Асимптотические свойства оценки максимального правдоподобия

Пусть $F(x, \theta)$ - функция распределения выборки, $\theta = (\theta_1,\ \dots,\ \theta_k)^T$

$\widehat{\theta}_n$ - оценка максимального правдоподобия при выборке объёма $n$

Свойство состоятельности состоит в том, что $\widehat{\theta}_n \to^{\mathbb{P}_\theta} \theta$ при $n \to \infty$

Свойство асимптотической нормальности

$\sqrt{n}(\widehat{\theta}_n - \theta) \to^d \xi,\ Low(\xi) = N(0, \Sigma(\theta))$, где $\Sigma(\theta)$ - матрица ковариаций $\xi$

Для оценки максимального правдоподобия $\Sigma(\theta) = I^{-1}(\theta)$, где $I(\theta)$ - информационная матрица Фишера, $(I(\theta))_{ij} = E_\theta \frac{\partial \ln L}{\partial \theta_i} \cdot \frac{\partial \ln L}{\partial \theta_j}$

Теорема (сформулирована в общем виде, а доказана в частном)

Пусть для регулярной модели $F(x, \theta)$ с векторным параметром $\theta = (\theta_1,\ \dots,\ \theta_k)^T$ и есть производные по параметрам, а также локальный максимум функции правдоподобия единственный и достигается во внутренней точке $\widehat{\theta}_n$ параметрического множества, тогда

1) $\widehat{\theta}_n$ является состоятельной оценкой для $\theta$

2) Если дополнительно существуют производные $\frac{\partial^3 f(x,\theta}{\partial \theta_i \partial \theta_j \partial \theta_k}$ и существует функция $M(x)$ такая, что при всех $\theta$ и $x$ $|\frac{\partial^3 f(x,\theta}{\partial \theta_i \partial \theta_j \partial \theta_k}| \le M(x)$ и $E_\theta M(X) < +\infty$, то $\sqrt{n}(\widehat{\theta}_n - \theta) \to \xi,\ Low(\xi) = N(0, I^{-1}(\theta)$

3) Если $\tau(\theta)$ - дифференцируемая функция от параметра, то $\widehat{\tau}_n = \tau(\widehat{\theta}_n)$ является состоятельной оценкой $\tau(\theta)$ и $\sqrt{n}(\widehat{\tau}_n - \tau(\theta)) \to^d \nu,\ Low(\nu) = N(0, \sigma^2(\theta)$, где $\sigma^2(\theta) = (grad\ \tau(\theta))^T(\theta) I^{-1}(\theta)(grad\ \tau(\theta))$

Доказательство

1) Возьмём тождественную функцию от параметра: $\tau(\theta) = \theta$

Тогда $\widehat{\tau}_n = \tau(\widehat{\theta}_n) = \widehat{\theta}_n$ по свойству оценки максимального правдоподобия является эффективной оценкой для $\tau$, то есть $E_\theta \widehat{\theta}_n = E_\theta \widehat{\tau}_n = \tau(\theta) = \theta$ - состоятельность оценки максимального правдоподобия доказана

Будем доказывать 2) и 3) в случае скалярного параметра

Условие на производную в случае скалярного параметра запишется таким образом: $\exists M(x)\ :\ |\frac{\partial^3 f(x,\theta}{\partial \theta_i \partial \theta_j \partial \theta_k}| \le M(x)$ и $E_\theta M(X) < +\infty$

Утверждения 2) и 3) в скалярном случае:

2) $\sqrt{n}(\widehat{\theta}_n - \theta) \to^d \xi\ Low(\xi) = N(0, \frac{1}{i_1(\theta)})$, где $i_1(\theta) = E_\theta(\frac{\partial \ln L(X_1, \theta)}{\partial \theta})^2 = D_\theta U_1$

3) $\sigma^2(\theta) = \frac{(\tau'(\theta))^2}{i_1(\theta)}$

$0 = U(\widehat{\theta}_n)$ - максимум функции

$0 = U(\widehat{\theta}_n) = U(\theta) + U'(\theta)(\widehat{\theta}_n - \theta) + \frac{1}{2}U''(\theta^*)(\widehat{\theta}_n - \theta)^2$, $\theta^*$ лежит между $\widehat{\theta}_n$ и $\theta$

$\widehat{\theta}_n - \theta = -U(\theta) \cdot (U'(\theta) + \frac{1}{2}U''(\theta^*)(\widehat{\theta}_n - \theta))^{-1}$

$\sqrt{n}(\widehat{\theta}_n - \theta) = -\sqrt{n}U(\theta) \cdot (U'(\theta) + \frac{1}{2}U''(\theta^*)(\widehat{\theta}_n - \theta))^{-1}$

Заметим, что в силу центральной предельной теоремы для одинаково распределённых слагаемых

$\frac{U(\theta)}{\sqrt{n}\sqrt{i_1(\theta)}} = \frac{\sum\limits_k \frac{\partial \ln L(X_k, \theta}{\partial \theta}}{\sqrt{n}\sqrt{i_1(\theta)}}$ сходится по распределению к случайной величине $\xi$, имеющей стандартное нормальное распределение

$\epsilon_n = \frac{1}{2}U''(\theta^*)(\widehat{\theta}_n - \theta)$

Мы знаем, что $\widehat{\theta}_n - \theta \to^{\mathbb{P}_\theta} 0$ - состоятельность оценки (утверждение 1)

$\frac{1}{n}U''(\theta^*) = \frac{1}{n}\sum\limits_{i=1}^n \frac{\partial^3}{\partial \theta^3} \ln L(X_i, \theta^*)$, причём $\theta^*$ разная для разных $X_i$

$|\frac{U''(\theta)}{n}| \le \frac{1}{n}\sum\limits_{i=1}^n M(X_i) \to^{\mathbb{P}_\theta} E_\theta M(X_1) < +\infty$

$\sqrt{n}(\widehat{\theta}_n - \theta) = \sqrt{n}U(\theta)(-\frac{U'(\theta)}{n} - \frac{\epsilon_n}{n})^{-1}n^{-1}$

$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{U(\theta)}{\sqrt{n}}(-\frac{U'(\theta)}{n} - \frac{\epsilon_n}{n})^{-1}$ - надо оценить множитель перед скобкой и 2 дроби в скобке

$\frac{U(\theta)}{\sqrt{n}\sqrt{i_1(\theta)}}$ сходится к стандартной нормальной величине, $\frac{U(\theta)}{\sqrt{n}i_1(\theta)}$ сходится к распределению $N(0, \frac{1}{i_1(\theta)})$

$-\frac{U'(\theta)}{n} = \frac{1}{n}\sum\limits_{i=1}^{n}(-\frac{\partial^2}{\partial\theta^2}\ln L(X_i, \theta)) \to^{\mathbb{P}} -E_\theta\frac{\partial^2}{\partial\theta^2}\ln L(X_1, \theta) = D_\theta U(X_1, \theta) = i_1(\theta)$

$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{U(\theta)}{\sqrt{n}}(-\frac{U'(\theta)}{n} - \frac{\epsilon_n}{n})^{-1}$, множитель перед скобкой сходится по распределению к $\xi \sim N(0, \frac{1}{i_1(\theta)})$, первая дробь в скобке стремится по вероятности к единице

Осталось доказать, что $\frac{1}{n}\epsilon_n = \frac{1}{2n}U''(\theta^*)(\widehat{\theta}_n - \theta) \to^{\mathbb{P}} 0$

$\mathbb{P}_\theta(\{\frac{1}{2n}|U''(\theta^*)(\widehat{\theta}_n - \theta)| > \delta\}) \leq \mathbb{P}_\theta(\{\frac{1}{2n}\sum\limits_{i=1}^{n}M(X_i) \cdot |\widehat{\theta}_n - \theta| > \delta\})$

$\eta_n = \frac{1}{2}|\widehat{\theta}_n - \theta| \to^{\mathbb{P}} 0$

$\xi_n = \frac{1}{n}\sum\limits_{i=1}^{n}M(X_i) \to^{\mathbb{P}} E_\theta M(X_1) = c > 0$

$\eta_n \to^{\mathbb{P}} 0,\ \xi_n \to^{\mathbb{P}} c > 0$

$\mathbb{P}_\theta(\{\xi_n\eta_n > \delta\}) = \mathbb{P}_\theta(\{\xi_n\eta_n > \delta\} \cap \{|\xi_n - c| > \frac{c}{2}\}) + \mathbb{P}_\theta(\{\xi_n\eta_n > \delta\} \cap \{|\xi_n - c| < \frac{c}{2}\})$

$\mathbb{P}_\theta(\{\xi_n\eta_n > \delta\} \cap \{|\xi_n - c| > \frac{c}{2}\}) \leq \mathbb{P}_\theta(\{|\xi_n - c| > \frac{c}{2}\}) \to 0$

$\mathbb{P}_\theta(\{\xi_n\eta_n > \delta\} \cap \{|\xi_n - c| < \frac{c}{2}\}) \leq \mathbb{P}_\theta(\{\frac{3}{2}\eta_n > \delta\}) \to 0$

Продолжение в следующей лекции