Теория Вероятностей и Статистика

Дельта метод и инвариантность оценок метода максимального правдоподобия

Потанин Богдан Станиславович

старший преподаватель, кандидат экономических наук

2021-2022

Инвариантность

Мотивация

- Вспомним, что характеристики распределения (вероятности, математическое ожидание, медиана и т.д.) являются функциями от параметров распределений. Ранее мы использовали это для того, чтобы оценивать характеристики распределений при помощи оценок параметров распределений.
- Если при этом мы использовали состоятельную оценку параметра распределения, то, при соблюдении некоторых условий, получали состоятельную оценку характеристики распределения. Например, в случае с экспоненциальным распределением полученная при помощи метода моментов состоятельная оценка $\hat{\lambda}_n = 1/\overline{X}_n$ использовалась для получения состоятельной оценки дисперсии $Var(X_1) = 1/\hat{\lambda}_n^2 = \overline{X}_n^2$.
- Ранее мы показали, что оценки, полученные с помощью метода максимального правдоподобия, обладают рядом хороших свойств: состоятельность, асимптотическая эффективность и асимптотическая нормальность.
- Будут ли сохраняться эти благоприятные свойства для оценок характеристик распределения, если для построения этих оценок мы воспользуемся ММП оценками? Свойство инвариантности гарантирует, при определенных условиях, положительный ответ на данный вопрос. Например, поскольку $\hat{\lambda}_n = 1/\overline{X}_n$ является также ММП оценкой, то оценка $Var(X_1) = 1/\hat{\lambda}_n^2 = \overline{X}_n^2$ будет не только состоятельной, но и асимптотически нормальной и асимптотически эффективной.

Дельта метод

Формулировка

- Рассмотрим последовательность случайных величин $X_1, X_2, ...$ такую, что:

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

- Тогда для функции $g(.)$ с ненулевой производной $g'(\mu) \neq 0$ справедливо:

$$\sqrt{n}\left(g\left(X_n\right) - g\left(\mu\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\left(g'(\mu)\right)^2\right)$$

- На практике при достаточно большом $n \geq 100$ можно предположить, что при соблюдении обозначенных условий:

$$X_n \dot{\sim} \mathcal{N}\left(\mu, \sigma^2/n\right) \implies g(X_n) \dot{\sim} \mathcal{N}\left(g\left(\mu\right), \sigma^2\left(g'(\mu)\right)^2/n\right)$$

Пример: имеется последовательность $X_1, X_2, ...$, где $X_i \sim U(2, 8)$. При помощи дельта метода найдем асимптотическое распределение $g\left(\overline{X}_n\right) = \left(\overline{X}_n\right)^3$. Используя ЦПТ, нетрудно показать, что $\overline{X}_n \dot{\sim} \mathcal{N}\left(5, 3/n\right)$, где $\mu = 5$ и $\sigma^2 = 3$. Поскольку $g'(5) = 3 \times 5^2 = 75 \neq 0$, то вследствие дельта метода:

$$\left(\overline{X}_n\right)^3 \dot{\sim} \mathcal{N}\left(5^3, 3 \times 75^2/n\right) = \mathcal{N}\left(125, 16875/n\right)$$

Для примера рассчитаем следующую вероятность:

$$P\left(\left(\overline{X}_{1000}\right)^3 \leq 130\right) \approx \Phi\left(\frac{130 - 125}{\sqrt{16875/1000}}\right) \approx \Phi(1.217) \approx 0.888$$

- Имеется последовательность Хи-квадрат случайных величин $\chi^2_1, \chi^2_2, \ldots$, где $\chi^2_i \sim \chi^2(i)$. Используя ЦПТ, нетрудно показать, что $X_n = \left(\chi^2_n/n\right) \dot{\sim} \mathcal{N}\left(1, 2/n\right)$, где $\mu = 1$ и $\sigma^2 = 2$. С помощью дельта метода найдем приблизительное распределение для $g(\chi^2_n) = \sqrt{\chi^2_n}$. Сперва рассмотрим $g(X_n)$ и, учитывая что $g'(1) = 1/(2\sqrt{1}) = 0.5 \neq 0$, получаем:

$$\sqrt{X_n} \dot{\sim} \mathcal{N}\left(\sqrt{1}, (2/n) \times 0.5^2\right) = \mathcal{N}\left(1, 0.5/n\right) \implies \sqrt{\chi^2_n} = \sqrt{nX_n} = \sqrt{n}\sqrt{X_n} \dot{\sim} \sqrt{n}\mathcal{N}\left(1, 0.5/n\right) = \mathcal{N}\left(\sqrt{n}, 0.5\right)$$

  Для примера рассчитаем вероятность:

$$P(\sqrt{\chi^2_{100}} \leq 10.5) \approx \Phi\left(\frac{10.5 - \sqrt{100}}{\sqrt{0.5}}\right) = \Phi\left(\sqrt{0.5}\right) \approx 0.76$$

  Примечание: убедитесь, что по аналогии не удастся аппроксимировать распределение $\sin(\chi^2_n)$.

- Найдем асимптотическое распределение оценки параметра экспоненциального распределения $\hat{\lambda}_n = 1/\overline{X}_n$. В силу ЦПТ $\overline{X}_n \dot{\sim} \left(1/\lambda, (1/\lambda^2)/n\right)$, а значит, полагая $g\left(\overline{X}_n\right) = \hat{\lambda}_n$ и применяя дельта метод, имеем $g'(1/\lambda) = \lambda^2 \neq 0$, откуда:

$$\hat{\lambda}_n \dot{\sim} \mathcal{N}\left(1/(1/\lambda), \left(\left(\lambda^2\right)^2/\lambda^2\right)/n\right) = \mathcal{N}\left(\lambda, \lambda^2/n\right)$$

Дельта метод

Подготовка к доказательству

- **Теорема о среднем значении**: пусть имеется функция $f(x)$, непрерывная на интервале $[a, b]$ и дифференцируемая на открытом интервале (a, b), где $b > a$. Тогда существует константа $c \in (a, b)$, такая, что:
$$f'(c) = \frac{f(b) - f(a)}{b - a}$$
- **Лемма**: пусть дана последовательность $X_1, X_2, ...$, такая, что:
$$\sqrt{n}(X_n - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$
Тогда $X_n \xrightarrow{p} \mu$.
Интуиция (не доказательство):
$$\lim_{n\to\infty} P(|X_n - \mu| > \varepsilon) = \lim_{n\to\infty} P(\sqrt{n}|X_n - \mu| > \sqrt{n}\varepsilon) \underset{\text{грубый переход}}{=} 2 - \lim_{n\to\infty} \Phi\left(\sqrt{n}\varepsilon\right) = 2 - 2 = 0$$

Дельта метод

Доказательство

Для простоты предположим непрерывность функции $g(.)$ и, применяя теорему о среднем (mean value theorem) значении, зададим случайную величину $\tilde{\mu}$, такую, что:

$$g'(\tilde{\mu}) = \frac{g(X_n) - g(\mu)}{X_n - \mu}$$

Из сформулированной ранее леммы известно, что $X_n \xrightarrow{p} \mu$. Покажем, что из этого следует $\tilde{\mu} \xrightarrow{p} \mu$. Поскольку $\tilde{\mu}$ лежит между $\overline{X}_n$ и μ, то $|X_n - \mu| > |\tilde{\mu} - \mu|$, а значит при любом $\varepsilon > 0$:

$$\lim_{n\to\infty} P(|\tilde{\mu} - \mu| > \varepsilon) > \lim_{n\to\infty} P(|X_n - \mu| > \varepsilon) = 0$$

Используя непрерывность $g(.)$ и теорему Манна-Вальда получаем $g'(\tilde{\mu}) \xrightarrow{p} g'(\mu)$.
Пользуясь записанным с помощью теоремы о среднем значении выражением, получаем:

$$g(X_n) - g(\mu) = g'(\tilde{\mu})(X_n - \mu) \implies \sqrt{n}\,(g(X_n) - g(\mu)) = \sqrt{n}\,(g'(\tilde{\mu})(X_n - \mu))$$

Поскольку по условию теоремы $\sqrt{n}\,(X_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, то применяя теорему Слуцкого получаем:

$$\sqrt{n}\,(g(X_n) - g(\mu)) = g'(\tilde{\mu})\sqrt{n}\,(X_n - \mu) \xrightarrow{d} g'(\mu)N\left(0, \sigma^2\right) = N\left(0, \sigma^2\left(g'(\mu)\right)^2\right)$$

Инвариантность оценок метода максимального правдоподобия

Формулировка

- Рассмотрим ММП оценку $\hat{\theta}_n$ параметра θ.
- Если функция $g(.)$ монотонна, то $g(\hat{\theta}_n)$ также является ММП оценкой, а значит обладает присущими ММП оценкам привлекательными свойствами: состоятельность, асимптотическая нормальность и асимптотическая эффективность.
- Пользуясь асимптотической нормальностью ММП оценок и дельта методом можно найти асимптотическое распределение функций от ММП оценок:

$$\sqrt{n}\left(\hat{\theta}_n-\theta\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{1}{i(\theta)}\right)\implies\sqrt{n}\left(g(\hat{\theta}_n)-g(\theta)\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\left(g'(\theta)\right)^2}{i(\theta)}\right)$$

- На практике предполагается, что $g(\hat{\theta}_n)\dot{\sim}\mathcal{N}\left(g(\theta),\frac{\left(g'(\theta)\right)^2}{ni(\theta)}\right)$ или $g(\hat{\theta}_n)\dot{\sim}\mathcal{N}\left(g(\hat{\theta}_n),\frac{\left(g'(\hat{\theta}_n)\right)^2}{ni(\hat{\theta}_n)}\right)$.
- Асимптотическая дисперсия и ее оценка имеют вид $As.Var(g(\hat{\theta}_n))=\frac{\left(g'(\theta)\right)^2}{ni(\theta)}$ и $\widehat{As.Var}(g(\hat{\theta}_n))=\frac{\left(g'(\hat{\theta}_n)\right)^2}{ni(\hat{\theta}_n)}$.

  Пример: по выборке из распределения Пуассона с помощью метода максимального правдоподобия была найдена оценка $\hat{\lambda}_n=\overline{X}_n$. Найдем асимптотическое распределение и оценку асимптотической дисперсии оценки вероятности того, что наблюдение примет нулевое значение. В силу монотонности экспоненциальной функции применима инвариантность, вследствие которой ММП оценкой $P(X_1=0)=e^{-\lambda}$ будет $\hat{P}(X_1=0)=e^{-\hat{\lambda}_n}$. Напомним, что $i(\lambda)=1/\lambda$ и вычислим $g'(\lambda)=P'(X_1=0)=-e^{-\lambda}$, откуда:

$$e^{-\hat{\lambda}_n}\dot{\sim}\mathcal{N}\left(e^{\lambda},\lambda e^{-2\lambda}/n\right)\qquad\widehat{As.Var}\left(e^{\hat{\lambda}_n}\right)=\overline{X}_ne^{-2\overline{X}_n}/n$$

Инвариантность оценок метода максимального правдоподобия

Дополнительный пример

Время (в часах) на прохождение миссии в игре случайно взятым игроком является экспоненциальной случайной величиной с параметром λ. По выборке из времени, затраченного игроками на прохождение миссии, найдите ММП оценку дисперсии времени, затрачиваемого игроками на прохождение миссии, а также асимптотическую дисперсию данной оценки и ее оценку. По выборке из $n = 1000$ наблюдений с реализацией выборочного среднего $\overline{x}_n = 0.5$ приблизительно рассчитайте вероятность того, что ММП оценка дисперсии наблюдений превысит 0.26.

Решение: поскольку ММП оценка имеет вид $\hat{\lambda}_n = 1/\overline{X}_n$ и оцениваемая дисперсия является монотонной функцией $Var(X_1) = 1/\lambda^2$, то в силу инвариантности $\widehat{Var}(X_1) = 1/\hat{\lambda}_n^2 = \left(\overline{X}_n\right)^2$. Поскольку $i(\lambda) = 1/\lambda^2$ и $g'(\lambda) = Var'(X_1) = -2\lambda^{-3}$, то:

$$\widehat{Var}(X_1)\dot{\sim}\mathcal{N}\left(1/\lambda^2, (-2\lambda^{-3})^2/\left(n\left(1/\lambda^2\right)\right)\right) = \mathcal{N}\left(\lambda^{-2}, 4\lambda^{-4}/n\right)$$

$$As.Var(\widehat{Var}(X_1)) = 4\lambda^{-4}/n \implies \widehat{As.Var}(\widehat{Var}(X_1)) = 4\overline{X}_n^4/n$$

Поскольку $\overline{x}_n = 0.5$, то $\hat{\lambda}_n(x) = 1/0.5 = 2$, откуда:

$$\widehat{Var}(X_1)\dot{\sim}\mathcal{N}\left(1/2^2, 4 \times 2^{-4}/1000\right) = \mathcal{N}(0.25, 0.00025)$$

Используя полученную информацию приблизительно рассчитаем искомую вероятность:

$$P(\widehat{Var}(X_1) > 0.26) \approx 1 - \Phi\left(\frac{0.26 - 0.25}{\sqrt{0.00025}}\right) \approx 0.263545$$