

Содержание

1	Предельные теоремы и законы больших чисел	2
2	Вариационные ряды и их характеристики	2
2.1	Генеральная и выборочная совокупности, объём выборки.	2
2.2	Вариационный ряд, варианта, частота. Виды вариационных рядов. Гистограмма, полигон.	2
2.3	Формулы числовых характеристик. Эмперическая функция распределения (ЭФР). Свойства ЭФР.	4
2.4	Эмперическая функция распределения ЭФР. Свойства ЭФР. . .	5
3	Оценки параметров распределения	6
3.1	Понятия статистики, оценки, выборочной характеристики. . . .	6
3.2	Несмещённые, состоятельные и эффективные оценки	6
3.3	Теорема о единственности эффективной оценки	7

1 Предельные теоремы и законы больших чисел

2 Вариационные ряды и их характеристики

2.1 Генеральная и выборочная совокупности, объём выборки.

Рассмотрим постановку задачи математической статистики: по результатам наблюдения за некоторой случайной величиной ξ требуется сделать выводы о неизвестном законе распределения этой величины $\mathcal{L}(x, \Theta)$ либо о неизвестных параметрах $\Theta_1, \dots, \Theta_n$ известного распределения.

Пусть ξ — случайная величина с некоторой (теоретической) функцией распределения $F_\xi(x) = P\{\xi < x\}$, $x \in R$.

Определение:

Совокупность n независимых одинаково распределённых случайных величин X_1, X_2, \dots, X_n называется выборкой (выборочной совокупностью), извлечённой из распределения случайной величины ξ .

Определение:

Под генеральной совокупностью понимается множество всех возможных значений случайной величины ξ .

Определение:

Объёмом совокупности называется количество всех её элементов, объём выборки или выборочной совокупности обозначается n , генеральной совокупности — N .

2.2 Вариационный ряд, варианта, частота. Виды вариационных рядов. Гистограмма, полигон.

Определение:

Пусть x_1, x_2, \dots, x_n — выборка из генеральной совокупности значений.

Вариационным рядом называется последовательность $x_1^*, x_2^*, \dots, x_n^*$ элементов выборки расположенных в порядке неубывания, т.е. $x_1^* \leq x_2^* \leq \dots \leq x_n^*$.

x_i^* — варианта.

n_i — частота появления варианты x_i^* в выборке.

Определение:

Точечным вариационным рядом называется:

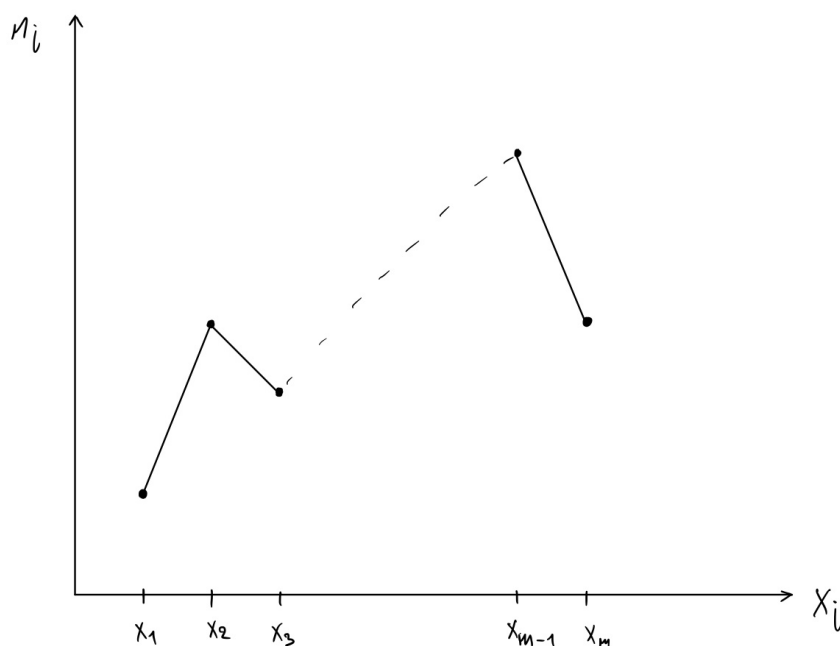
x_i	x_1	x_2	\dots	x_m
n_i	n_1	n_2	\dots	n_m

x_i — варианты, n_i — частота соответствующей варианты.

m — количество групп (различных вариантов (вариант в таблице)).

$n = \sum_{i=1}^m n_i$, где n_i — объём выборки.

Для графического представления точечных вариационных рядов используется полигон частот — ломанная с вершинами в точках (x_i, n_i) .



Определение:

Интервальным вариационным рядом называется:

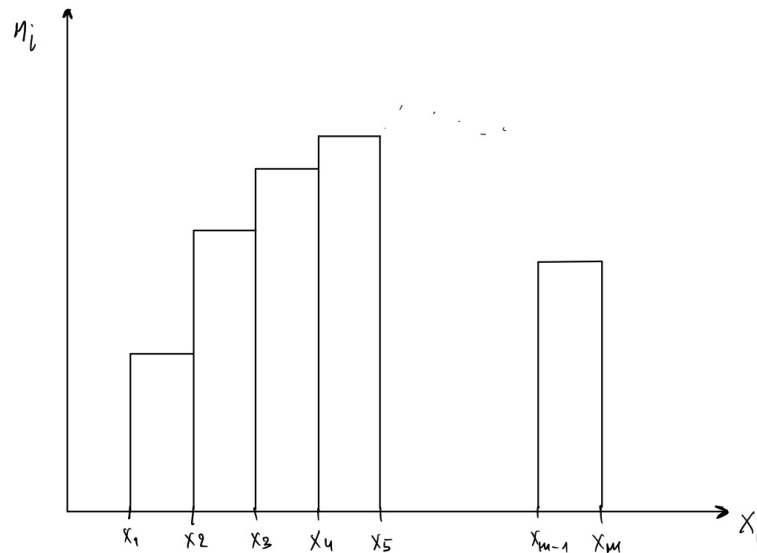
x_i	$[x_1, x_2]$	$(x_2, x_3]$	\dots	$(x_m, x_{m+1}]$
n_i	n_1	n_2	\dots	n_m

x_i — варианты, n_i — частота.

m — количество групп (интервалов).

$n = \sum_{i=1}^m n_i$, где n_i — объём выборки.

Для графического представления интервальных вариационных рядов используется гистограмма частот — фигура, составленная из прямоугольников, одной стороной которых служат интервалы $(x_i, x_{i+1}]$, а длина второй равна n_i .



2.3 Формулы числовых характеристик. Эмперическая функция распределения (ЭФР). Свойства ЭФР.

Определение:

Выборочным средним называется величина:

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

Если данные представлены в виде точечного или интервального вариационного ряда, то для вычисления используют формулу:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i * n_i,$$

где m — количество групп в точечном или интервалов в интервальном вариационном ряду, n_i — частота, т.е. количество элементов выборки, принадлежащих i -той группе или i -тому интервалу, x_i — варианта для точечного ряда и середина i -того интервала для интервального ряда.

Определение:

Выборочной дисперсией (смещённой) называется величина:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2.$$

Она характеризует среднее из квадратов отклонений наблюдаемой величины от выборочного среднего. Величина $S = \sqrt{S^2}$ называется выборочным средним

квадратическим отклонением (смещённым) величин выборки от выборочного среднего.

Если данные представлены в виде точечного или интервального вариационного ряда, то для вычисления используют формулу:

$$S^2 = \frac{1}{n} \sum_{i=1}^m (x_i - \bar{x})^2 n_i.$$

или

$$S^2 = \left(\frac{1}{n} \sum_{i=1}^m x_i^2 n_i \right) - \bar{x}^2.$$

Здесь m — количество групп в точечном или интервалах в интервальном вариационных рядах, n_i — частота. т.е. количество элементов выборки, принадлежащих i -той группе или i -тому интервалу, x_i — варианта для точечного ряда и середина i -того интервала для интервального ряда.

Определение:

Выборочной дисперсией (несмещённой) называется величина:

$$\bar{\sigma} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})^2.$$

Аналогично, величина $\bar{\sigma} = \sqrt{\bar{\sigma}^2}$ называется выборочным несмещённым средним квадратическим отклонением.

Очевидно, что смещённая и несмещённая выборочные дисперсии связаны формулой:

$$\bar{\sigma}^2 = \frac{n}{n-1} S^2.$$

2.4 Эмпирическая функция распределения ЭФР. Свойства ЭФР.

Эмпирическая функция распределения ЭФР

Эмпирической функцией распределения (ЭФР) называется функция

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n e(x - X_i),$$

где $e(x) = 1$, при $x > 0$ $e(x) = 0$, при $x \leq 0$.

Таким образом, если X_i , то $e(x) = 1$, если $X_i \geq x$, то $e(x) = 0$, а сумма $e(x - X_i)$ будет равна количеству элементов выборки, которые приняли значение, строго меньше некоторого $x \in R$.

Пусть x_1, x_2, \dots, x_n — реализация выборки X_1, X_2, \dots, X_n , т.е. наблюдавшиеся значения случайной величины ξ .

Обозначим $\mu(x)$ — число элементов выборки, строго меньших $x \in R$. Тогда эмпирическая функция распределения $\tilde{F}_n(x)$ может быть определена как

$$\tilde{F}_n(x) = \frac{\mu(x)}{n}.$$

Свойства ЭФР:

1. $0 \leq \tilde{F}_n(x) \leq 1$, т.к. $0 \leq \mu(x) \leq n$;
2. неубывающая непрерывная слева функция;
3. $\tilde{F}_n(x)$ — ступенчатая функция для всех типов распределений;
4. $\tilde{F}_n(x)$ сходится по распределению к $F_\xi(x)$.

3 Оценки параметров распределения

3.1 Понятия статистики, оценки, выборочной характеристики.

Определение:

Пусть $g(t_1, \dots, t_n)$ — непрерывная функция. Оценкой Θ назовём $\tilde{\Theta} = g(X_1, \dots, X_n)$. Если $g(X_1, \dots, X_n) = T$ некоторая функция, то T — статистика.

Определение:

Выборочными характеристиками называются функции от наблюдений (точечные оценки), приближённо оценивающие соответствующие числовые характеристики случайной величины.

3.2 Несмещённые, состоятельные и эффективные оценки

Определение:

Оценка $\tilde{\Theta}_n$ называется несмещённой оценкой параметра Θ , если $M(\tilde{\Theta}_n) = \Theta$.

Определение:

Оценка $\tilde{\Theta}_n$ называется состоятельной оценкой параметра Θ , если $\tilde{\Theta}_n$ сходится по вероятности к Θ .

Определение:

Оценка $\tilde{\Theta}_n$ называется эффективной, или оптимальной, или наилучшей несмещённой оценкой с минимальной дисперсией (НОМД), если $M(\tilde{\Theta}_n) = \Theta$ и $D(\tilde{\Theta}_n) = \inf_{\tilde{\Theta}_n^*} D\tilde{\Theta}_n^*$.

3.3 Теорема о несмещённой состоятельной оценке мат. ожидания

Пусть $\xi \sim L(x, \Theta)$, $L(x, \Theta)$ — закон распределения известен с точностью до параметра.

$$\bar{\Theta} = (\Theta_1, \dots, \Theta_n)$$

По результатам $\mathbf{X}_1, \dots, \mathbf{X}_n$ наблюдений за ξ требуется построить оценку Θ .

Теорема:

Пусть $\mathbf{X}_1, \dots, \mathbf{X}_n \sim L_\xi(x, \Theta)$, где ξ — случайная величина с $M\xi = a < +\infty$ $D\xi = \sigma^2$. Тогда выборочное среднее $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ является несмещённой и состоятельной оценкой $M\xi$.

Доказательство:

$$\begin{aligned} M\bar{x} &= M\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n Mx_i = \\ &= |x_i \text{ — н. о. р. случайной величины} \Rightarrow Mx_i = M\xi \quad \forall i| = \frac{1}{n} \sum_{i=1}^n M\xi = \\ &= \frac{1}{n} * n * a = a \Rightarrow \bar{x} \text{ — несмещённое} \end{aligned}$$

Состоятельность $\forall \epsilon > 0 \quad P\{|\bar{x} - a| < \epsilon\} \rightarrow_{n \rightarrow \infty} 1$

По неравенству Чебышёва:

$$\begin{aligned} \forall \epsilon > 0 \quad P\{|\bar{x} - a| < \epsilon\} &\geq 1 - \frac{D\bar{x}}{\epsilon^2} = 1 - \frac{1}{\epsilon^2} D\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \\ &= | \text{н. о. р.} \quad Dx_i = \sigma^2 | = \\ &= 1 - \frac{1}{n^2} \epsilon^2 D\xi = 1 - \frac{n\sigma^2}{n^2\epsilon^2} = 1 - \frac{\sigma^2}{n\epsilon^2} \rightarrow_{n \rightarrow \infty} 1 - 0 \end{aligned}$$

Т.к. $P(A) \leq 1 \quad \forall A$, то $\lim_{n \rightarrow \infty} P\{|\bar{x} - a| < \epsilon\} = 1 \Rightarrow \bar{x}$ — состоятельная.

