L04 Asymptotically efficient estimators

1. Efficient estimators

   (1) Efficient estimators

$$\begin{aligned} \widehat{\theta} \text{ is an efficient estimator for } \theta &\stackrel{def}{\iff} E_\theta(\widehat{\theta}) = \theta \text{ and } \mathrm{Cov}_\theta(\widehat{\theta}) = \mathrm{CRLB}(\theta) \\ &\implies \widehat{\theta} \text{ is the best one in UE}(\theta) \text{ by MSCPE risk.} \end{aligned}$$

   Comment: $E_\theta(\widehat{\theta})$ and $\mathrm{Cov}_\theta(\widehat{\theta})$ are functions of θ. $E_\theta(\widehat{\theta}) = \theta$ and $\mathrm{Cov}_\theta(\widehat{\theta}) = \mathrm{CRLB}(\theta)$ mean $E_\theta(\widehat{\theta}) \equiv \theta$ and $\mathrm{Cov}_\theta(\widehat{\theta}) \equiv \mathrm{CRLB}(\theta)$ for all θ.

   (2) Efficiency (function)
   For $\widehat{\theta} \in R^1$, $e_{\widehat{\theta}}(\theta) = \dfrac{\mathrm{CRLB}(\theta)}{\mathrm{var}_\theta(\widehat{\theta})}$ is the efficiency function for $\widehat{\theta}$. $0 < e_{\widehat{\theta}}(\theta) \le 1$ and

$$e_{\widehat{\theta}}(\theta) = 1 \iff \mathrm{var}_\theta(\widehat{\theta}) = \mathrm{CRLB}(\theta).$$

   So $\widehat{\theta}$ is an efficient estimator for $\theta \in R$ iff $\widehat{\theta} \in \mathrm{UE}(\theta)$ with efficiency 1.

   (3) Relative efficiecy
   For estimators $\widehat{\theta}$ and $\widetilde{\theta}$ for $\theta \in R^1$

$$e_{(\widehat{\theta},\widetilde{\theta})}(\theta) = \frac{e_{\widehat{\theta}}(\theta)}{e_{\widetilde{\theta}}(\theta)} = \frac{\mathrm{var}_\theta(\widetilde{\theta})}{\mathrm{var}_\theta(\widehat{\theta})} \text{ is the relative efficiency of } \widehat{\theta} \text{ to } \widetilde{\theta}.$$

   Then $e_{(\widehat{\theta},\widetilde{\theta})}(\theta) > 0$ and

$$\begin{aligned} e_{(\widehat{\theta},\widetilde{\theta})}(\theta) < 1 \text{ at } \theta &\iff \mathrm{var}_\theta(\widehat{\theta}) > \mathrm{var}_\theta(\widetilde{\theta}) \text{ at } \theta \\ e_{(\widehat{\theta},\widetilde{\theta})}(\theta) > 1 \text{ at } \theta &\iff \mathrm{var}_\theta(\widehat{\theta}) < \mathrm{var}_\theta(\widetilde{\theta}) \text{ at } \theta \\ e_{(\widehat{\theta},\widetilde{\theta})}(\theta) = 1 \text{ at } \theta &\iff \mathrm{var}_\theta(\widehat{\theta}) = \mathrm{var}_\theta(\widetilde{\theta}) \text{ at } \theta \end{aligned}$$

   So $\widehat{\theta} \in \mathrm{UE}(\theta)$ dominates $\widetilde{\theta} \in \mathrm{UE}(\theta)$ iff $e_{(\widehat{\theta},\widetilde{\theta})}(\theta) \ge 1$ for all θ.

2. Asymptotically efficient estimators

   (1) Asymptotically efficient estimators
   $\widehat{\theta}_n$ is an asymptotically efficient for $\theta \stackrel{def}{\iff} \widehat{\theta}_n \stackrel{p}{\longrightarrow} \theta$ and $\mathrm{Cov}(\sqrt{n}\,\widehat{\theta}_n) \longrightarrow I^{-1}(\theta)$.

   Comments: From efficient estimator to asymptotically efficient estimator the condition $E_\theta(\widehat{\theta}_n) = \theta$ changed to $\widehat{\theta}_n \stackrel{p}{\longrightarrow} \theta$ and the condition $\mathrm{Cov}_\theta(\widehat{\theta}_n) = \mathrm{CRLB}(\theta)$ changed to $n\,\mathrm{Cov}_\theta(\widehat{\theta}_n) \longrightarrow I^{-1}(\theta)$.
   The second condition can not be $\mathrm{Cov}(\widehat{\theta}_n) \longrightarrow \mathrm{CRLB}(\theta)$ since $\mathrm{CRLB}(\theta) = [nI(\theta)]^{-1}$ depends on n. So we multiply both sides by n to cancel it on the right-hand side, and the left-hand side becomes $n\,\mathrm{Cov}(\widehat{\theta}_n) = \mathrm{Cov}(\sqrt{n}\ \widehat{\theta}_n)$.

   (2) Asymptotically efficient estimator for $\theta \in R$

$$\begin{aligned} \text{With } \theta \in R \qquad & \lim_{n\to\infty}\left[\mathrm{var}_\theta(\sqrt{n}\ \widehat{\theta}_n)\right] = I^{-1}(\theta) \text{ for all } \theta \\ \iff\ & \lim_{n\to\infty}\left[n \cdot \mathrm{var}_\theta(\widehat{\theta}_n) \cdot I(\theta)\right] = 1 \text{ for all } \theta \\ \iff\ & \lim_{n\to\infty} \frac{1}{n\cdot \mathrm{var}_\theta(\widehat{\theta}_n)\cdot I(\theta)} = 1 \text{ for all } \theta \\ \iff\ & \lim_{n\to\infty} \frac{[nI(\theta)]^{-1}}{\mathrm{var}_\theta(\widehat{\theta}_n)} = 1 \text{ for all } \theta \\ \iff\ & \lim_{n\to\infty} e_{\widehat{\theta}_n}(\theta) \longrightarrow 1 \text{ for all } \theta. \end{aligned}$$

So $\widehat{\theta}_n$ is an asymptotically efficient estimator for $\theta \in R$ iff $\widehat{\theta}_n$ is a consistent estimator for θ and $e_{\widehat{\theta}_n}(\theta) \longrightarrow 1$ for all θ.

(3) Chebyshev inequality

Suppose random variable $X \geq 0$. Then $P(X > \epsilon) \leq \frac{E(X)}{\epsilon}$.

Proof. Let $f(x)$ be the pdf of X. Then

$P(X > \epsilon) = \int_{x>\epsilon} f(x)dx \leq \int_{x>\epsilon} \frac{x}{\epsilon} f(x)dx \leq \int_r \frac{x}{\epsilon} f(x)dx = \frac{E(X)}{\epsilon}$.

(4) Theorem

If $\widehat{\theta}_n$ is efficient estimator for $\theta \in R^k$, then $\widehat{\theta}_n$ is asymptotically efficient estimator for θ.

Proof. Note that $E_\theta(\widehat{\theta}_n) = \theta$ and $\text{Cov}_\theta(\widehat{\theta}_n) = \text{CRLB}(\theta) = \frac{I^{-1}(\theta)}{n}$.

$$\begin{aligned}
\forall\, \epsilon > 0,\ P(\|\widehat{\theta}_n - \theta\| > \epsilon) &= P(\|\widehat{\theta}_n - \theta\|^2 > \epsilon^2) \leq \frac{E\big(\|\widehat{\theta}_n - \theta\|^2\big)}{\epsilon^2} \\
&= \frac{E\big\{\text{tr}\big[(\widehat{\theta}_n-\theta)'(\widehat{\theta}_n-\theta)\big]\big\}}{\epsilon^2} = \frac{\text{tr}\big\{E\big[(\widehat{\theta}_n-\theta)(\widehat{\theta}_n-\theta)'\big]\big\}}{\epsilon^2} \\
&= \frac{\text{tr}\big[\text{Cov}(\widehat{\theta}_n)\big]}{\epsilon^2} = \frac{\text{tr}\big(I^{-1}/n\big)}{\epsilon^2} \\
&= \frac{\text{tr}(I^{-1}(\theta))}{n\epsilon^2} \longrightarrow 0 \text{ as } n \longrightarrow \infty.
\end{aligned}$$

Thus $\widehat{\theta}_n \xrightarrow{p} \theta$.

$\text{Cov}(\widehat{\theta}_n) = \text{CRLB}(\theta) = \frac{I^{-1}(\theta)}{n} \Longrightarrow n \cdot \text{Cov}_\theta(\widehat{\theta}_n) = I^{-1}(\theta)$.

So $\lim_{n\to\infty} \Big[\text{Cov}_\theta(\sqrt{n}\ \widehat{\theta}_n)\Big] = I^{-1}(\theta)$.

Hence $\widehat{\theta}_n$ is asymptotically efficient estimator for θ.

3. More on asymptotically efficient estimators

(1) Theorem

If $E_\theta(\widehat{\theta}_n) = \theta$ and $\lim_{n\to\infty}[\text{Cov}_\theta(\sqrt{n}\ \widehat{\theta}_n)] = I^{-1}(\theta)$, then $\widehat{\theta}_n$ is asymptotically efficient estimator for θ.

Proof. Left as HW.

(2) Example

$\widehat{\theta}_n = \begin{pmatrix} \overline{X}_n \\ S_n^2 \end{pmatrix}$ is BLUE for $\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$ in $N(\mu, \sigma^2)$ since $E_\theta(\widehat{\theta}_n) = \theta$ and $\widehat{\theta}_n$ is a function of sufficient and complete statistic $S = \begin{pmatrix} \sum X_i \\ \sum X_i^2 \end{pmatrix}$

$\widehat{\theta}_n$ is not an efficient estimator for θ since

$$\text{Cov}_\theta(\widehat{\theta}_n) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{pmatrix} \neq \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix} = \text{CRLB}(\theta).$$

$\widehat{\theta}_n$ is an asymptotically efficient estimator for θ since $E_\theta(\widehat{\theta})n) = \theta$ and

$$\lim_{n\to\infty} [\text{Cov}_\theta(\sqrt{n}\ \widehat{\theta}_n)] = \lim_{n} n \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} = I^{-1}(\theta).$$

L05 Convergence in distributions

1. Convergence in distributions

   (1) Definition
   Let $F_n(x) = P(X_n \leq x$ component wise) and $F(x) = P(X \leq x$ component wise) be the cumulative distribution functions of X_n and X. X_n converges to X in distributions denoted as $X_n \xrightarrow{d} X$ if $F_n(x) \longrightarrow F(x)$ at all continuity point x for $F(x)$.

   (2) Sufficient and necessary condition
   $X_n \xrightarrow{d} X \Longleftrightarrow P(X_n \in G) \longrightarrow P(X \in G)$ for all open sets G.

   (3) Usage
   If $X_n \xrightarrow{d} X$ and X is a continuous random vector with continuous cdf $F(x)$, then $P(X_n \in G) \longrightarrow P(X \in G)$ for all G, $E(X_n) \longrightarrow E(X)$ and $\text{Cov}(X_n) \longrightarrow \text{Cov}(X)$.
   So $P(X_n \in G) \approx P(X \in G)$, $E(X_n) \approx E(X)$ and $\text{Cov}(X_n) \approx \text{Cov}(X)$.

2. Relations

   (1) From convergence in probability to that in distribution
   $X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X$.

   (2) For $X_n \xrightarrow{d} X$ there is no base to consider $X_n \xrightarrow{p} X$.
   $X_n \xrightarrow{p} X \Longleftrightarrow P(\|X_n - X\| > \epsilon) \longrightarrow 0$ for all ϵ. So X and X_i must be defined in the same probability space for $i = 1, 2, ...$
   But with $X_n \xrightarrow{d} X \Longleftrightarrow P(X_n \in G) \longrightarrow P(X \in G)$, $X, X_1, X_2, ...$ could be in different probability space so that there is no base to consider $P(\|X_n - X\| > \epsilon)$.

   (3) A spacial case
   With non-random C, $\quad X_n \xrightarrow{p} C \Longleftrightarrow X_n \xrightarrow{d} C$.
   Non-random C can be regarded as a variable defined in every probability space so that even if $X_1, ..., X_n$ are in the different probability spaces, but $P(\|X_n - C\| > \epsilon)$ can be calculated.

   (4) Skorokhod representation
   For $X_n \xrightarrow{d} X$, using inverse of cdf and uniform distributions one can create $Y, Y_1, ...$ in a probability space such that $Y_n \overset{d}{=} X_n$ and $Y \overset{d}{=} X$ and $Y_n \xrightarrow{a.s.} Y$.

3. Properties

   (1) $X_n \xrightarrow{d} X \Longleftrightarrow g(X_n) \xrightarrow{d} g(X)$ for all continuous $g(\cdot)$
   ⇒: By Skrokohod representation,
   If $X_n \xrightarrow{d} X$, then there exist Y_n and Y such that $X_n \overset{d}{=} Y_n \xrightarrow{a.s.} Y \overset{d}{=} X$.
   So $g(X_n) \overset{d}{=} g(Y_n) \xrightarrow{a.s.} g(Y) \overset{d}{=} g(X)$. Thus $g(X_n) \xrightarrow{d} g(X)$.
   ⇐: Take $g(x) = x$.
   Comment: The above property is shared by both almost sure convergence and the convergence in probability. But here is a special $g(\cdot)$,

$$X_n \xrightarrow{d} X \Longleftrightarrow \alpha' X_n \xrightarrow{d} \alpha' X \text{ for all vector } \alpha.$$

(2) $X_n \xrightarrow{d} X \iff X_{n_k} \xrightarrow{d} X$ for all subsequence X_{n_k}.

Comment: This property is also common for the convergence w.p.1 and in probability.

(3) $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ Y \end{pmatrix} \Longrightarrow X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$

Proof. $X_n = g(X_n \, Y_n)$ is a continuous function. Conclusion follows from (1).

Comment: This property is common for convergence wp1 and in probability.

(4) For $\begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ and $\begin{pmatrix} X \\ Y \end{pmatrix}$, $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ may not imply $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ Y \end{pmatrix}$.

Example: For $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 8 & 2 \\ 2 & 5 \end{pmatrix}\right)$ and $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 8 & 0 \\ 0 & 5 \end{pmatrix}\right)$,

$X_n \sim N(0, 8) \xrightarrow{d} X \sim N(0, 8)$ and $Y_n \sim N(0, 5) \xrightarrow{d} Y \sim N(0, 5)$.

But $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \not\xrightarrow{d} \begin{pmatrix} X \\ Y \end{pmatrix}$.

(5) Slutsky theorem

For $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \in R^2$ and $\begin{pmatrix} X \\ C \end{pmatrix} \in R^2$, if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} C$, then $X_n + Y_n \xrightarrow{d} X + C$.

(6) With $\begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ and $\begin{pmatrix} X \\ C \end{pmatrix}$, if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} C$, then $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ C \end{pmatrix}$.

Proof: By the sufficient and necessary condition in the comment after (1),

we need to show $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} X \\ C \end{pmatrix}$ for all vector $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

$X_n \xrightarrow{d} X \Longrightarrow \alpha' X_n \xrightarrow{d} \alpha' X$ for all vector α

$Y_n \xrightarrow{d} C \Longrightarrow \beta' Y_n \xrightarrow{d} \beta' C$ for all vector β

By Slutsky theorem, $\alpha' X_n + \beta' Y_n \xrightarrow{d} \alpha' X + \beta' C$.

Thus $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} X \\ C \end{pmatrix}$ for all vector $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

.

Corollary: $g(X_n \, Y_n) \xrightarrow{d} g(X, C)$ for all continuous $g(\cdot, \cdot)$.

Ex: If $X_n \xrightarrow{d} X$, show that $\frac{2n}{n+k} X_n \xrightarrow{d} 2X$.

Proof. $X_n \xrightarrow{d} X$ and $\frac{2n}{n+k} \longrightarrow 2$. So $\begin{pmatrix} X_n \\ \frac{2n}{n+k} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ 2 \end{pmatrix}$. Hence $\frac{2n}{n+k} X_n \xrightarrow{d} 2X$.