L21: Central limit theorem and best asymptotically normal estimator

1. Best asymptotically normal estimator

   (1) Best normal estimator

   $$\begin{aligned} & \widehat{\theta}_n \text{ is a best normal estimator for } \theta \overset{def}{\Longleftrightarrow} \widehat{\theta}_n \sim N(\theta,\, \text{CRLB}(\theta)) \\ \Longleftrightarrow\ & \widehat{\theta}_n \text{ is normally distributed; } E(\widehat{\theta}_n) = \theta; \text{ and } \text{Cov}(\widehat{\theta}) = \text{CRLB}(\theta) \\ \Longleftrightarrow\ & \widehat{\theta}_n \text{ is a normal unbiased estimator for } \theta \text{ and it is an efficient statistic} \\ \Longleftrightarrow\ & \widehat{\theta}_n \text{ is a normal and efficient estimator for } \theta \\ \Longrightarrow\ & \widehat{\theta}_n \text{ is MVUE for } \theta. \end{aligned}$$

   (2) Best asymptotically normal estimator

   $\widehat{\theta}_n$ is a best asymptotically normal estimator for $\theta \overset{def}{\Longleftrightarrow} \widehat{\theta}_n \sim AN(\theta,\, \text{CRLB}(\theta))$.

   (3) Properties

   If $\widehat{\theta}_n$ is a best asymptotically normal estimator for θ, then

   (i) $\widehat{\theta}_n$ is asymptotically normal,

   (ii) $\widehat{\theta}_n$ is a consistent estimator for θ,

   (iii) $n[\text{Cov}(\widehat{\theta}_n) - \text{CRLB}(\theta)] \longrightarrow 0$.

   Proof: (ii) $\widehat{\theta}_n \sim AN(\theta,\, \text{CRLB}(\theta)) = AN\left(\theta,\, \frac{I^{-1}(\theta)}{n}\right) \Longrightarrow \sqrt{n}(\widehat{\theta}_n - \theta) \sim AN(0,\, I^{-1}(\theta))$

   $\Longrightarrow \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0,\, I^{-1}(\theta)) \Longrightarrow \sqrt{n}(\widehat{\theta}_n - \theta) = O_p(1)$.

   So $\widehat{\theta}_n - \theta = \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\widehat{\theta}_n - \theta) = o_p(1)O_p(1) = o_p(1)$. Thus $\widehat{\theta}_n \xrightarrow{p} \theta$.

   (iii) $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0,\, I^{-1}(\theta)) \Longrightarrow \text{Cov}\left(\sqrt{n}(\widehat{\theta}_n - \theta\right) \longrightarrow I^{-1}(\theta)$

   $\Longrightarrow n\text{Cov}(\widehat{\theta}_n) - I^{-1}(\theta) \longrightarrow 0 \Longrightarrow n \cdot [\text{Cov}(\widehat{\theta}_n) - \text{CRLB}(\theta)] \longrightarrow 0$.

   Comment: Best asymptotically normal estimator for θ is an asymptotically normal and asymptotically efficient estimator for θ.

2. Central limit theorem

   (1) Central Limit Theorem

   Suppose $X_1, ..., X_n$ are iid with $X \sim (\theta,\, \Sigma)$. Then $\frac{X_1 + \cdots + X_n}{n} \sim AN\left(\theta,\, \frac{1}{n}\Sigma\right)$.

   Write $\frac{X_1 + \cdots + X_n}{n}$ as $\overline{X}_n$. Then CLT says $\overline{X}_n \sim AN\left(\theta,\, \frac{1}{n}\Sigma\right)$, i.e.,

   $$\overline{X}_n - \theta \sim AN(0,\, \frac{1}{n}\Sigma) \Longleftrightarrow \sqrt{n}(\overline{X}_n - \theta) \sim AN(0,\, \Sigma) \Longrightarrow \sqrt{n}(\overline{X}_n - \theta) \xrightarrow{d} N(0,\, \Sigma).$$

   (2) A lemma

   If $X_1, ..., X_n$ are iid with $X \sim (\theta,\, \Sigma)$ where $\Sigma = I^{-1}(\theta)$, then $\overline{X}_n$ is a best asymptotically normal estimator for θ.

   Proof: By CLT, $\overline{X}_n \sim AN\left(\theta,\, \frac{1}{n}\Sigma\right) = AN\left(\theta,\, \frac{1}{n}I^{-1}(\theta)\right) = AN(\theta,\, \text{CRLB}(\theta))$.
   Thus $\overline{X}_n$ is the best asymptotically normal estimator for θ. □

   Ex1: For $X \sim \text{Pisson}(\lambda) \sim (\lambda,\, \lambda)$. It can be shown that $I(\lambda) = \frac{1}{\lambda} = [\text{var}(X)]^{-1}$.
   Thus the sample mean is a best asymptotically normal estimator for λ.

3. Some transformations

(1) $X_n \sim AN(\mu_n, a_n^2\Sigma)$ and $\frac{a_n}{b_n} \longrightarrow 1$, then $X_n \sim AN(\mu_n, b_n^2\Sigma)$.

Proof: $X_n \sim AN(\mu_n, a_n^2\Sigma) \Longrightarrow \frac{X_n-\mu_n}{a_n} \sim AN(0, \Sigma) \Longrightarrow \frac{X_n-\mu_n}{a_n} \xrightarrow{d} X \sim N(0, \Sigma)$.

With $\frac{a_n}{b_n} \longrightarrow 1$, $Y_n = \begin{pmatrix} \frac{X_n-\mu_n}{a_n} \\ \frac{a_n}{b_n} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ 1 \end{pmatrix} = Y$.

So $\frac{X_n-\mu_n}{a_n} \cdot \frac{a_n}{b_n} = g(Y_n) \xrightarrow{d} g(Y) = X \cdot 1 = X \sim N(0, \Sigma)$.

Thus $\frac{X_n-\mu_n}{b_n} \sim AN(0, \Sigma)$. Hence $X_n \sim AN(\mu_n, b_n^2\Sigma)$.

(2) $X_n \sim AN(\mu_n, a_n^2\Sigma)$ and $\frac{\mu_n-\nu_n}{a_n} \longrightarrow 0$, then $X_n \sim AN(\nu_n, a_n^2\Sigma)$.

Proof: $X_n \sim AN(\mu_n, a_n^2\Sigma) \Longrightarrow \frac{X_n-\mu_n}{a_n} \sim AN(0, \Sigma) \Longrightarrow \frac{X_n-\mu_n}{a_n} \xrightarrow{d} X \sim N(0, \Sigma)$.

With $\frac{\mu_n-\nu_n}{a_n} \longrightarrow 0$, $Y_n = \begin{pmatrix} \frac{X_n-\mu_n}{a_n} \\ \frac{\mu_n-\nu_n}{a_n} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ 0 \end{pmatrix} = Y$.

So $\frac{X_n-\mu_n}{a_n} + \frac{\mu_n-\nu_n}{a_n} = g(Y_n) \xrightarrow{d} g(Y) = X + 0 = X \sim N(0, \Sigma)$.

Thus $\frac{X_n-\nu_n}{a_n} \sim AN(0, \Sigma)$. Hence $X_n \sim AN(\nu_n, a_n^2\Sigma)$.

(3) Transformation $A_nX_n + \beta_n$ with random A_n and β_n

$X_n \sim AN(\mu, \Sigma)$, $A_n \xrightarrow{p} A$ and $\beta_n \xrightarrow{p} \beta$ where A and β are non-random. Then $A_nX_n + \beta_n \sim AN(A\mu + \beta, A\Sigma A')$.

Pf: $X_n \sim AN(\mu, \Sigma) \Longrightarrow X_n \xrightarrow{d} X \sim N(\mu, \Sigma)$.

With $\begin{pmatrix} \text{vec}(A_n) \\ \beta_n \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \text{vec}(A) \\ \beta \end{pmatrix}$, $Y_n = \begin{pmatrix} X_n \\ \text{vec}(A_n) \\ \beta_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ \text{vec}(A) \\ \beta \end{pmatrix} = Y$.

So $A_nX_n + \beta_n = g(Y_n) \xrightarrow{d} g(Y) = AX + \beta \sim N(A\mu + \beta, A\Sigma A')$.

Thus $A_nX_n + \beta_n \sim AN(A\mu + \beta, A\Sigma A')$.