

# Оглавление

<b>Занятие 1. Выборочные характеристики</b>	<b>1</b>
Контрольные вопросы и задания . . . . .	3
Аудиторные задачи . . . . .	4
Домашнее задание . . . . .	8



# Занятие 1. Выборочные характеристики

## Контрольные вопросы и задания

**Приведите определение выборки, вариационного ряда, статистики, порядковой статистики, эмпирической функции распределения.**

$x_1, \dots, x_n$  — наблюдаемые значения — независимые одинаково распределённые случайные величины с неизвестной функцией распределения  $F(x)$ .

Такой набор случайных величин называется выборкой из распределения  $F$ .

Вариационный ряд — последовательность  $x_{(1)}, \dots, x_{(n)}$ , полученная в результате расположения в порядке неубывания исходной последовательности независимых одинаково распределённых случайных величин  $x_1, \dots, x_n$ .

Статистикой называют функцию  $S$  от выборки  $X = (x_1, x_2, \dots, x_n)$  такую, что  $S(X) = S(x_1, x_2, \dots, x_n)$ .

Вариационный ряд и его члены являются порядковыми статистиками.

Эмпирической (выборочной) функцией распределения, построенной по выборке  $x_1, \dots, x_n$  называется функция

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{x_k \leq x}, \quad x \in \mathbb{R}.$$

**Какими свойствами обладает эмпирическая функция распределения?**

Есть множество полной вероятности, на котором эмпирическая функция распределения аппроксимирует функцию распределения, то есть почти наверное  $F_n \Rightarrow F, n \rightarrow \infty$ .

Запишите выражения для выборочного среднего, выборочной дисперсии, выборочных моментов.

$$\frac{1}{n} \sum_{k=1}^n x_k$$

— выборочное среднее.

Выборочная дисперсия

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Выборочные моменты в математической статистике — это оценка теоретических моментов распределения на основе выборки.

Выборочный момент порядка  $k$  — это случайная величина

$$a_n(k) = \frac{1}{n} \sum_{i=1}^n x_i^k.$$

## Аудиторные задачи

### 1.4

*Задание.* Пусть  $X_1, \dots, X_n$  — выборка из равномерного распределения на отрезке  $[0, \theta]$  с неизвестным параметром  $\theta$ . Какие из приведённых ниже функций являются статистиками?

- a)  $\bar{X}$ ;
- b)  $5X_{(n)}$ ;
- c)  $\theta/2$ ;
- d)  $X_1/\theta$ ;
- e)  $X_{(1)} + X_1 + X_n$ .

*Решение.*

- a) Да;
- b) да;
- c) нет, так как не функция от выборки;
- d) функция не только от выборки (зависит от неизвестного параметра). Отсюда следует, что это не статистика;
- e) да.

### 1.5

*Задание.* Пусть  $X_1, \dots, X_n$  — выборка из распределения Пуассона с параметром  $\lambda$ . Вычислите математическое ожидание и дисперсию статистики

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Выясните, имеет ли статистика  $\bar{X}$  распределение Пуассона.

*Решение.* Все  $X_i$  одинаково распределены. Отсюда следует, что все математические ожидания одинаковы

$$M\bar{X} = \frac{1}{n} \sum_{i=1}^n MX_i = \frac{1}{n} \cdot nMX_1 = MX_1 = \lambda.$$

Для всякой выборки справедливо  $M\bar{X} = MX_1$ .

Из независимости  $X_i$  получаем

$$D\bar{X} = \frac{1}{n^2} \sum_{i=1}^n DX_i.$$

Так как  $X_i$  одинаково распределены, то все дисперсии одинаковы

$$D\bar{X} = \frac{DX_1}{n} = \frac{\lambda}{n}.$$

Математическое ожидание и дисперсия для распределения Пуассона совпадают. Отсюда следует, что эта случайная величина не имеет распределения Пуассона.

$\bar{X}$  не обязательно будет принимать целые значения.

### 1.6

*Задание.* Вычислите математическое ожидание статистик:

a)  $S^2 = \overline{X^2} - (\bar{X})^2$ ;

b)  $S_0^2 = 1/(n-1) \cdot \sum_{i=1}^n (X_i - \bar{X})^2$ .

*Решение.*

a) Распишем каждую из величин

$$S^2 = \overline{X^2} - (\bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2.$$

Распишем квадрат

$$\frac{1}{n} \sum_{i=1}^n X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n X_i^2 + 2 \sum_{i,j=1, i < j}^n X_i X_j \right).$$

Берём слева и справа математическое ожидание. Из того, что случайные величины в выборке одинаково распределены

$$MS^2 = MX_1^2 - \frac{1}{n^2} [nMX_1 + 2C_n^2 (MX_1)^2].$$

Подставляем  $C_n^2$  и группируем

$$MX_1^2 - \frac{1}{n^2} [nMX_1 + 2C_n^2 (MX_1)^2] = \frac{n-1}{n} \cdot MX_1^2 - \frac{n-1}{n} (MX_1)^2.$$

Вынесем общий множитель за скобки

$$\frac{n-1}{n} \cdot MX_1^2 - \frac{n-1}{n} (MX_1)^2 = \frac{n-1}{n} [MX_1^2 - (MX_1)^2] = \frac{n-1}{n} \cdot SX_1.$$

Эта оценка смещена асимптотически;

b) выразим  $S_0$  через  $S$ . Раскроем квадрат

$$S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2).$$

Имеем сумму  $n$  одинаковых слагаемых

$$S_0^2 = \frac{1}{n-1} (n\bar{X}^2 - 2(\bar{X})^2 n + n\bar{X}^2) = \frac{n}{n-1} [\bar{X}^2 - (\bar{X})^2] = \frac{n-1}{n} \cdot S^2.$$

Отсюда следует, что

$$MS_0^2 = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot DX_1 = DX_1.$$

## 1.7

*Задание.* Найдите в терминах функции распределения  $F$  выборки  $X_1, \dots, X_n$ :

- a) распределение  $k$ -ой порядковой статистики  $X_{(k)}$ ;
- b) вероятность  $P(X_{(k)} < y, X_{(k+1)} \geq y)$ .

*Решение.*

- a) Сделали упорядочивание случайных величин

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)} \leq \dots \leq X_{(n)}.$$

По определению  $F_{X_{(k)}}(y) = P(X_{(k)} \leq y) = P\{\text{хотя бы } k \text{ элементов выборки не превышает } y\} =$

$$= \sum_{i=k}^n P(A_i),$$

где  $A_i = \{\text{ровно } i \text{ элементов выборки не превышают } y\}$ .

Есть  $n$  испытаний, успех —  $X_i \leq y$ .

Вероятность успеха — это  $F(y)$ , вероятность неудачи — это  $[1 - F(y)]$ .

Это биномиальное распределение

$$F_{X_{(k)}} = \sum_{i=k}^n C_n^i F^i(y) [1 - F(y)]^{n-i};$$

б) согласно с предыдущим пунктом  $P(X_{(k)} < y, X_{(k+1)} \geq y) = P\{\text{ровно } i \text{ элементов выборки не превышает } y\} = C_n^k F^k(y) [1 - F(y)]^{n-k}$ .

## 1.8

*Задание.* Пусть  $(-0.8; 2.9; 4.5; -5.7; 1.1; -3.2)$  — наблюдаемые значения выборки. Составьте вариационный ряд, постройте эмпирическую функцию распределения  $F_6(x)$  и её график. Вычислите выборочное среднее и выборочную дисперсию.

*Решение.* Вариационный ряд:  $(-5.7; -3.2; -0.8; 1.1; 2.9; 4.3)$ .

Эмпирическая функция распределения (рис. 1).

$$F_6(y) = \frac{1}{6} \sum_{i=1}^6 \mathbb{1}\{x_i \leq y\} = \begin{cases} 0, & x < -5.7; \\ \frac{1}{6}, & -5.7 \leq x < -3.2, \\ \frac{2}{6}, & -3.2 \leq x < -0.8, \\ \frac{3}{6}, & -0.8 \leq x < 1.1, \\ \frac{4}{6}, & 1.1 \leq x < 2.9, \\ \frac{5}{6}, & 2.9 \leq x < 4.3, \\ 1, & x \geq 4.3. \end{cases}$$

Выборочное среднее

$$\bar{X} = \frac{1}{6} (-5.7 - 3.2 - 0.8 + 1.1 + 2.9 + 4.3) = \frac{1}{6} (-9.7 + 8.3) = -\frac{1}{6} \cdot 1.4 = -0.23.$$

Выборочная дисперсия

$$\hat{\sigma}^2 = \frac{1}{5} \sum_{i=1}^6 (X_i + 0.23)^2.$$

Она является несмещённой.

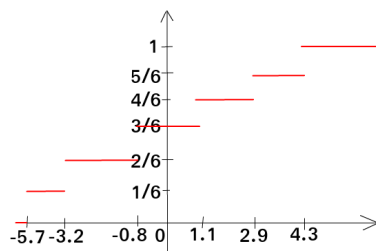


Рис. 1: Эмпирическая функция распределения

### 1.9

*Задание.* Вычислите вероятность  $P(F_n(y) < F_n(z))$ .

*Решение.*

а)  $y \geq z$ . Событие невозможное, потому что  $F_n(y) \geq F_n(z)$ ;

б) рассмотрим случай, когда  $y < z$ .

Тогда искомая вероятность равна  $P(F_n(y) < F_n(z)) = P\{\text{в } (y, z) \text{ попал хотя бы 1 элемент выборки}\} = 1 - P\{\text{в } (y, z) \text{ ни один элемент выборки не попал}\}$ . Случайные величины одинаково распределены, поэтому  $1 - P\{\text{в } (y, z) \text{ ни один элемент выборки не попал}\} =$   
 $= [1 - P\{x_i \notin (y, z)\}]^n = 1 - [1 - P\{x_1 \in (y, z)\}]^n =$   
 $= 1 - [1 - P(x_1 < z) + P(x_1 < y)]^n = 1 - [1 - F(z) + F(y)]^n.$

## Домашнее задание