STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Jan 25

Lecturer: Xianyang Zhang

1 Central Limit Theorem

For each n, let $Y_{n,1}, Y_{n,2}, \cdots, Y_{n,k_n}$ be independent random vectors with finite covariances. If

1. $\sum_{i=1}^{k_n} \mathbb{E}||Y_{n,i}||^2 \mathbf{1}\{||Y_{n,i}|| > \epsilon\} \to 0 \ \forall \ \epsilon > 0$,
2. $\sum_{i=1}^{k_n} \mathrm{Cov}(Y_{n,i}) \to \Sigma$.

Then

$$\sum_{i=1}^{k_n} (Y_{n,i} - \mathbb{E}Y_{n,i}) \xrightarrow{d} N(0, \Sigma).$$

2 Application to Linear Regression

Let $\beta = (\beta_1, \beta_2, \cdots, \beta_p)^\top \in \mathbb{R}^p$ be an unknown vector and

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ \vdots & \vdots & \vdots & \ddots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \in \mathbb{R}^{n \times p}$$

be a known matrix of full rank. Consider the linear model

$$Y = \mathbf{X}\beta + \mathbf{e},$$

where the error vector $\mathbf{e} = (e_1, e_2, \cdots, e_n)^\top$ has i.i.d components with zero mean and variance σ^2. The least squares estimate $\hat{\beta}$ of β is given by

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y.$$

The estimator is unbiased (i.e., $\mathbb{E}\hat{\beta} = \mathbb{E}\beta$) and $\hat{\beta}$ has a covariance matrix $\sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$. Some algebra yields that

$$(\mathbf{X}^\top \mathbf{X})^{\frac{1}{2}} (\hat{\beta} - \beta) = (\mathbf{X}^\top \mathbf{X})^{-\frac{1}{2}} \mathbf{X}^\top e.$$

Let $\mathbf{A} = (\mathbf{X}^\top \mathbf{X})^{-\frac{1}{2}} \mathbf{X}^\top = [a_{n1} \ a_{n2} \cdots a_{nn}]$, where $a_{ni} \in \mathbb{R}^p$. Observe that

$$\sum_{i=1}^{n} ||a_{ni}||^2 = \mathrm{tr}(\mathbf{A}^\top \mathbf{A}) = \mathrm{tr}(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) = \mathrm{tr}(\mathbf{I}_p) = p.$$

The quantity $(\mathbf{X}^\top \mathbf{X})^{\frac{1}{2}} (\hat{\beta} - \beta)$ can be re-written as

$$(\mathbf{X}^\top \mathbf{X})^{\frac{1}{2}} (\hat{\beta} - \beta) = \sum_{i=1}^{n} a_{ni} e_i.$$

We find sufficient conditions under which the above sequence is asymptotically normal. To this end, we shall require

$$\sum_{i=1}^{n} \mathbb{E}||a_{ni} e_i||^2 \mathbf{1}\{||a_{ni} e_i|| > \epsilon\} \to 0.$$

This quantity can be upper-bounded as

$$\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}||a_{ni}e_i||^2\mathbf{1}\{||a_{ni}e_i||>\epsilon\}\\
=&\sum_{i=1}^{n}||a_{ni}||^2\mathbb{E}|e_i|^2\mathbf{1}\{||a_{ni}|||e_i|>\epsilon\}\\
\leq&\max_i\mathbb{E}|e_i|^2\mathbf{1}\{||a_{ni}|||e_i|>\epsilon\}\sum_{i=1}^{n}||a_{ni}||^2\\
\leq&\mathbb{E}|e_1|^2\mathbf{1}\{\max_i||a_{ni}|||e_1|>\epsilon\}\sum_{i=1}^{n}||a_{ni}||^2\\
=&p\mathbb{E}|e_1|^2\mathbf{1}\{\max_i||a_{ni}|||e_1|>\epsilon\}.
\end{aligned}$$

If $\mathbb{E}|e_1|^2<\infty$, then it suffices to have

$$\mathbf{1}\{\max_i||a_{ni}|||e_1|>\epsilon\}\to^{a.s}0,$$

which happens when

$$\max_i||a_{ni}||\to 0 \text{ as } n\to\infty.$$

3 Delta Method

Let $\{X_n\},\{Y_n\},\{Z_n\}$ be sequences of random variables such that

$$\begin{aligned}
&\frac{X_n-Y_n}{Z_n}\xrightarrow{d}T,\\
&X_n\xrightarrow{p}a,\\
&Z_n\xrightarrow{p}0,
\end{aligned}$$

where a is a constant. Let $f:\mathbb{R}\to\mathbb{R}$ be a function that is differentiable at a. Then we have

$$\frac{f(X_n)-f(Y_n)}{Z_n}\xrightarrow{d}f'(a)T.$$

4 Differentiability of Multi-variate Function

A function $f:\mathbb{R}^k\to\mathbb{R}^m$ is differentiable at θ if there exists a matrix $\mathbf{D}_\theta\in\mathbb{R}^{m\times k}$ such that

$$f(\theta+h)-f(\theta)=\mathbf{D}_\theta h+o(||h||),$$

for $h\in\mathbb{R}^k$. Here $\mathbf{D}_\theta$ is called the gradient when $m=1$.

5 Multi-variate Delta Method

Let $f:\mathbb{R}^k\to\mathbb{R}^m$ be a vector-valued function that is differentiable at θ. Suppose the random vectors T_n takes values in the domain of f. If there are numbers $\{r_n\}$ such that $r_n\to\infty$ and

$$r_n(T_n-\theta)\xrightarrow{d}T,$$

as $n\to\infty$. Then we have

$$r_n(f(T_n)-f(\theta))\xrightarrow{d}\mathbf{D}_\theta T.$$

5.1 Proof

Under the assumption $r_n(T_n - \theta) \xrightarrow{d} T$ and $r_n \to \infty$, we have

$$T_n - \theta = O_p(1/r_n) = o_p(1).$$

Let $R(h) = f(\theta + h) - f(\theta) - \mathbf{D}_\theta h$. Then $R(h) = o(||h||)$. Now, choosing $h_n = T_n - \theta$ and using a lemma from previous lecture, we have

$$R(h_n) = o_p(||h_n||).$$

Hence

$$\begin{aligned} r_n(f(T_n) - f(\theta)) =& r_n(f(\theta + h_n) - f(\theta)) \\ =& r_n[\mathbf{D}_\theta(T_n - \theta) + o_p(||T_n - \theta||)] \\ =& \mathbf{D}_\theta r_n(T_n - \theta) + o_p(r_n||T_n - \theta||). \end{aligned}$$

As $r_n(T_n - \theta) \xrightarrow{d} T$ and $o_p(r_n||T_n - \theta||) = o_p(O_p(1)) = o_p(1)$. Hence, we have

$$r_n(f(T_n) - f(\theta)) \xrightarrow{d} \mathbf{D}_\theta T.$$

6 Sample Variance

The sample variance of n observations $X_1, X_2 \cdots X_n$ is defined as

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \overline{X^2} - \bar{X}^2,$$

where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $\overline{X^2} = n^{-1}\sum_{i=1}^{n} X_i^2$. Define $S^2 = \phi(\bar{X}, \overline{X^2})$, where $\phi(X, Y) := Y - X^2$. Suppose X_i are i.i.d and $\mathbb{E}X_1^k = \alpha_k$. Also, assume the first four moments $\{\alpha_i\}_{i=1}^4$ are finite. Then, by central limit theorem, we have

$$\sqrt{n}\left(\begin{pmatrix}\bar{X} \\ \overline{X^2}\end{pmatrix} - \begin{pmatrix}\alpha_1 \\ \alpha_2\end{pmatrix}\right) \xrightarrow{d} N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\alpha_2 - \alpha_1^2 & \alpha_3 - \alpha_1\alpha_2 \\ \alpha_3 - \alpha_1\alpha_2 & \alpha_4 - \alpha_2^2\end{pmatrix}\right).$$

The function ϕ is differentiable at (α_1, α_2) and its gradient at that point is given by $\nabla\phi(\alpha_1, \alpha_2) = (-2\alpha_1, 1)$. Let the vector $(T_1, T_2)^\top$ possess the normal distribution described above. Then, applying the delta method, we get

$$\sqrt{n}(\phi(\bar{X}, \overline{X^2}) - \phi(\alpha_1, \alpha_2)) \xrightarrow{d} -2\alpha_1 T_1 + T_2.$$