

Seminar 3

Carlos Buitrago

Table of contents

1	Метод моментов	1
1.1	Теоретические и выборочные моменты	1
1.2	Обобщённая идея моментов	3
1.3	Пример: распределение Коши	3
1.4	Сильная состоятельность оценки метода моментов	4
1.4.1	Почему это верно?	4
1.5	Асимптотическая нормальность оценки метода моментов	4
2	Квантили	5
2.1	Квантиль распределения	5
2.2	Выборочный квантиль	5
2.3	Асимптотическая нормальность выборочного квантиля	5
2.4	Выборочная медиана	6
2.5	Пример: медиана нормального распределения	6
3	Метод максимального правдоподобия	7
3.1	Функция правдоподобия	7
3.2	Логарифмическая функция правдоподобия	8
3.3	Оценка максимального правдоподобия	8
3.4	Пример: $\text{Bin}(m, p)$, параметр p неизвестен	8
3.5	Пример: нормальное распределение	9
4	Связь трёх идей	9

1 Метод моментов

Пусть $X_1, \dots, X_n$ — независимая одинаково распределённая выборка из семейства распределений $\{P_\theta : \theta \in \Theta\}$, где параметр θ неизвестен.

Идея метода моментов: теоретические моменты распределения выражаются через неизвестный параметр, а затем заменяются соответствующими выборочными моментами.

1.1 Теоретические и выборочные моменты

k -м теоретическим моментом называется

$$m_k(\theta) = \mathbb{E}_\theta X^k,$$

если это математическое ожидание существует.

Соответствующий **выборочный момент** имеет вид

$$\widehat{m}_k = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Если параметр одномерный, обычно достаточно одного уравнения

$$\widehat{m}_1 = m_1(\theta).$$

Если $\theta = (\theta_1, \dots, \theta_k)$ — k -мерный параметр, используют систему

$$\begin{cases} \widehat{m}_1 = m_1(\theta), \\ \vdots \\ \widehat{m}_k = m_k(\theta). \end{cases}$$

Если отображение

$$m : \Theta \rightarrow m(\Theta), \quad m(\theta) = (m_1(\theta), \dots, m_k(\theta))$$

обратимо, то оценка метода моментов записывается как

$$\widehat{\theta}_{\text{ММ}} = m^{-1}(\widehat{m}_1, \dots, \widehat{m}_k).$$

i Пример: распределение Пуассона

Пусть

$$X_1, \dots, X_n \sim \text{Pois}(\lambda).$$

Так как

$$\mathbb{E}X = \lambda,$$

приравниваем теоретический и выборочный моменты:

$$\overline{X} = \lambda.$$

Следовательно,

$$\widehat{\lambda}_{\text{ММ}} = \overline{X}.$$

i Пример: $U[0, \theta]$

Если $X \sim U[0, \theta]$, то

$$\mathbb{E}X = \frac{\theta}{2}.$$

Поэтому

$$\overline{X} = \frac{\theta}{2} \implies \widehat{\theta}_{\text{ММ}} = 2\overline{X}.$$

1.2 Обобщённая идея моментов

Не обязательно использовать только степени X^k . Можно выбрать функцию g такую, что

$$m(\theta) = \mathbb{E}_\theta g(X)$$

существует и однозначно определяет θ . Тогда можно решить уравнение

$$\frac{1}{n} \sum_{i=1}^n g(X_i) = m(\theta).$$

Это особенно полезно для тяжёлохвостых распределений, у которых обычные моменты могут не существовать.

1.3 Пример: распределение Коши

Пусть

$$X \sim \text{Cauchy}(0, \theta), \quad f_\theta(x) = \frac{\theta}{\pi(x^2 + \theta^2)}, \quad \theta > 0.$$

Обычные положительные целые моменты распределения Коши не существуют. Возьмём вместо них ограниченную функцию

$$g(x) = \frac{x^2}{1 + x^2}.$$

Тогда

$$\mathbb{E}_\theta g(X) = \mathbb{E}_\theta \frac{X^2}{1 + X^2} = \frac{\theta}{1 + \theta}.$$

Обозначим

$$M_n = \frac{1}{n} \sum_{i=1}^n \frac{X_i^2}{1 + X_i^2}.$$

Из моментного уравнения

$$M_n = \frac{\theta}{1 + \theta}$$

получаем

$$\hat{\theta} = \frac{M_n}{1 - M_n}.$$

1.4 Сильная состоятельность оценки метода моментов

! Теорема

Пусть

$$m : \Theta \rightarrow m(\Theta)$$

— биекция, причём функция m^{-1} определена и непрерывна в каждой точке множества $m(\Theta)$.
Предположим также, что

$$\mathbb{E}_\theta |g_i(X_1)| < \infty, \quad i = 1, \dots, k,$$

для всех $\theta \in \Theta$, где

$$m_i(\theta) = \mathbb{E}_\theta g_i(X_1).$$

Тогда оценка метода моментов является **сильно состоятельной**:

$$\hat{\theta}_n \xrightarrow{a.s.} \theta.$$

1.4.1 Почему это верно?

По усиленному закону больших чисел

$$\frac{1}{n} \sum_{j=1}^n g_i(X_j) \xrightarrow{a.s.} \mathbb{E}_\theta g_i(X_1) = m_i(\theta).$$

Следовательно,

$$\widehat{m}_n \xrightarrow{a.s.} m(\theta).$$

По непрерывности m^{-1} ,

$$\hat{\theta}_n = m^{-1}(\widehat{m}_n) \xrightarrow{a.s.} m^{-1}(m(\theta)) = \theta.$$

1.5 Асимптотическая нормальность оценки метода моментов

! Теорема

В условиях предыдущей теоремы предположим дополнительно, что m^{-1} дифференцируема в точке $m(\theta)$ и

$$\mathbb{E}_\theta [g_i(X_1)^2] < \infty, \quad i = 1, \dots, k.$$

Тогда оценка метода моментов асимптотически нормальна.

Действительно, многомерная ЦПТ даёт

$$\sqrt{n}(\widehat{m}_n - m(\theta)) \xrightarrow{d} \mathcal{N}_k(0, \Sigma),$$

где

$$\Sigma = \text{Cov}_\theta(g(X_1)).$$

Применяя многомерный дельта-метод к m^{-1} , получаем

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}_k(0, Dm^{-1}(m(\theta)) \Sigma Dm^{-1}(m(\theta))^\top).$$

2 Квантили

2.1 Квантиль распределения

💡 Определение

Для $p \in (0, 1)$ величина

$$z_p = \inf\{x : F_P(x) \geq p\}$$

называется **p -квантилем** распределения P .

Например, при $p = 1/2$ получаем медиану распределения.

2.2 Выборочный квантиль

Пусть

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

— вариационный ряд выборки $X_1, \dots, X_n$.

💡 Определение

Статистика

$$z_{n,p} = \begin{cases} X_{(\lfloor np \rfloor + 1)}, & np \notin \mathbb{Z}, \\ X_{(np)}, & np \in \mathbb{Z}, \end{cases}$$

называется **выборочным p -квантилем**.

2.3 Асимптотическая нормальность выборочного квантиля

! Теорема

Пусть $X_1, \dots, X_n$ — выборка из распределения P с плотностью $\rho(x)$. Пусть z_p — p -квантиль распределения, причём ρ непрерывна в точке z_p и

$$\rho(z_p) > 0.$$

Тогда

$$\sqrt{n}(z_{n,p} - z_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{\rho^2(z_p)}\right).$$

Отсюда, при больших n ,

$$z_{n,p} \approx \mathcal{N}\left(z_p, \frac{p(1-p)}{n\rho^2(z_p)}\right).$$

2.4 Выборочная медиана

💡 Определение

Выборочной медианой для выборки $X_1, \dots, X_n$ называется

$$\hat{\mu} = \begin{cases} X_{(k+1)}, & n = 2k + 1, \\ \frac{X_{(k)} + X_{(k+1)}}{2}, & n = 2k. \end{cases}$$

Для непрерывного распределения с медианой μ и плотностью $\rho(\mu) > 0$ теорема о выборочном квантиле при $p = 1/2$ даёт

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4\rho^2(\mu)}\right).$$

2.5 Пример: медиана нормального распределения

Пусть

$$X_i \sim \mathcal{N}(\mu, \sigma^2).$$

Медиана равна μ , а

$$\rho(\mu) = \frac{1}{\sigma\sqrt{2\pi}}.$$

Поэтому

$$\frac{1}{4\rho^2(\mu)} = \frac{\pi\sigma^2}{2},$$

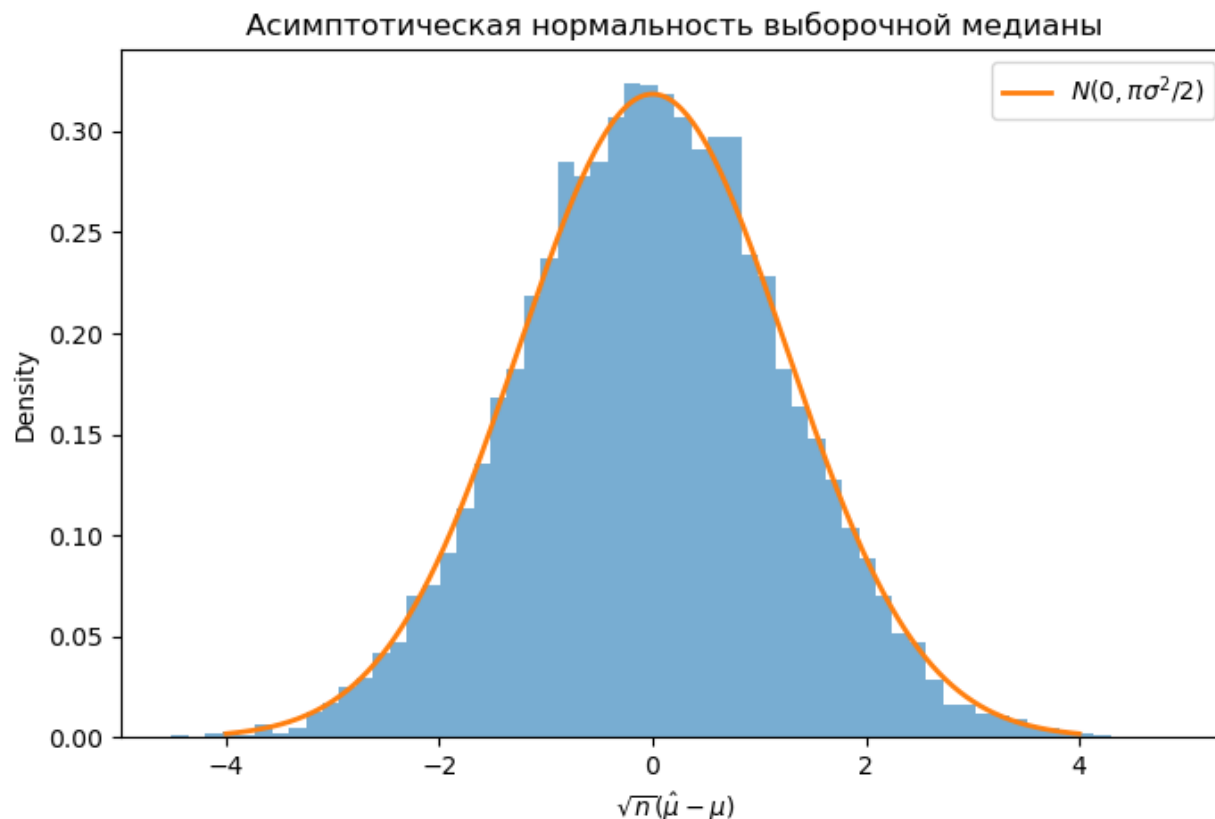
и

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi\sigma^2}{2}\right).$$

Для сравнения,

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Таким образом, при нормальной модели выборочное среднее асимптотически эффективнее медианы, но медиана значительно устойчивее к выбросам.



3 Метод максимального правдоподобия

Пусть

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

— доминируемое семейство распределений с плотностью $\rho_\theta(x)$.

3.1 Функция правдоподобия

Для независимой одинаково распределённой выборки $X_1, \dots, X_n$ совместная плотность имеет вид

$$f_\theta(X_1, \dots, X_n) = \prod_{i=1}^n \rho_\theta(X_i).$$

Если рассматривать наблюдения $x_1, \dots, x_n$ как фиксированные, а θ — как переменную, эта же функция называется **функцией правдоподобия**:

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n \rho_\theta(x_i).$$

3.2 Логарифмическая функция правдоподобия

Вместо произведения обычно удобнее максимизировать его логарифм:

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log \rho_{\theta}(X_i).$$

Поскольку логарифм строго возрастает,

$$\arg \max_{\theta \in \Theta} L(\theta) = \arg \max_{\theta \in \Theta} \ell(\theta).$$

3.3 Оценка максимального правдоподобия

💡 Определение

Оценкой максимального правдоподобия (MLE) параметра θ называется статистика

$$\hat{\theta}_{\text{MLE}}(\mathbf{X}) \in \arg \max_{\theta \in \Theta} L(\theta; X_1, \dots, X_n).$$

На практике обычно:

1. записывают $L(\theta)$;
2. переходят к $\ell(\theta) = \log L(\theta)$;
3. находят критические точки из $\ell'(\theta) = 0$;
4. проверяют границы пространства параметров и убеждаются, что найден максимум.

Важно: если пространство параметров дискретно, дифференцирование может быть неприменимо. Тогда удобно сравнивать соседние значения правдоподобия, например $L(\theta + 1)/L(\theta)$.

3.4 Пример: $\text{Bin}(m, p)$, параметр p неизвестен

Пусть

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bin}(m, p),$$

где m известно. Тогда

$$L(p) = \prod_{i=1}^n \binom{m}{X_i} p^{X_i} (1-p)^{m-X_i}.$$

С точностью до слагаемого, не зависящего от p ,

$$\ell(p) = \left(\sum_{i=1}^n X_i \right) \log p + \left(nm - \sum_{i=1}^n X_i \right) \log(1-p).$$

Из $\ell'(p) = 0$ получаем

$$\hat{p}_{\text{MLE}} = \frac{\bar{X}}{m}.$$

3.5 Пример: нормальное распределение

Пусть

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Если оба параметра неизвестны, максимизация логарифмического правдоподобия приводит к

$$\hat{\mu}_{\text{MLE}} = \bar{X},$$

и

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

Обратите внимание: MLE для σ^2 использует делитель n , а не $n - 1$, поэтому при конечном n эта оценка смещена.

4 Связь трёх идей

В этой части курса появляются три разных способа строить оценки:

- **метод моментов:** сопоставляем теоретические и эмпирические средние некоторых функций $g(X)$;
- **выборочные квантили:** оцениваем положение распределения через порядковые статистики;
- **метод максимального правдоподобия:** выбираем параметр, при котором наблюдаемая выборка наиболее правдоподобна.

Эти методы могут приводить как к одинаковым, так и к различным оценкам. После построения оценки нас интересуют её свойства: состоятельность, смещение, дисперсия, асимптотическая нормальность и эффективность.