L19: Slutsky's theorem

1. Slutsky's Theorem

(1) Slutsky's Theorem

Suppose $X_n \xrightarrow{d} X \in R$ and $Y_n \xrightarrow{d} C \in R$. Then $X_n + Y_n \xrightarrow{d} X + c$.

Proof: With continuity point t for $F_{X+c}(\cdot)$, We show that $F_{X_n+Y_n}(t) \longrightarrow F_{X+c}(t)$.

$$\begin{aligned} F_{X_n+Y_n}(t) &= P(X_n + Y_n \le t) \\ &= P(X_n + Y_n \le t,\, |c - Y_n| < \epsilon_1) + P(X_n + Y_n \le t,\, |c - Y_n| \ge \epsilon_1) \\ &\le P(X_n \le t - c + \epsilon_1) + P(|c - Y_n| \ge \epsilon_1) \\ &= F_{X_n}(t - c + \epsilon_1) + P(|c - Y_n| \ge \epsilon_1) \end{aligned}$$

$$\begin{aligned} F_{X_n}(t - c - \epsilon_2) &= P(X_n \le t - c - \epsilon_2) \\ &\le P(X_n \le t - c - \epsilon_2,\, |Y_n - c| < \epsilon_2) + P(|Y_n - c| \ge \epsilon_2) \\ &\le P(X_n + Y_n \le t) + P(|Y_n - c| \ge \epsilon_2) \\ &= F_{X_n+Y_n}(t) + P(|Y_n - c| \ge \epsilon_2). \end{aligned}$$

So $F_{X_n}(t-c-\epsilon_2) - P(|Y_n - c| \ge \epsilon_2) \le F_{X_n+Y_n}(t) \le F_{X_n}(t-c+\epsilon_1) + P(|c - Y_n| \ge \epsilon_1)$.
$Y_n \xrightarrow{d} c \Longrightarrow Y_n \xrightarrow{p} c$. So $P(|c - Y_n| \ge \epsilon_1) \longrightarrow 0$ and $P(|c - Y_n| \ge \epsilon_2) \longrightarrow 0$.
t is continuity point for $F_{X+c}(\cdot) \Longrightarrow t - c$ is a continuity point for $F_X(\cdot)$. Keep $t - c - \epsilon_2$ and $t - c + \epsilon_1$ as continuity points of $F_X(\cdot)$. By $X_n \xrightarrow{d} X$,
$F_{X_n}(t - c + \epsilon_1) \longrightarrow F_X(t - c + \epsilon_1)$ and $F_{X_n}(t - c - \epsilon_2) \longrightarrow F_X(t - c - \epsilon_2)$. So

$$F_X(t - c - \epsilon_2) \le \lim_n F_{X_n+Y_n}(t) \le F_X(t - c + \epsilon_1).$$

Let $\epsilon_1 \longrightarrow 0$ and $\epsilon_2 \longrightarrow 0$. Because $F_X(\cdot)$ is continuous at $t - c$, it follows that $F_{X_n+Y_n}(t) \longrightarrow F_X(t - c) = F_{X+c}(t)$. □

(2) A corollary

$X_n \xrightarrow{d} X \in R^p$ and $Y_n \xrightarrow{d} C \in R^q \Longleftrightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ C \end{pmatrix}$.

Proof: $\Leftarrow$ is trivial. Only show $\Rightarrow$. For all $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in R^{p+q}$, $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \begin{pmatrix} X_n \\ Y_n \end{pmatrix} = \alpha' X_n + \beta' C$.

But $X_n \xrightarrow{d} X \Longrightarrow \alpha' X_n \xrightarrow{d} \alpha' X$ and $Y_n \xrightarrow{d} C \Longrightarrow \beta' Y_n \xrightarrow{d} \beta' C$.
By Slutsky's theorem, $\alpha' X_n + \beta' Y_n \xrightarrow{d} \alpha' X + \beta' C = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \begin{pmatrix} X \\ C \end{pmatrix}$.
So $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \begin{pmatrix} X \\ C \end{pmatrix}$. By Cramer-Wold device $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ C \end{pmatrix}$. □

Comment: Further $g\left(\begin{pmatrix} X_n \\ Y_n \end{pmatrix}\right) \xrightarrow{d} g\left(\begin{pmatrix} X \\ C \end{pmatrix}\right)$ for continuous $g(\cdot)$.

Ex1: $X_n \xrightarrow{d} X \in R^k$ and $Y_n \xrightarrow{d} C \in R^k$. Show that $X_n + Y_n \xrightarrow{d} X + C$.

Proof: $X_n \xrightarrow{d} X \in R^k$ and $Y_n \xrightarrow{d} C \in R^k$. By the Corollary $\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ C \end{pmatrix}$.

With continuous $g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x + y \in R^k$, $g\left(\begin{pmatrix} X_n \\ Y_n \end{pmatrix}\right) \xrightarrow{d} g\left(\begin{pmatrix} X \\ C \end{pmatrix}\right)$, i.e.,
$X_n + Y_n \xrightarrow{d} X + C$.

2. Convergence in distribution and boundedness in probability

(1) $X_n \xrightarrow{d} X \in R \Longrightarrow X_n = O_p(1)$.

Proof: Let $F_n(\cdot)$ and $F(\cdot)$ be the cdfs of X_n and X respectively. $\forall \epsilon > 0$,
$F(-\infty) = 0 \Longrightarrow \exists$ continuity point $-M_1 < 0$ for $F(\cdot)$ such that $F(-M_1) < \frac{\epsilon}{2}$.
$F(\infty) = 1 \Longrightarrow \exists$ continuity point $M_2 > 0$ for $F(\cdot)$ such that $F(M_2) > 1 - \frac{\epsilon}{2}$.
$X_n \xrightarrow{d} X \Longrightarrow F_n(-M_1) \longrightarrow F(-M_1) < \frac{\epsilon}{2}$, and $F_n(M_2) \longrightarrow F(M_2) > 1 - \frac{\epsilon}{2}$.
$\exists\, N_1 > 0$, $\forall\, n > N_1$, $F_n(-M_1) < \frac{\epsilon}{2}$; $\exists\ N_2 > 0$, $\forall\, n > N_2$, $F_n(M_2) > 1 - \frac{\epsilon}{2}$.
Let $M = \max(M_1, M_2)$ and $N = \max(N_1, N_2)$. $\forall\, n > N$,

$$\begin{aligned} P(|X_n| \le M) &= P(-M \le X_n \le M) \\ &\ge P(-M_1 < X_n \le M_2) = F_n(M_2) - F_n(-M_1) \\ &\ge 1 - \tfrac{\epsilon}{2} - \tfrac{\epsilon}{2} = 1 - \epsilon \end{aligned}$$

Thus $X_n = O_p(1)$

(2) $X_n \xrightarrow{d} X \in R^k \Longrightarrow X_n = O_p(1)$.

Proof: $X_n \xrightarrow{d} X \in R^k \implies (X_n)_i \xrightarrow{d} X_i$ for all $i = 1, ..., k$
$\overset{(i)}{\Longrightarrow}$ $(X_n)_i = O_p(1)$ for all $i = 1, ..., k \Longleftrightarrow X_n = O_p(1)$.

Ex2: A chain: $X_n \xrightarrow{as} X \Longrightarrow X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X \Longrightarrow X_n = O_p(1)$.

Ex3: Suppose $Y_n \xrightarrow{p} 0 \in R$ and $\frac{\widehat{\theta}_n - \theta}{Y_n} \xrightarrow{d} Z \in R^k$.

Show that $\widehat{\theta}_n$ is a consistent estimator for θ.

Pf: $Y_n \xrightarrow{p} 0 \in R \Longrightarrow Y_n = o_p(1)$; $\quad \frac{\widehat{\theta}_n - \theta}{Y_n} \xrightarrow{d} Z \in R^k \Longrightarrow \frac{\widehat{\theta}_n - \theta}{Y_n} = O_p(1)$

Thus $\quad \widehat{\theta}_n - \theta = Y_n \frac{\widehat{\theta}_n - \theta}{Y_n} = o_p(1)\, O_p(1) = o_p(1)$.

So $\widehat{\theta}_n \xrightarrow{p} \theta$. Hence $\widehat{\theta}_n$ is a consistent estimator for θ.

Comment: $\begin{pmatrix} \frac{\widehat{\theta}_n - \theta}{Y_n} \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z \\ 0 \end{pmatrix}$.

L20 Asymptotic Normality

1. Normal distributions

   (1) Traditional definition

   $X \sim N(\mu, \Sigma)$
   $\overset{def}{\iff} X$ has pdf $f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]$
   $\implies E(X) = \mu$ and $\mathrm{Cov}(X) = \Sigma > 0$.

   $X \in N(\mu, \Sigma)$
   $\iff BX + b \sim N(B\mu + b, B\Sigma B')$ for all full row-rank B and all b
   $\implies$ dimention of $BX + b$ is less than or equal to that of X.

   (2) Generalized normal distribution

   $X \sim N(\mu, \Sigma)$
   $\overset{def}{\iff} \exists\ Z \in N(0, I_r)$ such that $X \overset{d}{=} AZ + \mu$ and $AA' = \Sigma \geq 0$
   $\implies P(X \in D) = P(Z \in \{z : Az + \mu \in D\})$

   $X \in N(\mu, \Sigma)$ with $\mu \in R^p$ and $\Sigma \geq 0$
   $\iff BX + b \sim N(B\mu + b, B\Sigma B')$ for all B and all b
   $\overset{*}{\implies} X \in \mu + L(\Sigma)$ where $L(\Sigma) = \{\Sigma u : u \in R^p\}$.

   Proof: $X = \mu + Az \in \mu + L(A) = \mu + L(AA') = \mu + L(\Sigma)$.

   Ex1: $Z \sim N(0, 1^2) \implies X \overset{d}{=} \begin{pmatrix}1\\2\end{pmatrix} Z + \begin{pmatrix}1\\1\end{pmatrix} \sim N\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1 & 2\\2 & 4\end{pmatrix}\right)$

   $X \in \begin{pmatrix}1\\1\end{pmatrix} + L\left(\begin{pmatrix}1 & 2\\2 & 4\end{pmatrix}\right) = \begin{pmatrix}1\\1\end{pmatrix} + L\left(\begin{pmatrix}1\\2\end{pmatrix}\right) = \left\{\begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}1\\2\end{pmatrix} t : t \in R\right\}$.

   With $X = \begin{pmatrix}Z+1\\2Z+1\end{pmatrix}$, its support is clearly given by $\left\{\begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}1\\2\end{pmatrix} t : t \in R\right\}$

2. Asymptotic normality

   (1) Definitions

   $$X_n \sim AN(\mu_n, \Sigma_n) \overset{def}{\iff} \exists\ Z_n \xrightarrow{d} Z \sim N(0, I_r) \text{ such that } X_n \overset{d}{=} B_n Z_n + \mu_n \text{ and } B_n B_n' = \Sigma_n.$$

   (2) A special case: $\mu_n = \mu$ and $\Sigma_n = \Sigma$

   (i) $X_n \sim N(\mu, \Sigma) \implies X_n \xrightarrow{d} N(\mu, \Sigma)$.

   (ii) $X_n \xrightarrow{d} N(\mu, \Sigma), \Sigma > 0 \implies X_n \sim AN(\mu, \Sigma) \implies X_n \xrightarrow{d} N(\mu, \Sigma)$.

   Proof: (i) For $X_n \sim N(\mu, \Sigma)$, let $X \overset{d}{=} X_n$.
   Then $P(X_n \in A) = P(X \in A) \longrightarrow P(X \in A)$.
   Thus $X_n \xrightarrow{d} X \sim N(\mu, \Sigma)$.

(ii) First $\Rightarrow$: From $X_n \xrightarrow{d} X \sim N(\mu, \Sigma)$ with $\Sigma > 0$,

$Z_n = \Sigma^{-1/2}(X_n - \mu) \xrightarrow{d} Z = \Sigma^{-1/2}(X - \mu) \sim N(0, I)$.

But $X_n = \Sigma^{1/2} Z_n + \mu$ and $\Sigma^{1/2}(\Sigma^{1/2})' = \Sigma$. So $X_n \sim AN(\mu, \Sigma)$.

Second $\Rightarrow$: From $X_n \sim AN(\mu, \Sigma)$,

$\exists\ Z_n \xrightarrow{d} Z \sim N(0, I)$ such that $X_n = BZ_n + \mu$ and $BB' = \Sigma$.

So $X_n = BZ_n + \mu = g(Z_n) \xrightarrow{d} g(Z) = BZ + \mu \sim N(\mu, \Sigma)$.

3. Transformation and applications

(1) Transformation

$$X_n \sim AN(\mu_n, \Sigma_n) \Longleftrightarrow A_n X_n + \beta_n \sim AN\left(A_n\mu_n + \beta_n,\ A_n\Sigma_n A_n'\right) \text{ for all } A_n \text{ and } \beta_n$$

Proof: $\Leftarrow$: Take $A_n = I$ and $\beta_n = 0$.

$\Rightarrow$: From $X_n \sim AN(\mu_n, \Sigma_n)$,

$\exists\ Z_n \xrightarrow{d} Z \sim N(0, I)$ such that $X_n = B_n Z_n + \mu_n$ and $B_n B_n' = \Sigma_n$.

Let $Y_n = AX_n + \beta_n$. We show that $Y_n \sim AN(A_n\mu_n + \beta_n,\ A_n\Sigma_n A_n')$.

With $Z_n \xrightarrow{d} Z \sim N(0, I)$,

$Y_n = AX_n + \beta_n = A_n B_n Z_n + A_n\mu_n + \beta_n$ and $(A_nB_n)(A_nB_n)' = A_n\Sigma_n A_n'$.

By definition, $Y_n \sim AN(A_n\mu_n + \beta_n,\ A_n\Sigma_n A_n')$.

Ex2: $X_n \sim AN(\mu, a_n^2\Sigma) \Longrightarrow \frac{X_n - \mu}{a_n} \sim AN(0, \Sigma) \Longrightarrow \frac{X_n - \mu}{a_n} \xrightarrow{d} X \sim N(0, \Sigma)$

$\Longrightarrow g\left(\frac{X_n - \mu}{a_n}\right) \xrightarrow{d} g(X)$.

(2) Approximation

If $X_n \xrightarrow{d} X$, then the distribution of X_n with large n can be approximated by that of X. Hence $P(X_n \in A) \approx P(X \in A)$, $E(X_n) \approx E(X)$, and $\mathrm{Cov}(X_n) \approx \mathrm{Cov}(X)$.

If $Z_n \xrightarrow{d} Z$, then the distribution Z_n is approximated by that of Z, and further the distribution of $X_n = B_n Z_n + \mu_n$ is approximated by that of $X(n) = B_n Z + \mu_n$. Hence $P(X_n \in A) \approx P(X(n) \in A)$, $E(X_n) \approx E(X(n))$ and $\mathrm{Cov}(X_n) \approx \mathrm{Cov}(X(n))$.

(3) Application of $AN(\mu_n, \Sigma_n)$

If $X_n \sim AN(\mu_n, \Sigma_n)$, then

$\exists\, Z_n \xrightarrow{d} Z \sim N(0, I)$ such that $X_n \overset{d}{=} B_n Z_n + \mu_n$ and $B_n B_n' \Sigma_n$.

So the distribution of Z_n is approximated by that of $Z \in N(0, I)$;

The distribution of $X_n = B_n Z_n + \mu_n$ is approximated by that of $B_n Z + \mu_n$.

But $B_n Z + \mu_n \sim N(\mu_n, \Sigma_n)$. Thus

$P(X_n \in A) \approx P(N(\mu_n, \Sigma_n) \in A)$, $E(X_n) \approx \mu_n$ and $\mathrm{Cov}(X_n) \approx \Sigma_n$.