



*«Московский государственный технический
университет имени Н.Э. Баумана»
(МГТУ им. Н.Э. Баумана)*

Факультет: Информатика и системы управления

Кафедра: ИУ7

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Студент группы ИУ7-63,
Степанов Александр Олегович

Преподаватель:
Власов Павел Александрович

2020 г.

Оглавление

1	Предельные теоремы теории вероятности	2
1.1	Неравенства Чебышева	2
1.2	Виды сходимости последовательности случайных величин	3
1.3	Закон больших чисел	4
1.4	Центральная предельная теорема	8
2	Математическая статистика	13
2.1	Основные понятия выборочной теории	13
2.1.1	Основные определения	13
2.1.2	Предварительная обработка результатов эксперимента	16
2.2	Точечные оценки	23
2.2.1	Основные понятия	23
2.2.2	Несмещенность точечной оценки	24
2.2.3	Состоятельность точечной оценки	26
2.2.4	Эффективность точечной оценки	27

1 | Пределные теоремы теории вероятности

1.1 Неравенства Чебышева

теорема: 1-е неравенство Чебышева

Пусть

1. X - случайная величина
2. $X \geq 0$ (т.е. $P\{X < 0\} = 0$)
3. $\exists MX$

Тогда

$$\forall \varepsilon > 0 P\{X \geq \varepsilon\} \leq \frac{MX}{\varepsilon}$$

Доказательство:

$$\begin{aligned} MX &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} x f_X(x) dx = \text{свойство аддитивности} = \\ &= \int_0^{\varepsilon} x f_X(x) dx + \int_{\varepsilon}^{+\infty} x f_X(x) dx \geq \int_{\varepsilon}^{+\infty} x dF_X(x) \geq \varepsilon \int_{\varepsilon}^{+\infty} f_X(x) dx = \end{aligned}$$

$$= \varepsilon P\{X \geq \varepsilon\} \Rightarrow P\{X \geq \varepsilon\} \leq \frac{MX}{\varepsilon}$$

Теорема: 2-е неравенство Чебышева

Пусть

1. X - случайная величина

2. $\exists MX, \exists DX$

Тогда

$$\forall \varepsilon > 0 P\{|X - MX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}$$

Доказательство

$$DX = M[(X - MX)^2] = \left| \begin{array}{l} 1) Y = (X - MX)^2 \geq 0 \\ 2) \exists DX \Rightarrow \exists MX \Rightarrow \delta = \varepsilon^2 > 0 \\ 3) M[Y] \geq \delta P\{Y \geq 0\} \end{array} \right| \geq \varepsilon^2 \cdot P\{Y \geq \varepsilon^2\} =$$

$$= \varepsilon^2 P\{(X - MX)^2 \geq \varepsilon^2\} = \varepsilon^2 P\{|X - MX| \geq \varepsilon\} \Rightarrow$$

$$\Rightarrow P\{|X - MX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}$$

1.2 Виды сходимости последовательности случайных величин

Пусть X_1, X_2, \dots - последовательность случайных величин, заданных на одном вероятностном пространстве

Определение: Говорят, что последовательность X_1, X_2, \dots сходится по вероятности к случайной величине Z , заданной на том же вероятностном пространстве, если

$$\forall \varepsilon > 0 P\{|X_n - Z| \geq \varepsilon\} \rightarrow_{n \rightarrow \infty} 0$$

Обозначается

$$X_n \xrightarrow[n \rightarrow \infty]{P} Z$$

Замечание: Сходимость последовательности $X_n, n \in \mathcal{N}$, к случайной величине Z по вероятности означает, что при достаточно больших n вероятность отклонения случайной величины X_n от случайной величины Z будет меньше (любого) наперед заданного числа

Определение: Говорят, что последовательность X_1, X_2, \dots слабо сходится к случайной величине Z , если в каждой точке x непрерывности функции распределения $F_Z(x)$ числовая последовательность значений $F_{X_n}(x)$ сходится к $F_Z(x)$, то есть

$$(\forall x \in \mathcal{R})((F_Z \text{ непрерывная в точке } x) \Rightarrow (F_{X_n}(x) \xrightarrow[n \rightarrow \infty]{} F_Z(x)))$$

Обозначается

$$X_n \Rightarrow_{n \rightarrow \infty} Z$$

1.3 Закон больших чисел

Определение: Говорят, что последовательность случайных величин X_1, X_2, \dots удовлетворяет закону больших чисел, если

$$\forall \varepsilon > 0 P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n m_i\right| \geq \varepsilon\right\} \xrightarrow[n \rightarrow \infty]{} 0$$

Замечание:

1. Рассмотрим случайную величину $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

Выполнение для последовательности X_1, X_2, \dots ЗБЧ означает, что при достаточно больших n случайная величина \overline{X}_n практически теряет случайный

характер и принимает неслучайное значение $\frac{1}{n} \sum_{i=1}^n m_i$ с вероятностью, близкой к 1

2. Выполнение для последовательности X_1, X_2, \dots ЗБЧ означает, что

$$Y_n = \frac{1}{n} \sum_{i=1}^n (X_i - m_i) \xrightarrow{P}_{n \rightarrow \infty} 0$$

Вопрос: Какими свойствами должна обладать последовательность X_1, X_2, \dots , чтобы она удовлетворяла ЗБЧ?

Теорема: Чебышева (ЗБЧ в форме Чебышева)

Пусть

1. X_1, X_2, \dots – последовательность независимых случайных величин
2. $EX_i = m_i, EX_i^2 = \sigma_i^2, i \in \mathcal{N}$
3. Дисперсии ограничены в сов-ти, то есть

$$\exists c > 0 \sigma_i^2 \leq c, i \in \mathcal{N}$$

Тогда последовательность X_1, X_2, \dots удовлетворяет ЗБЧ

Доказательство:

1. Рассмотрим случайную величину

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, n \in \mathcal{N}$$

Тогда

$$M[\overline{X}_n] = M\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n MX_i = \frac{1}{n} \sum_{i=1}^n m_i$$

$$D[\overline{X}_n] = D\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} D\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n DX_i = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2$$

2. Запишем для случайной величины $\overline{X_n}$ 2-е неравенство Чебышева

$$\forall \varepsilon > 0 P\{|\overline{X_n} - M\overline{X_n}| \geq \varepsilon\} \leq \frac{D\overline{X_n}}{\varepsilon^2}$$

То есть

$$\forall \varepsilon > 0 P\{|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n m_i| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2$$

Так как по условию дисперсии ограничены в сов-ти, то

$$\sum_{i=1}^n \sigma_i^2 \leq \sum_{i=1}^n C \leq C \cdot n$$

поэтому

$$\forall \varepsilon > 0 0 \leq P\{|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n m_i| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} \frac{C \cdot n}{n^2}$$

Устремим $n \rightarrow \infty$

По теореме о 2х милиционерах

$$\forall \varepsilon > 0 P\{|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n m_i| \geq \varepsilon\} \xrightarrow{n \rightarrow \infty} 0$$

Замечание: Теорема Чебышева носит достаточный характер, то есть если для некоторой последовательности случайных величин выполнены ее условия, то последовательность удовлетворяет ЗБЧ. В этом случае также говорят, что последовательность удовлетворяет ЗБЧ в форме Чебышева. Если для последовательности не выполнены условия теоремы Чебышева, то она не удовлетворяет ЗБЧ в форме Чебышева, но, возможно, удовлетворяет ЗБЧ.

Следствие:

Пусть

1. X_1, X_2, \dots – последовательность независимых одинаково распределенных случайных величин
2. $m = MX_i, i \in \mathcal{N}$

Тогда

$$\forall \varepsilon > 0 P\left\{\left|\frac{1}{n} \sum X_i - m\right| \geq \varepsilon\right\} \xrightarrow{n \rightarrow \infty} 0$$

Доказательство: Заметим, что $m_i \equiv m, i \in \mathcal{N}$, поэтому $\frac{1}{n} \sum m_i = m$, и исп. m теоремы Чебышева

Следствие: ЗБЧ в форме Бернулли

Пусть

1. Проводится n испытаний по схеме Бернулли с вероятностью успеха p
2. Наблюденная частота успеха

$$r_n = \frac{\left\{\text{число успехов в серии } n \text{ испытаний}\right\}}{n}, n \in \mathcal{N}$$

Тогда

$$r_n \xrightarrow[n \rightarrow \infty]{P} p$$

Доказательство

1. Пусть

$$X_i = \begin{cases} 1, & \text{если в } i\text{-м испытании серии произошёл успех} \\ 0, & \text{иначе} \end{cases}$$

X_i	0	1
P	q	p

где $q = 1 - p$ – вероятность неудачи

Тогда $\sum_{i=1}^n X_i$ – число успехов в серии из n испытаний

$$\frac{1}{n} \sum_{i=1}^n X_i = r_n$$

2. Применим к последовательности X_1, X_2, \dots ЗБЧ в форме Чебышева

(a) X_1, X_2, \dots - независимы (определение схемы Бернулли)

(b) $MX_i = p, DX_i = q, i \in N$

(c) Так как $DX_i = pq$, то дисперсии определены в совокупности

Применяя к последовательности следствие 1 получаем:

$$\forall \varepsilon > 0 \quad P\left\{\left|\underbrace{\frac{1}{n} \sum X_i}_{r_n} - \underbrace{m}_p\right| \geq \varepsilon\right\} \xrightarrow{n \rightarrow \infty} 0$$

Таким образом

$$r_n \xrightarrow[n \rightarrow \infty]{P} p$$

1.4 Центральная предельная теорема

Рассмотрим последовательность X_1, X_2, \dots случайных величин, обладающих следующими свойствами:

1. $X_i, i \in \mathcal{N}$ – независимы

2. $X_i, i \in \mathcal{N}$ – одинаково распределены

3. $\exists MX_i = m, \exists DX = \sigma_i^2, i \in \mathcal{N}$

Обозначим:

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$$

Тогда (смотреть расс-я из доказательства теоремы Чебышева)

$$M\overline{X_n} = m$$

$$D\overline{X_m} = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Рассмотрим случайную величину

$$Y_n = \frac{\overline{X_n} - M\overline{X_n}}{\sqrt{D[\overline{X_n}]} } = \frac{\overline{X_n} - m}{\sigma/\sqrt{n}}, n \in \mathcal{N}$$

$$MY_n = 0, DY_n = 1$$

Теорема ЦПТ

Пусть выполнены условия (1) – (3). Тогда последовательность Y_n слабо сходится к случайной величине $Z \sim N(0, 1)$

То есть

$$(\forall x \in \mathcal{R})(F_{Y_n}(x) \rightarrow_{n \rightarrow \infty} \Phi(x))$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt - \text{функция распределения случайной величины } Z$$

Доказательство: без доказательства

Замечание:

1. Следствие 1 ЗБЧ утверждает, что если выполнены (1) – (3), то последовательность $\overline{X_n} \xrightarrow{P_{n \rightarrow \infty}} m$ ЦПТ уточняет характер этой сходимости

Пример: ЭВМ производит суммирование $n = 10^4$ чисел, каждое из которых округлено с точностью до 10^{-4} . Считая, что ошибки округления независимы и равномерно распределены на $(-0.5 \cdot 10^{-4}, 0.5 \cdot 10^{-4})$. Найти диапазон, в котором с вероятностью 0.95 будет заключена ошибка всех сумм.

(а) Пусть $X_i, i = \overline{1; n}$ - ошибка округления i -го числа

- i. все X_i независимы (по условию)
- ii. $X_i \sim R(-0.5 \cdot 10^{-4}, 0.5 \cdot 10^{-4}) \Rightarrow$ одинаково распределены
- iii.

$$MX_i = \frac{a+b}{2} = 0$$

$$FX_i = \frac{(b-a)^2}{12} = \frac{(10^{-4})^2}{12} \approx 10^{-9}$$

Таким образом последовательность X_i удовлетворяет ЦПТ

(b) $S = \sum_{i=1}^m X_i$ – ошибка всей суммы

$$Y_n = \frac{\frac{1}{n} \sum_{i=1}^n X_i - MX_i}{\sqrt{DX_i/n}} \sim N(0, 1)$$

(c) Изображение с фото

Таким образом $\varepsilon = U_{0.975}$ – квантиль уровня 0.975 распределения $N(0, 1)$ из таблицы $U_{0.975} = 1.96$

Таким образом

$$0.95 = P\{|Y_n| < 1.96\} = P\left\{\left|\frac{\frac{1}{n}S - 0}{\sqrt{DX_i/n}}\right| < 1.96\right\} = P\left\{\left|\frac{S}{\sqrt{DX_i}}\right| < 1.96\right\} =$$

$$= P\{|S| < 1.96 \cdot \sqrt{nDX_i}\} = P\{|S| < 1.96\sqrt{10^4 \cdot 10^{-9}}\} = P\{|S| < 6.2 \cdot 10^{-3}\}$$

Ответ: с вероятностью 0.95

Следствие. Интегральная Теорема Мдавра-Лапласа

Пусть:

1. проводится $n \gg 1$ испытаний по схеме испытаний Бернулли с вероятностью успеха p
2. k – число успехов в этой серии

Тогда

$$P\{k_1 \leq k \leq k_2\} \approx \Phi_0(x_2) - \Phi_0(x_1)$$

,где

$$x_i = \frac{k_i - np}{\sqrt{npq}}, i = \overline{1; 2}$$

$$\Phi_0(x) = \frac{1}{2\pi} \int_0^x e^{-\frac{t^2}{2}} dt, q = 1 - p$$

Доказательство

1. Рассмотрим случайную величину

$$X_i = \begin{cases} 1, & \text{если в } i\text{-м испытании произошел успех} \\ 0, & \text{иначе} \end{cases}$$

Тогда

X_i	0	1
P	q	p

$$\left. \begin{aligned} MX_i &= p, \\ DX_i &= pq \end{aligned} \right\} (*)$$

2. $k = \sum_{i=1}^n X_i$ ФОТО

Пример 1.4.1 В эксперименте Пирсона по подбрасыванию правильной монеты после $n = 24000$ испытаний герб выпал 12012 раз. Какова вероятность того, что при повторном испытании отклонение относительной частоты успеха будет таким же или больше

Решение

1. Имеем схему испытаний Бернулли

$$n = 24000 \gg 1$$

$$p = q = \frac{1}{2}$$

Пусть

$$A = \{\text{отклонение не меньше}\}$$

Тогда

$$\overline{A} = \{\} = \{|k - 12000| < 12\}$$

$$P(\overline{A}) = P\{119888 < k < 12012\} = \Phi_0(x_2) - \Phi_0(x_1) \approx 2\Phi_0(0.155) \approx 0.123$$

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} \approx 0.155$$

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = -0.155$$

$$P(A) = 1 - P(\overline{A}) \approx 0.877$$

2 | Математическая статистика

2.1 Основные понятия выборочной теории

2.1.1 Основные определения

Теория вероятностей – является разделом «чистой» математики, который строится дедуктивно, исходя из вполне определенного набора аксиом.

Математическая статистика – раздел «прикладной» математики, который строится индуктивно: от наблюдений к гипотезе; при этом аргументация основана на выводах теории вероятностей.

Типовая задача ТВ	Типовая задача МС
1) при подбрасывании монеты вероятность выпадения герба равна p . Какова вероятность того, что после n подбрасываний герб выпадет m раз?	Известно, что после n подбрасываний монеты герб выпал m раз. Чему равна вероятность выпадения герба в отдельном испытании?
2) Закон распределения случайной величины $X \sim N(m, \sigma^2)$. Какова вероятность того, что $\{X \in [a, b]\}$	После n наблюдений за реализациями случайной величины X получены значения x_1, \dots, x_n . Какой закон распределения имеет случайная величина X ?

Основная задача МС – разработка методов получения научно-обоснованных выводов о массовых процессах или явлениях по результатам наблюдений или экспериментов. При этом указанные выводы относятся не к отдельным экспериментам, а являются утверждениями о вероятностных характеристиках изучаемого яв-

ления.

Одной из «общих» задач математической статистики является следующая задача: X – случайная величина, закон распределения которой неизвестен. Требуется по результатам наблюдений «найти» закон распределения случайной величины X . У этой задачи есть две модификации:

1. **1-я основная задача МС:** закон распределения случайной величины X неизвестен вообще; требуется его «найти»
2. **2-я основная задача МС:** известен общий вид закона распределения случайной величины X , однако неизвестны значения одного или нескольких параметров этого закона. Требуется оценить значения неизвестных параметров

Пример 2.1.1 Пусть $X \sim N(t, \sigma^2)$, где значения t, σ^2 неизвестны. Требуется оценить t и σ .

Определение 2.1.1 Пусть X – случайная величина, закон распределения которой неизвестен или известен не полностью

Множество всех возможных значений случайной величины X называется **генеральной совокупностью**

Определение 2.1.2 Случайной выборкой объема n из генеральной совокупности X называется случайный вектор

$$\vec{X} = (X_1, \dots, X_n)$$

где

1. Случайные величины $X_i, i = \overline{1; n}$ независимы в совокупности
2. $X_i \sim X, i = \overline{1; n}$ (X_i имеет то же распределение, что и X)

Определение 2.1.3 Выборкой объема n из генеральной совокупности X называется (любая) реализация случайной выборки X

Замечание 2.1.1 1. На практике в результате наблюдений (экспериментов) исследователь получает реализацию \vec{x} – числовой вектор. Случайная выборка \vec{X} используется для теоретических рассуждений

2. Пусть $F(t)$ – функция распределения случайной величины X . Тогда

$$F_{\vec{X}}(t_1, \dots, t_n) = F(t_1) \cdot F(t_2) \cdot \dots \cdot F(t_n)$$

– функция распределения выборки из генеральной совокупности X

Определение 2.1.4 Множество всех возможных значений выборки \vec{X} называется выборочным пр-вом и обозначается \mathcal{X}_n

Определение 2.1.5 Любую функцию $g(\vec{X})$ случайной выборки \vec{X} называется статистикой или выборочной характеристикой. Закон распределения случайной величины $g(\vec{X})$ называется выборочным законом распределения

Замечание 2.1.2 1. Значение $g(\vec{x})$ статистики g на выборке \vec{x} называется выборочным законом этой статистики

2. Пусть $\vec{x} = (x_1, \dots, x_n)$ – реализация выборки \vec{X} . Это позволяет считать, что закон распределения случайной величины X моделируется дискретной случайной величиной, закон распределения которой задан таблицей

Значение	x_1	x_2	\dots	x_i	\dots	x_n
P	$\frac{1}{n}$	$\frac{1}{n}$	\dots	$\frac{1}{n}$	\dots	$\frac{1}{n}$

ФОТО

Определение 2.1.6 Выборочным средним (выборочным математическим ожиданием) называется статистика

$$\hat{m}(\vec{X}) = \frac{1}{n} \sum_{i=1}^n X_i$$

которая обозначается $\hat{m} = \overline{X}$

Определение 2.1.7 Выборочной дисперсией называют статистику

$$\hat{\sigma}^2(\vec{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Определение 2.1.8 Выборочным начальным моментом порядка j называют статистику

$$\hat{m}_j(\vec{X}) = \frac{1}{n} \sum_{i=1}^n X_i^j$$

Определение 2.1.9 Центральным выборочным моментом порядка j называют статистику

$$\nu_j(\vec{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^j$$

ФОТО

2.1.2 Предварительная обработка результатов эксперимента

Вариационный ряд

Пусть \vec{x} – выборка из генеральной совокупности X ,

$$\vec{x} = (x_1, \dots, x_n)$$

Расположим значения x_1, \dots, x_n в порядке неубывания

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

где $x_{(i)}$ – i -й элемент полученной последовательности

Определение 2.1.10 Последовательность значений $x_{(1)}, \dots, x_{(n)}$, удовлетворяющее (*), называют вариационным рядом выборки \vec{x} , $x_{(i)}$ называют i -м членом вариационного ряда

Определение 2.1.11 Вариационным рядом, отвечающим случайной выборке \vec{X} , называется последовательность случайных величин

$$X_{(1)}, \dots, X_{(n)},$$

где случайная величина $X_{(i)}$ для каждой реализации \vec{x} случайной выборки \vec{X} принимает значение $x_{(i)}$, равное значению i -го числа вариационного ряда выборки $\vec{x}, i = \overline{1; n}$

Замечание 2.1.3 1.

$$P\{X_{(i)} \leq X_{(i+1)}\} = 1$$

2.

$$F_{X_{(n)}}(t) = P\{X_{(n)} < t\} = P\{X_1 < t, \dots, X_n < t\} = F(t) \cdot \dots \cdot F(t) = [F(t)]^n$$

где F - функция распределения случайной величины X

3.

$$F_{X_{(1)}}(t) = P\{X_{(1)} < t\} = 1 - P\{X_{(1)} \geq t\} = 1 - P\{X_1 \geq t, \dots, X_n \geq t\} = 1 - [1 - F(t)]^n$$

Статистический ряд

Среди элементов выборки \vec{x} могут встречаться одинаковые. Предположим, что среди значений x_1, \dots, x_n выделены m попарно различных значений $z_{(1)} < \dots < z_{(m)}$, причем каждый элемент x_i совпадает с некоторым $z_{(j)}$. Эти значения группируют в таблицу, которая называется **статистическим рядом выборки \vec{x}** .

$z_{(1)}$	\dots	$z_{(i)}$	\dots	$z_{(m)}$
n_1	\dots	n_i	\dots	n_m

Здесь n_i – число элементов выборки \vec{x} , которые равны $z_{(i)}$

Замечание 2.1.4

$$\sum_{i=1}^m n_i = n$$

Определение 2.1.12 Число n_i называется частотой значения $z_{(i)}$, а величина $\frac{n_i}{n}$ – относительной частотой этого значения.

Эмпирическая функция распределения

Пусть $\vec{x} = (x_1, \dots, x_n)$ – выборка из генеральной совокупности X . Обозначим $n(x, \vec{x})$ – число элементов вектора \vec{x} , которые имеют значения меньше x .

Определение 2.1.13 Эмпирической функцией распределения называют функцию

$$F_n : \mathcal{R} \rightarrow \mathcal{R}$$

определенную условием

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

Замечание 2.1.5 1. F_n обладает всеми свойствами функции распределения

2. F_n кусочно-постоянна и скачкообразно изменяет свои значения в точке $z_{(i)}$

3. Если все элементы выборки \vec{x} попарно различны, то

$$F_n(x) = \begin{cases} 0, x \leq x_{(1)} \\ \frac{1}{n}, x_{(1)} < x \leq x_{(2)} \\ \dots \\ \frac{i}{n}, x_{(i)} < x \leq x_{(i+1)}, i = \overline{1; n-1} \\ \dots \\ 1, x > x_{(n)} \end{cases}$$

4. Эмпирическая функция распределения $F_n(x)$ позволяет интерпретировать выборку \vec{x} как реализацию дискретной случайной величины \tilde{X} , реализация которой имеет вид

\tilde{X}	$z_{(1)}$	\dots	$z_{(m)}$
P	$\frac{n_1}{n}$	\dots	$\frac{n_m}{n}$

В дальнейшем это позволит рассматривать численные характеристики случайной величины \tilde{X} как приближенные значения численных характеристик случайной величины X .

Выборочная функция распределения

Пусть $\vec{X} = (X_1, \dots, X_n)$ – случайная выборка из генеральной совокупности X .

Обозначим $n(x, \vec{X})$ – случайная величина, которая для каждой реализации \vec{x} случайной выборки \vec{X} принимает значение, равное $n(x, \vec{x})$

Определение 2.1.14 Выборочной функцией распределения называется функция

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

Замечание 2.1.6 1. При каждом фиксированном $x \in \mathcal{R}$ функция \hat{F}_n может принимать лишь значения из множества

$$0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$$

Обозначим $p = P\{X_i < x\}$ ($x \in \mathcal{R}$ некоторое фиксированное значение). Величина p не зависит от i , так как все X_i одинаково распределены. Тогда, для любого фиксированного $x \in \mathcal{R}$

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

Так как X_i не только одинаково распределены, но и независимы, то указанная вероятность равно вероятности появления ровно k успехов в серии из n испытаний по схеме Бернулли, где успех – событие $\{X < x\}$

$$C_n^k p^k q^{n-k},$$

где $p = P\{X_i < x\} = P\{X < x\}$, $q = 1 - p$. Таким образом для каждого фиксированного x случайная величина $Y = n(x, \vec{X})$ имеет распределение $Y \sim B(n, p)$

Теорема 2.1.1 Для любого $x \in \mathcal{R}$ последовательность $\hat{F}_n(x, \vec{X})$ при $x \rightarrow \infty$ сходится по вероятности к значению $F(x)$, где F - теоретическая («истинная») функция распределения случайной величины X , то есть

$$\hat{F}_n(x, \vec{X}) \xrightarrow[n \rightarrow \infty]{P} F(x)$$

Доказательство

При каждом фиксированном $x \in \mathcal{R}$ величина $\hat{F}_n(x, \vec{X})$ равна относительной (наблюдённой) частоте успеха в серии из n испытаний с вероятностью успеха $p = P\{X < x\}$ (см. замечание). Из ЗБЧ в форме Бернулли следует, что

$$\hat{F}_n(\underbrace{x}_{\text{наблюдённая частота}}, \vec{X}) \xrightarrow[n \rightarrow \infty]{P} \underbrace{p}_{\text{теоретическая частота}} = P\{X < x\} = F(x)$$

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:

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

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

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

Определение 2.1.15 Интервальным статистическим рядом называют таблицу

J_1	\dots	J_i	\dots	J_p
n_1	\dots	n_i	\dots	n_p

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

Замечание 2.1.7 1. Очевидно, что $\sum_{i=1}^p n_i = n$

2. Для выборка p – числа интервалов можно пользоваться формулой

$$p = [\log_2 n] + 1,$$

где $[a]$ – целая часть числа a

Эмпирическая плотность

Предположим, что для выборки \vec{x} построен интервальный статистический ряд

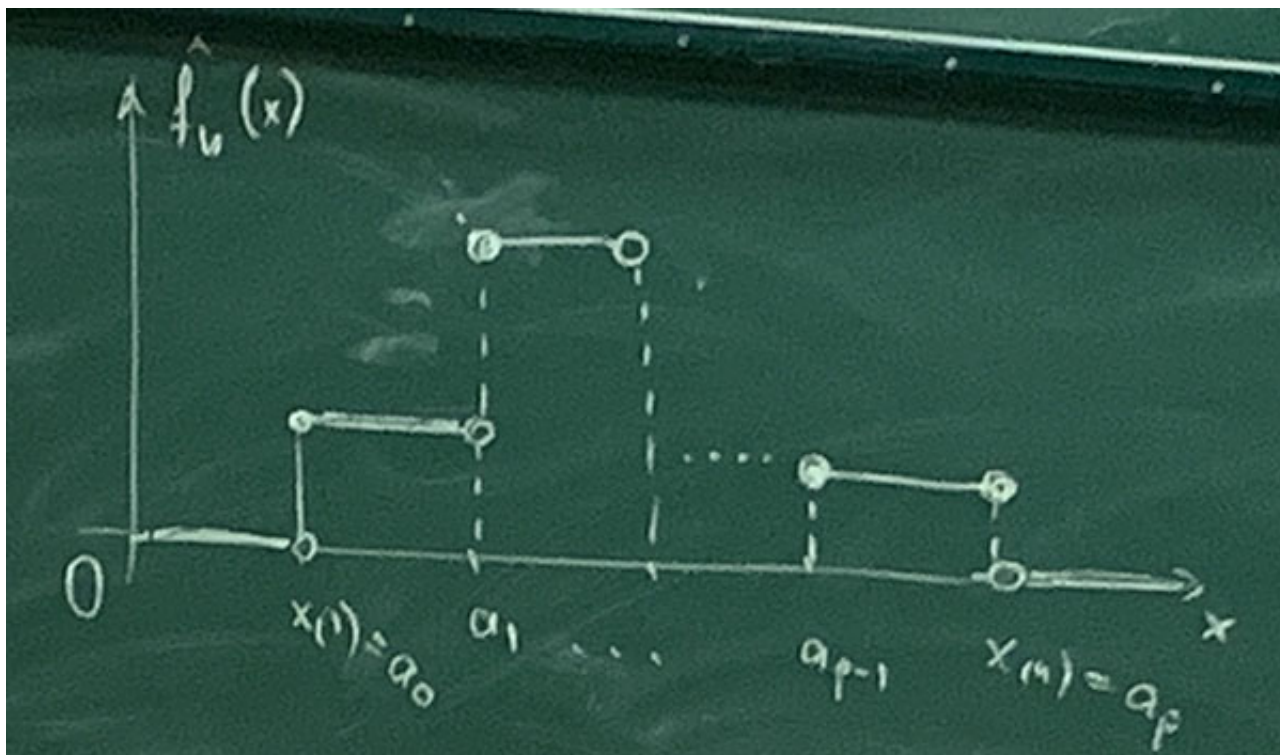
$$(J_i, n_i), i = \overline{1; p}$$

Определение 2.1.16 Эмпирической плотностью (отвечающей выборке \vec{x}) называют функцию

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

Определение 2.1.17 Гистограммой называют график эмпирической плотности

Замечание 2.1.8 1. \hat{f}_n является кусочно-постоянной функцией



2.

$$\int_{-\infty}^{+\infty} \hat{f}_n(x) dx = \sum_{i=1}^p (\text{Площадь прямоугольников}) = \frac{n_1}{n\Delta} \cdot \Delta + \dots + \frac{n_p}{n\Delta} \cdot \Delta = \dots = 1$$

3. Эмпирическая плотность является выборочным аналогом функции плотности распределения вероятностей. Аналогично свойству выборочной функции распределения мгновенно доказывается, что для большой n

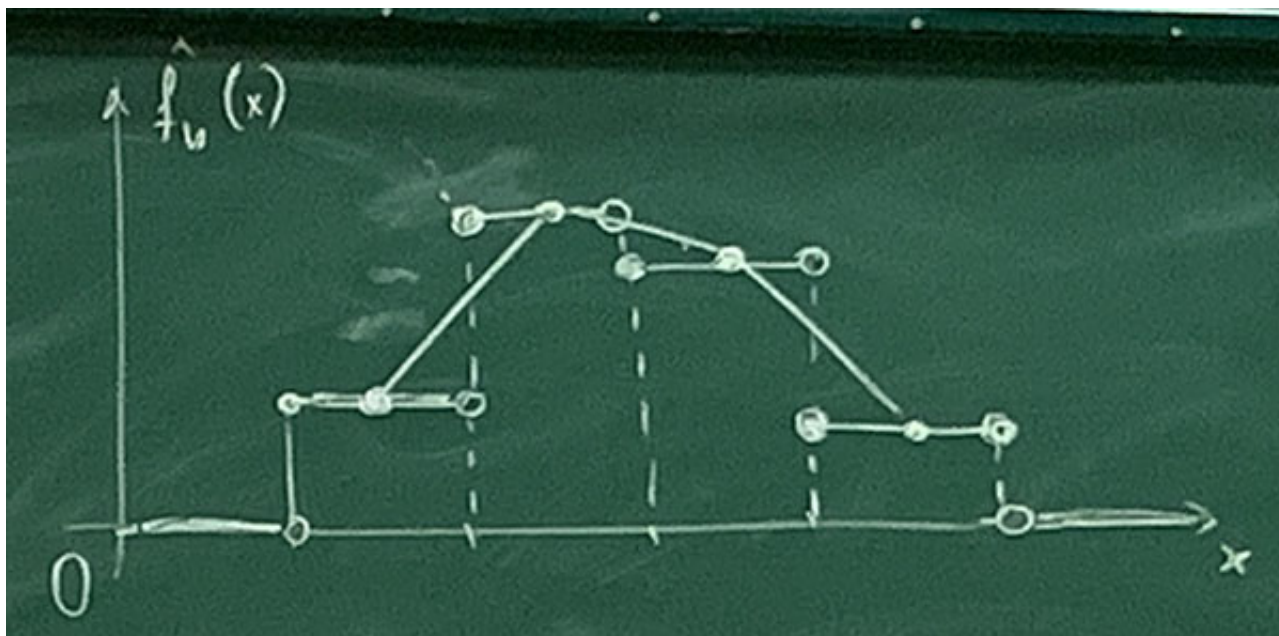
$$\hat{f}_n(x) \approx f(x),$$

где f – теоретическая («истинная») функция плотности распределения случайной величины X

Полигон частот

Предположим, что для выборки X построена гистограмма.

Определение 2.1.18 Полигоном частот называется ломаная, звенья которой соединяют середины верхних сторон соседних прямоугольников гистограммы.



2.2 Точечные оценки

2.2.1 Основные понятия

Рассмотрим 2-ю основную задачу математической статистики.

Дано: X – случайная величина, закон распределения которой известен с точностью до вектора $\vec{\Theta} = (\Theta_1, \dots, \Theta_r)$ неизвестных параметров (Это означает, что известен общий вид закона распределения случайной величины X , но неизвестны значения некоторых числовых параметров этого закона).

Требуется: оценить значение вектора $\vec{\Theta}$.

Для упрощения обозначений будем считать, что $r = 1$, т.е. $\vec{\Theta} = (\Theta)$

Определение 2.2.1 Точечной оценкой параметра Θ называется статистика $\hat{\Theta}(\vec{X})$, выборочное значение которой принимается в качестве значения этого параметра, то есть полагают

$$\Theta := \hat{\Theta}(\vec{x})$$

Пример 2.2.1 1. Пусть X – случайная величина, $m = MX$ – неизвестно. Для оценки параметра m можно предложить статистики:

$$\hat{m}_1(\vec{X}) = \overline{X} = \frac{1}{n} \sum_{i=1}^n X_i - \text{выборочное среднее}$$

$$\hat{m}_2(\vec{X}) = \frac{X_{(1)} + X_{(n)}}{2}$$

$$\hat{m}_3(\vec{X}) = \begin{cases} X_{(\frac{n+1}{2})}, & \text{если } n - \text{нечетное} \\ \frac{1}{2} \left[X_{(\frac{n}{2})} + X_{(\frac{n+2}{2})} \right], & \text{иначе} \end{cases}$$

$$\hat{m}_4(\vec{X}) = \sin(\overline{X})^2 + \sigma^2(\vec{X}) \cdot \exp(-\overline{X} + X_{(1)}^2 - X_{(n)}^{15})$$

Очевидно, что статистики $\hat{m}_1, \hat{m}_2, \hat{m}_3$ из предыдущего примера будут давать более или менее удовлетворительный результат, а последняя статистика вряд ли будет давать что-то близкое к m . Тем не менее, все эти статистики являются точечными оценками для m , так как в определении ничего не сказано о том, что соответствующая статистика должна давать «хорошие» результаты. Для исследования качества построенной оценки используются характеристики:

1. несмещенность
2. состоятельность
3. эффективность

Замечание 2.2.1 Из замечания не следует, что статистика $\hat{\Theta}$ должна принимать значения хоть сколько-нибудь близкие к теоретическому («истинному») значению параметра Θ .

2.2.2 Несмещенность точечной оценки

Определение 2.2.2 Пусть $\hat{\Theta}(\vec{X})$ – точечная оценка неизвестного параметра Θ .

Оценка $\hat{\Theta}$ параметра Θ называется несмещенной, если ее математическое ожидание равно теоретическому значению неизвестного параметра, т.е.

$$\exists M[\hat{\Theta}(\vec{X})] = \Theta$$

Пример 2.2.2 1. Пусть X – случайная величина, $t = MX$ – неизвестный параметр

(a) Рассмотрим статистику

$$\hat{m}_1 = \vec{X}$$

$$M[\hat{m}_1] = M\left[\overline{X}\right] = M\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n MX_i = \frac{1}{n} \sum_{i=1}^n t = t$$

является несмещенной оценкой для t

(b) Рассмотрим статистику

$$\hat{m}_2 = X_1$$

Тогда

$$M[\hat{m}_2] = M[X_1] = MX = t$$

тоже несмещенная оценка

2. X – случайная величина, $\exists MX, \exists DX = \sigma^2$

Оказывается выборочная дисперсия

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

является смещенной оценкой для σ^2

ФОТО

Определение 2.2.3 Статистика σ^2 называется направленной выборочной дисперсией

2.2.3 Состоятельность точечной оценки

Определение 2.2.4 Пусть

1. X – случайная величина, общий вид закона распределения которой известен
2. Θ – неизвестный параметр
3. $\hat{\Theta}(\vec{X})$ – точечная оценка для Θ

Оценка $\hat{\Theta}(\vec{X})$ параметра Θ называется состоятельной, если

$$\hat{\Theta}(\vec{X}) \xrightarrow[n \rightarrow \infty]{P} \Theta,$$

то есть

$$\forall \varepsilon > 0 P\{|\hat{\Theta}(\vec{X}) - \Theta| > \varepsilon\} \xrightarrow[n \rightarrow \infty]{} 0$$

Пример 2.2.3 Пусть

1. X – случайная величина
2. $EX = m$
3. $DX < \infty$

Тогда $\hat{m}_1 = \bar{X}$ – состоятельная оценка для m

Доказательство:

1. Последовательность X_1, X_2, \dots является последовательностью независимых одинаково распределенных случайных величин (см. определение случайной выборки)
2. $EX_i = EX = m$
3. $DX < \infty$

В соответствии со следствием ЗБЧ в форме Чебышева для одинаково распределенных случайных величин:

$$\forall \varepsilon > 0 P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - m\right| > \varepsilon\right\} \xrightarrow{n \rightarrow \infty} 0,$$

То есть

$$\overline{X} \xrightarrow[n \rightarrow \infty]{P} m$$

Замечание 2.2.2 Состоятельность оценки $\hat{m}_1 = \overline{X}$ математического ожидания можно доказать, не предполагая существования конечной дисперсии случайной величины X

Пример 2.2.4 Пусть

$$1. X \sim N(m, \sigma^2)$$

$$2. \hat{m}_2(\vec{X}) = X_1 - \text{точечная оценка для } m$$

Покажем, что m_2 не является состоятельной.

Доказательство:

$$\begin{aligned} P\{|\hat{m}_2(\vec{X}) - m| > \varepsilon\} &= P\{|X_1 - m| > \varepsilon\} = \left| X_1 \sim X, \text{ то есть } f_{X_1}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \right| = \left| X_1 - m \right| \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\varepsilon}^{\varepsilon} e^{-\frac{t^2}{2\sigma^2}} dt = \underbrace{2\Phi_0\left(\frac{\varepsilon}{\sigma}\right)}_{\text{не зависит от } n} \neq 0 \quad (\text{при } n \rightarrow \infty) \end{aligned}$$

Замечание 2.2.3 1. Можно показать, что оценка

$$S^2(\vec{X}) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$$

Является состоятельной оценкой дисперсии случайной величины X , если у X \exists моменты до 4-го порядка включительно

2. Более общо: можно показать, что если у случайной величины X \exists начальный и центральные моменты, то соответствующие выборочные аналоги являются их состоятельными оценками. При этом оценки моментов порядка $k \geq 2$ будут смещенными.

2.2.4 Эффективность точечной оценки

Определение 2.2.5 Пусть

1. X – случайная величина
2. Θ неизвестный параметр закона распределения случайной величины X
3. $\hat{\Theta}_1(\vec{X}), \hat{\Theta}_2(\vec{X})$ – две несмещенные точечные оценки для Θ
4. $\exists D\hat{\Theta}, \exists D\hat{\Theta}_2$

Оценка $\hat{\Theta}_2$ называется более эффективной, чем $\hat{\Theta}_1$, если

$$D\hat{\Theta}_2 < D\hat{\Theta}_1$$

Определение 2.2.6 Несмещенная оценка $\hat{\Theta}(\vec{X})$ параметра Θ называется эффективной, если она обладает наименьшей дисперсией среди всех несмещенных оценок параметра Θ .

Замечание 2.2.4 Часто говорят об оценке, эффективной среди класса H оценок параметра Θ . Более точно, если все оценки из класса H являются несмещенными, то оценка $\tilde{\Theta} \in H$ называется эффективной в классе H , если она обладает наименьшей дисперсией среди всех оценок из H .

Пример 2.2.5 Пусть X – случайная величина. $t = MX$ H – класс линейных оценок для t .

Покажем, что $\hat{t}_1(\vec{X}) = \bar{X}$ является эффективной оценкой в классе H .

Решение:

1. Линейная оценка имеет вид:

$$\hat{\mu}(\vec{X}) = \lambda_1 X_1 + \dots + \lambda_n X_n$$

ФОТО

Теорема 2.2.1 *О единственности эффективной оценки*

Пусть

1. $\hat{\Theta}_1(\vec{X})$ и $\hat{\Theta}_2(\vec{X})$ – две эффективные оценки для параметра Θ

Тогда

$$\hat{\Theta}_1(\vec{X}) = \hat{\Theta}_2(\vec{X})$$

Замечание 2.2.5 $\hat{\Theta}_1$ и $\hat{\Theta}_2$ являются случайными величинами, потому равенство $\hat{\Theta}_1(\vec{X}) = \hat{\Theta}_2(\vec{X})$ понимается:

$$P\{\vec{X} \in \{\vec{x} : \hat{\Theta}_1(\vec{x}) \neq \hat{\Theta}_2(\vec{x})\}\} = 0$$