L06 Asymptotically normal distributions

1. Asymptotically normal distribution

   (1) Definition
   X_n has asymptotically normal distribution with mean μ and variance-covariance matrix Σ denoted as $X_n \sim AN(\mu, \Sigma)$ if $X_n \xrightarrow{d} X \sim N(\mu, \Sigma)$.
   Thus $X_n \sim AN(\mu, \Sigma) \Longrightarrow X_n = O_p(1)$, $E(X_n) \longrightarrow \mu$ and $\text{Cov}(X_n) \longrightarrow \Sigma$.

   (2) Definition
   For $X_n \sim AN(\mu_n \, \Sigma_n)$ write $Y_n = A_n X_n + b_n \sim AN(A_n \mu_n + b_n,\ A_n \Sigma_n A_n')$.
   $$\begin{aligned} \text{So } X_n \sim AN(\mu_n, \Sigma_n) \quad &\Longrightarrow \quad \Sigma_n^{-1/2}(X_n - \mu_n) \sim AN(0, I) \\ &\Longrightarrow \quad \Sigma_n^{-1/2}(X_n - \mu_n) \xrightarrow{d} X \sim N(0,\ I). \end{aligned}$$
   Thus $E[\Sigma_n^{-1/2}(X_n - \mu_n)] \longrightarrow 0$ and $\text{Cov}[\Sigma_n^{-1/2}(X_n - \mu_n)] \Longrightarrow I$.

2. Asymptotically normal estimators

   (1) Asymptotically normal estimators
   Let $\widehat{\theta}_n \sim AN\left(\theta, \frac{\Sigma}{a_n}\right)$ with $a_n \longrightarrow \infty$ be an estimator for θ. Then this estimator is called an asymptotically normal estimator. This estimator is a consistent estimator for θ.

   Proof. $\widehat{\theta}_n \sim AN\left(\theta, \frac{\Sigma}{a_n}\right) \Longrightarrow \sqrt{a_n}(\widehat{\theta}_n - \theta) \sim AN(0, \Sigma) \Longrightarrow \sqrt{a_n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, I)$
   $\Longrightarrow \sqrt{a_n}(\widehat{\theta}_n - \theta) = O_p(1)$.
   But $\widehat{\theta}_n - \theta = \frac{1}{\sqrt{a_n}}[\sqrt{a_n}(\widehat{\theta}_n - \theta)] = o_p(1)O_p(1) = o_p(1)$.
   So $\widehat{\theta}_n - \theta \xrightarrow{p} 0$, i.e, $\widehat{\theta}_n \xrightarrow{p} \theta$. Hence $\widehat{\theta}_n$ is a consistent estimator.

   (2) Best asymptotically normal estimator
   $\widehat{\theta}_n$ is a best asymptotically normal estimator for θ if $\widehat{\theta}_n \sim AN(\theta, \text{CRLB}(\theta))$. The best asymptotically normal estimator is an asymptotically efficient estimator.

   Proof. $\widehat{\theta}_n \sim AN(\theta, \text{CRLB}(\theta)) = AN\left(\theta, \frac{I^{-1}(\theta)}{n}\right)$.
   By (1) $\widehat{\theta}_n$ is a consistent estimator for θ.
   $$\begin{aligned} \widehat{\theta}_n \sim AN\left(\theta, \tfrac{I^{-1}}{n}\right) \quad &\Longrightarrow \quad \sqrt{n}(\widehat{\theta}_n - \theta) \sim AN(0, I^{-1}(\theta)) \\ &\Longrightarrow \quad \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, I^{-1}(\theta)) \\ &\Longrightarrow \quad \text{Cov}[\sqrt{n}(\widehat{\theta}_0 - \theta)] \longrightarrow I^{-1}(\theta). \end{aligned}$$
   So $\widehat{\theta}_n$ is an asymptotically efficient estimator for θ.

   (3) Central limit theorem
   $X_1, X_2, \cdots$ are iid with (μ, Σ). Then $\frac{X_1 + \cdots + X_n}{n} \sim AN\left(\mu \, \frac{\Sigma}{n}\right)$.

   Comment 1: Sample mean is an asymptotically normal estimator for population mean μ. This estimator is unbiased estimator, so it is asymptotically unbiased. By (1), this estimator is also consistent.

   Comment 2: Sample mean is a best asymptotically normal estimator for population mean if and only if $\Sigma = I^{-1}(\theta)$.

3. An example
In 7.5 p208, X has pdf $f(x;\,\theta) = \frac{1}{\sqrt{2\pi}\,x}\, e^{-\frac{1}{2}(\ln x-\theta)^2}$, $x > 0$, with parameter θ.

(1) Find Fisher information $I(\theta)$.

We notice that x appears in the pdf in $\ln x$. Let $Y = \ln X$. The the pdf for Y is

$$f_Y(y;\,\theta) = f(x;\,\theta)\left|\frac{dx}{dy}\right| = \frac{1}{\sqrt{2\pi}\,x} e^{-\frac{1}{2}(\ln x-\theta)^2}\, e^y = \frac{1}{\sqrt{2\pi}\,e^y} e^{-\frac{1}{2}(y-\theta)^2}\, e^y = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(y-\theta)^2}$$

where $-\infty < y < \infty$. So $Y \sim N(\theta,\, 1^2)$.

$f_Y(y;\,\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2}$ $\qquad$ $\ln f_Y(y;\,\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}(y-\theta)^2$

$\frac{\partial}{\partial\theta}\ln f_y(y;\,\theta) = y - \theta$ $\qquad$ $\frac{\partial}{\partial\theta}\ln f_Y(Y;\,\theta) = Y - \theta \sim N(0,\, 1^2)$ $\qquad$ So $I(\theta) = 1$.

(2) Find an asymptotically normal estimator for θ.

$X_1, ..., X_n$ is a random sample from X. So $Y_i = \ln X_i$, $i = 1, ..., n$, is a random sample from Y with mean θ. Thus mean of a sample from Y is an asymptotically normal estimator for θ.

$$\widehat{\theta}_n = \overline{y} = \frac{y_1 + \cdots + y_n}{n} = \frac{\ln x_1 + \cdots + \ln x_n}{n} = \frac{\ln(x_1 \cdots x_n)}{n} = \ln(x_1 \cdots x_n)^{1/n},$$

and

$$\widehat{\theta}_n \sim AN\left(\theta,\, \frac{1}{n}\right).$$

(3) Show that $\widehat{\theta}_n$ is best asymptotically normal estimator for θ.

$\widehat{\theta}_n = \overline{Y}_n \sim AN\left(\theta,\, \frac{1}{n}\right)$ by CLT.
From 1, $\text{CRLB}(\theta) = [nI(\theta)]^{-1} = n^{-1} = \frac{1}{n}$.
So $\widehat{\theta}_n \sim AN(\theta,\, \text{CRLB}(\theta))$.
Hence $\widehat{\theta}_n$ is best asymptotically normal estimator for θ.

L07 Delta method

1. Taylor expansion of $f(x) \in R^q$ with $x \in R^p$

   (1) Taylor expansion with remainder

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_q(x) \end{pmatrix} = \begin{pmatrix} f_1(\theta) + \left[\frac{\partial f_1(\theta)}{\partial \theta^T}\right](x-\theta) + R_1 \\ \vdots \\ f_q(\theta) + \left[\frac{\partial f_q(\theta)}{\partial \theta^T}\right](x-\theta) + R_q \end{pmatrix}$$
$$= f(\theta) + \left[\frac{\partial f(\theta)}{\partial \theta^T}\right](x-\theta) + R$$
$$= f(\theta) + D(\theta)(x-\theta) + R$$

   where $D(\theta) = \frac{\partial f(\theta)}{\partial \theta^T} \in R^{q\times p}$

   (2) Taylor expansion without remainder

$$f(x) = f(\theta) + D(\xi)(x-\theta)$$

   where $\xi = \lambda x + (1-\lambda)\theta$ for some $\lambda \in (0,\, 1)$.

2. Delta method

   (1) Theorem
   $X_n \in R^p$, $X_n \sim AN\left(\theta,\, \frac{\Sigma}{a_n}\right)$ where $a_n \longrightarrow \infty$ as $n \longrightarrow \infty$.
   $g(x) \in R^q$ is a continuous vector-valued function of $x \in R^p$ with continuous matrix-valued $D(x) = \frac{\partial g(x)}{\partial x^T} \in R^{q\times p}$. Then

$$g(X_n) \sim AN\left(g(\theta),\, D(\theta)\frac{\Sigma}{a_n}D^T(\theta)\right)$$

   (2) Analysis
   $X_n \sim AN\left(\theta,\, \frac{\Sigma}{a_n}\right) \Longleftrightarrow \sqrt{a_n}(X_n - \theta) \xrightarrow{d} N(0,\, \Sigma)$.
   $g(X_n) \sim AN\left(g(\theta),\, D(\theta)\frac{\Sigma}{a_n}D^T(\theta)\right) \Longleftrightarrow \sqrt{a_n}[g(X_n) - g(\theta)] \xrightarrow{d} N\left(0,\, D(\theta)\Sigma D^T(\theta)\right)$
   By Taylor expansion $g(X_n) = g(\theta) + D(\xi_n)(X_n - \theta)$,, $g(X_n) - g(\theta) = D(\xi_n)(X_n - \theta)$.
   Thus under $\sqrt{a_n}(X_n - \theta) \xrightarrow{d} N(0,\, \Sigma)$ we need to show

$$D(\xi_n)[\sqrt{a_n}(X_n - \theta)] \xrightarrow{d} N(0,\, D(\theta)\Sigma D^T(\theta))$$

3. Proof

   (1) $X_n \xrightarrow{p} \theta$
   $\sqrt{a_n}(X_n - \theta) \xrightarrow{d} N(0,\, \Sigma) \Longrightarrow \sqrt{a_n}(X_n - \theta) = O_p(1)$
   $a_n \longrightarrow \infty \Longrightarrow \frac{1}{\sqrt{a_n}} = o_p(1)$
   So $X_n - \theta = \frac{1}{\sqrt{a_n}}[\sqrt{a_n}(X_n - \theta)] = o_p(1)$.
   $X_n \xrightarrow{p} \theta$ follows.

(2) $\xi_n \xrightarrow{p} \theta$

$\xi_n = \lambda_n X_n + (1-\lambda_n)\theta \Longrightarrow \xi_n - \theta = \lambda_n(X_n - \theta)$.

Here $\lambda_n = O_p(1)$ and $X_n - \theta = o_p(1)$. So $\xi_n = o_p(1)$.

$\xi_n \xrightarrow{p} \theta$ follows.

(3) $D(\xi_n) \xrightarrow{p} D(\theta)$

$\xi_n \xrightarrow{p} \theta$ and $D(x)$ is a matrix-valued continuous function of x.

Thus $D(\xi_n) \xrightarrow{p} D(\theta)$.

(4) $D(\xi_n)[\sqrt{a_n}(X_n - \theta)] \xrightarrow{d} N(0,\, D(\theta)\Sigma D^T(\theta))$.

$$\begin{cases} \sqrt{a_n}(X_n - \theta) \xrightarrow{d} X \sim N(0,\,\Sigma) \\ D(\xi_n) \xrightarrow{p} D(\theta) \end{cases} \Longrightarrow \begin{pmatrix} \sqrt{a_n}(X_n-\theta) \\ \text{vec}[D(\xi_n)] \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ \text{vec}[D(\theta)] \end{pmatrix}.$$

$$\Longrightarrow D(\xi_n)[\sqrt{a_n}(X_n - \theta)] \xrightarrow{d} D(\theta)X \sim N(0,\, D(\theta)\Sigma D^T(\theta))$$

Ex1: From population $X \sim (\mu,\,\Sigma)$ where $\mu \in R^2$, by CLT $\overline{X}_n \sim AN\left(\mu,\, \frac{\Sigma}{n}\right)$.

(i) For $\theta = g(\mu) = \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1^2 \end{pmatrix}$, $D(\mu) = \begin{pmatrix} 1 & 1 \\ 2\mu_1 & 0 \end{pmatrix}$.

So $\begin{pmatrix} (\overline{X}_n)_1 + (\overline{X}_n)_2 \\ (\overline{X}_n)_1^2 \end{pmatrix} \sim AN\left(\theta,\, \frac{1}{n}\begin{pmatrix} 1 & 1 \\ 2\mu_1 & 0 \end{pmatrix} \Sigma \begin{pmatrix} 1 & 2\mu_1 \\ 1 & 0 \end{pmatrix}\right)$.

(ii) For $\theta = g(\mu) = \mu_1 + \mu_2^2$, $D(\mu) = (1,\, 2\mu_2)$.

So $(\overline{X}_n)_1 + (\overline{X}_n)_2^2 \sim AN\left(\theta,\, \frac{1}{n}(1,\, 2\mu_2)\Sigma \begin{pmatrix} 1 \\ 2\mu_2 \end{pmatrix}\right)$.

Ex2: From population $X \sim (\mu,\, \sigma^2)$, by CLT, $\overline{X}_n \sim AN\left(\mu,\, \frac{\sigma^2}{n}\right)$.

For $\theta = g(\mu) = \mu + \mu^2$, $D(\mu) = 1 + 2\mu$. So $\overline{X}_n + \overline{X}_n^2 \sim AN\left(\theta,\, \frac{1}{n}(1+2\mu)^2\sigma^2\right)$.