end);
  fprintf('%d\n', X(end));
  J(length(J)+1) = 20;

  j = 1;
  n = length(X);

  for i = 1:n
    if (j ~= m)
      if ((not (X(i) >= J(j) && X(i) < J(j+1))))
        j = j + 1;
        fprintf('[%.2f;%.2f)\n', J(j-1), J(j));
      end
    end
```

```
    count(j) = count(j) + 1;
  end

  fprintf('[%2.2f;%2.2f] -> %d\n', J(m), J(m + 1), count);

  Xbuf = count(1:m+2);
  for i = 1:m+2
    Xbuf(i) = count(i) / (n*delta);
  end

  stairs(J, Xbuf), grid;
end
```

Листинг 5: Отрисовка графов

```
function graphs(X)
  hold on;
  f(X, expect(X), ...
      unbisVariance(X), ...
      numSubintervals(length(X)));

  figure;
  empiricF(X);
  hold on;
  F(X, expect(X), ...
      unbisVariance(X), ...
      numSubintervals(length(X)));
end

function f(X, MX, DX, m)
  R = X(end) - X(1);
  delta = R/m;
  Sigma = sqrt(DX);
  Xn = (MX - R): delta/50 :(MX + R);
  Xn(length(Xn)+1) = 20;
  Y = normpdf(Xn, MX, Sigma);
  plot(Xn, Y);
end

function F(X, MX, DX, m)
  R = X(end) - X(1);
  delta = R/m;
  Xn = (MX - R): delta :(MX + R);
  Y = 1/2 * (1 + erf((Xn - MX) / sqrt(2*DX)));
  p2 = plot(Xn, Y,'Color',[.1 .7 .7],'LineWidth',1);
  hold off;
end

function empiricF(X)
  [yy, xx] = ecdf(X);
  yy(length(yy)+1) = 1;
  xx(length(xx)+1) = 20;
  stairs(xx, yy);
end
```

Результат выполнения программы

$$M_{\max} = 9.12$$

$$M_{\min} = 3.68$$

$$R = 5.44$$

$$\mu = 6.4596$$

$$S^2 = 1.1013$$

$[3.68; 4.36)$	3
$[4.36; 5.04)$	8
$[5.04; 5.72)$	20
$[5.72; 6.40)$	22
$[6.40; 7.08)$	30
$[7.08; 7.76)$	25
$[7.76; 8.44)$	9
$[8.44; 9.12]$	3

Таблица 1: Результаты расчетов для выборки

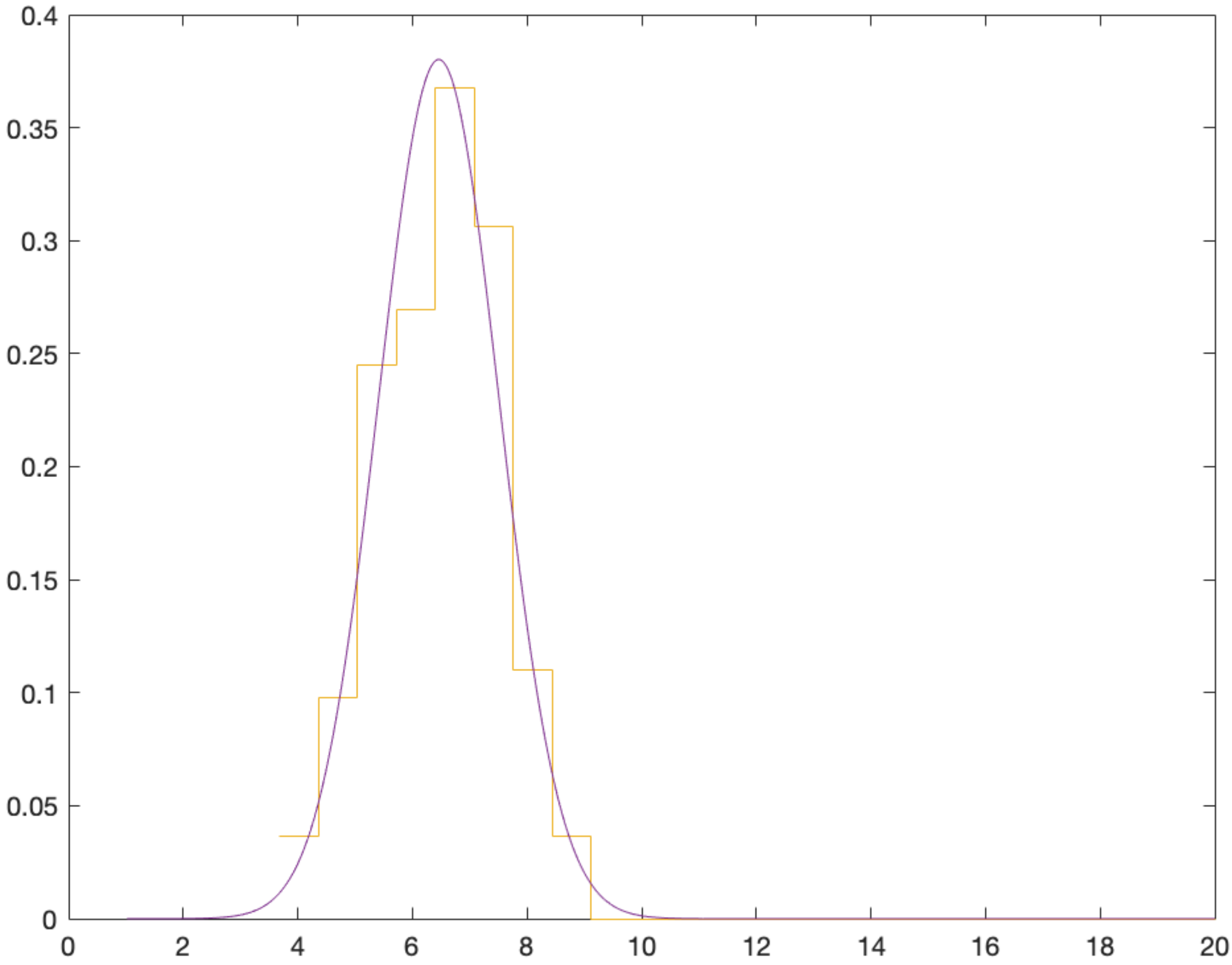

Рис. 1: Гистограмма

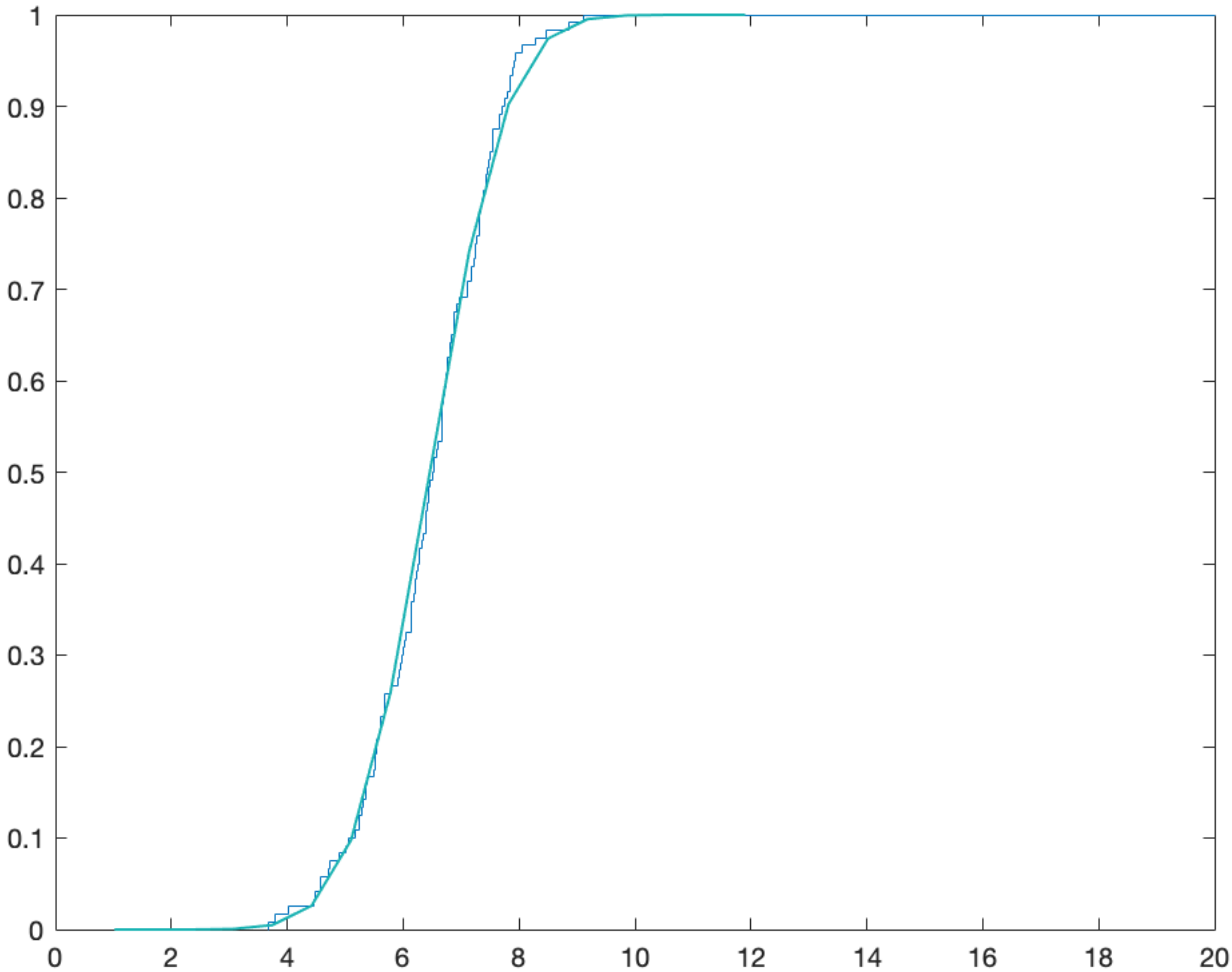

Рис. 2: Эмпирическая функция распределения