5(c)

35) The delta method is a special case since if T_n and θ are real valued, then $D_{g(\theta)} = g'(\theta)$.

See HW6

ex) Suppose $\sqrt{n}\left[\begin{pmatrix}\hat{\lambda}\\ \hat{\eta}\end{pmatrix}-\begin{pmatrix}\lambda\\ \eta\end{pmatrix}\right] \xrightarrow{D} N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \underbrace{\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}}_{\Sigma}\right]$

a) Find the limiting dist of

$$\sqrt{n}\left(\hat{\lambda}\hat{\eta} - \lambda\eta\right) = N(0, \tau^2).$$

soln $\underline{\theta} = \begin{pmatrix}\lambda\\ \eta\end{pmatrix}$ and $g(\underline{\theta}) = \lambda\eta$

$$D_{g(\underline{\theta})} = \left[\frac{\partial g}{\partial \lambda} \quad \frac{\partial g}{\partial \eta}\right] = [\eta \quad \lambda]$$

$$\tau^2 = D_{g(\underline{\theta})}\,\Sigma\, D_{g(\theta)}^T = [\eta \quad \lambda]\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}\begin{pmatrix}\eta\\ \lambda\end{pmatrix}$$

$$= (\eta\sigma_{11} + \lambda\sigma_{21} \quad \eta\sigma_{12} + \lambda\sigma_{22})\begin{pmatrix}\eta\\ \lambda\end{pmatrix}$$

$$= \eta^2 \sigma_{11} + \eta\lambda\sigma_{21} + \lambda\eta\sigma_{12} + \lambda^2\sigma_{22}$$

$$\boxed{\tau^2 = \eta^2 \sigma_{11}^2 + 2\eta\lambda\sigma_{12} + \lambda^2\sigma_{22}}$$

$$\left(\sigma_{\hat\lambda\hat\eta} = \sigma_{12} = \sigma_{21},\ \sigma_{11} = \sigma^2_{\hat\lambda},\ \sigma_{22} = \sigma^2_{\hat\eta}\right)$$

b) $$\underbrace{(1 \quad 1)}_{A} \sqrt{n}\left[\begin{pmatrix}\hat\lambda\\ \hat\eta\end{pmatrix} - \begin{pmatrix}\lambda\\ \eta\end{pmatrix}\right] =$$

$$\sqrt{n}\left[(\hat\lambda + \hat\eta) - (\lambda+\eta)\right] \xrightarrow{D} N(0, A\Sigma A^T)$$

$$\text{where } A\Sigma A^T = [1 \quad 1]\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} =$$

$$[\sigma_{11} + \sigma_{21} \quad \sigma_{12} + \sigma_{22}]\begin{pmatrix}1\\1\end{pmatrix} = \sigma_{11} + 2\sigma_{12} + \sigma_{22}$$

b) $$\overbrace{(1 \quad -1)}^{B} \sqrt{n}\left[\begin{pmatrix}\hat\lambda\\ \hat\eta\end{pmatrix} - \begin{pmatrix}\lambda\\ \eta\end{pmatrix}\right] =$$

$$\sqrt{n}\left[(\hat\lambda - \hat\eta) - (\lambda-\eta)\right] \xrightarrow{D} N(0, B\Sigma B^T)$$

52.5

where $B \Sigma B^T = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$= \begin{pmatrix} \sigma_{11} - \sigma_{21} & \sigma_{12} - \sigma_{22} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} =$

$\sigma_{11} - \sigma_{21} - \sigma_{12} + \sigma_{22} = \sigma_{11} - 2\sigma_{12} + \sigma_{22}$

CLTs, O_p, and o_p

1] For each n, let $W_{n1}, \dots, W_{nr_n}$ be independent. The prob dist may change with n. These RVs form a double array. Let $E[W_{nk}] = 0$, $V(W_{nk}) = E(W_{nk}^2) = \sigma_{nk}^2$ and $s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2 = V\left(\sum_{k=1}^{r_n} W_{nk}\right)$. $Z_n = \dfrac{\sum_{k=1}^{r_n} W_{nk}}{s_n}$ is the Z score of $\sum_{k=1}^{r_n} W_{nk}$.

ex

W_{11}

$W_{21} \quad W_{22}$

$W_{31} \quad W_{32} \quad W_{33}$

$W_{41} \quad W_{42} \quad W_{43} \quad W_{44}$

$\vdots$

double array = triangular array if $r_n = n$

2) <u>Lyapounov's CLT</u>: Under 1),

assume the $E[|W_{nk}|^{2+\delta}]$ exist for

some $\delta > 0$. Assume Lyapounov's

condition: $\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{E[|W_{nk}|^{2+\delta}]}{s_n^{2+\delta}} = 0$.

Then $Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1)$,

3) Special cases: i) $r_n = n$, $W_{nk} = W_k$

has $W_1, \ldots, W_n, \ldots$ independent,

ii) $W_{nk} = X_{nk} - E(X_{nk}) = X_{nk} - \mu_{nk}$

has $\frac{\sum_{k=1}^{r_n} (X_{nk} - \mu_{nk})}{s_n} \xrightarrow{D} N(0,1)$.

iii) Suppose $X_1, X_2, \ldots$ are ind. with $E(X_i) = \mu_i$, $V(X_i) = \sigma_i^2$. Let

$$Z_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{\left(\sum_{i=1}^{n} \sigma_i^2 \right)^{\frac{1}{2}}}$$ be the z score

of $\sum_{i=1}^{n} X_i$. Hence $E(Z_n) = 0$, $V(Z_n) = 1$.

Assume $E\left[|X_i - \mu_i|^3 \right] < \infty$ for $i \in \mathbb{N}$

and $$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} E\left[|X_i - \mu_i|^3 \right]}{\left(\sum_{i=1}^{n} \sigma_i^2 \right)^{3/2}} = 0. \quad (*)$$

Then $Z_n \xrightarrow{D} N(0,1)$.

proof of iii): Take $W_{nk} = X_k - \mu_k$, $\delta = 1$,

$s_n = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$ and apply Lyapounov's

CLT. Note that

$$\left(\sum_{i=1}^{n} \sigma_i^2\right)^{3/2} = (s_n^2)^{3/2} = s_n^3 = s_n^{2+1}$$

4) The (Lindeberg Lévy) CLT has the X_i iid with $V(X_i) = \sigma^2 < \infty$.

The Lyapounov CLT in 3 iii) has the X_i independent (not necessarily identically distributed) with $E[|X_i|^3] < \infty$ and satisfies (*),

ex) MS81 Qual problem: Suppose the $X_i \overset{ind}{\sim} Ber(p_i)$ with $\sum_{i=1}^{\infty} p_i q_i = \infty$ where $q_i = 1 - p_i$ and $V(X_i) = p_i q_i$.

Then $Z_n = \dfrac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} p_i}{\left(\sum_{i=1}^{n} p_i q_i\right)^{\frac{1}{2}}} \xrightarrow{D} N(0,1)$

Proof: Let $Y_i = |W_i| = |X_i - p_i|$.

Then $E[|X_i - p_i|^3] =$

$w_i = x_i - p_i$	$1-p_i$	$-p_i$
x	1	0
$f(x)$	p_i	q_i
y	$1-p_i$	q_i

$$E[|W_i|^3] = \sum_y y^3 f(y) = (1-p_i)^3 p_i + p_i^3 q_i$$

$$= q_i^3 p_i + p_i^3 q_i = p_i q_i \underbrace{(p_i^2 + q_i^2)}_{\le (p_i+q_i)^2 = 1} \le p_i q_i .$$

$$\therefore \sum_{i=1}^{n} E[|X_i - p_i|^3] \le \sum_{i=1}^{n} p_i q_i .$$

divide both sides by $\left(\sum_i p_i q_i\right)^{3/2}$

or $$\frac{\sum_{i=1}^{n} E[|X_i - p_i|^3]}{\left(\sum_{i=1}^{n} p_i q_i\right)^{3/2}} \le \frac{1}{\left(\sum_{i=1}^{n} p_i q_i\right)^{\frac{1}{2}}} \to 0.$$

$\therefore$ (*) in 3) iii) holds and $Z_n \xrightarrow{D} N(0,1)$.

$\square$

DeGroot

5) <u>Lindeberg CLT</u>: Let the W_{nk} satisfy 1) and <u>Lindeberg's condition</u>

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} E\left(W_{nk}^2 \, I\left[|W_{nk}| \geq \varepsilon s_n\right]\right)$$

$= 0$ for any $\varepsilon > 0$. Then $Z_n \xrightarrow{D} N(0,1)$.

Note: Sometimes called the Lindeberg–Feller CLT.

6) special case: Assume $r_n = n$ and the $W_{nk} = W_k$ are ind. If

$$(*)\quad \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{s_n^2} E\left(W_k^2 \, I\left[|W_k| \geq \varepsilon s_n\right]\right) = 0$$

$\forall \varepsilon > 0$, then $Z_n = \dfrac{\sum_{k=1}^{n} W_k}{s_n} \xrightarrow{D} N(0,1)$.

7) Lindeberg's Condition is nearly necessary for $Z_n = \dfrac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1)$.

55.5

8) a) <u>Uniformly bounded sequence:</u>

Let $r_n = n$, $W_{nk} = W_k$. If there is a constant $c > 0$ such that $P(|W_k| < c) = 1, \forall k$, and if $\Delta_n \to \infty$, then 6} holds.

proof: Once n is large enough so that

$$\underbrace{\varepsilon \Delta_n > c}_{\text{occurs since } \Delta_n \to \infty}, \quad I\{|W_k| \geq \varepsilon \Delta_n\} = 0$$

So $(*) = 0$.

b) If $W_1, W_2, \ldots$ are iid with $V(W_i) = \sigma^2 > 0$, then 6} holds.

9} Taking $W_i = X_i - \mu$ means

8b) $\Rightarrow$ (Lindeberg Lévy) CLT holds.

10} If the Lyapounov condition holds, then Lindeberg's condition holds.

So Lindeberg's CLT proves the Lyapounov's CLT.

ex] For each $n \in \mathbb{N}$, let W_{nk} be ind with $E[W_{nk}] = 0$, $V(W_{nk}) = \sigma_{nk}^2$, and $\Delta_n^2 = \sum_{k=1}^{n} \sigma_{nk}^2$. Suppose $|W_{nk}| \leq M_n$ and $\frac{M_n}{\Delta_n} \to 0$. Verify Lindeberg's condition holds (with using the fact that if Lyapounov's condition holds then Lindeberg's condition holds).

proof: $\frac{1}{\Delta_n^2} E\left[W_{nk}^2 \, I(|W_{nk}| \geq \varepsilon \Delta_n)\right]$

$$\leq \frac{1}{\Delta_n^2} E\left[M_n^2 \, I(|W_{nk}| \geq \varepsilon \Delta_n)\right]$$

$$= \frac{M_n^2}{\Delta_n^2} P\left[|W_{nk}| \geq \varepsilon \Delta_n\right]$$

$$\underset{\text{cheb}}{\leq} \frac{M_n^2}{\Delta_n^2} \frac{E(W_{nk}^2)}{\varepsilon^2 \Delta_n^2} = \left(\frac{M_n}{\Delta_n}\right)^2 \frac{\sigma_{nk}^2}{\Delta_n^2 \varepsilon^2}. \text{ So}$$

$$\frac{\sum_{k=1}^{r_n} E\left[W_{nk}^2\, I\left(|W_{nk}| \geq \varepsilon\, s_n\right)\right]}{s_n^2} \leq$$

$$\left(\frac{M_n}{s_n}\right)^2 \frac{1}{s_n^2\, \varepsilon^2} \underbrace{\sum_{k=1}^{r_n} \sigma_{nk}^2}_{s_n^2} = \left(\frac{M_n}{s_n}\right)^2 \frac{1}{\varepsilon^2} \to 0.$$

11) <u>Hájek Šidak CLT</u>: Let

$X_1, \ldots, X_n$ be iid, $E(X_i) = \mu$, $V(X_i) = \sigma^2$.

Let $\underline{c}_n = (c_{n1}, \ldots, c_{nn})^T$ be a vector of constants such that $\max_{1 \leq i \leq n} \frac{c_{ni}^2}{\sum_{j=1}^{n} c_{nj}^2}$ (*)

$\to 0$ as $n \to \infty$. Then

$$Z_n = \frac{\sum_{i=1}^{n} c_{ni}\,(X_i - \mu)}{\sigma \sqrt{\sum_{j=1}^{n} c_{nj}^2}} \xrightarrow{D} N(0,1).$$

Z score of the sum

12] a) $c_{ni} = \frac{1}{n} \Rightarrow$ usual CLT,

b) $W_{ni} = c_{ni}(X_i - \mu)$ are ind for each n,

$E[W_{ni}] = 0$, $V(W_{ni}) = c_{ni}^2 \sigma^2$,

$V\left(\sum_{i=1}^n W_{ni}\right) = \sum_{i=1}^n V(W_{ni}) = \sigma^2 \sum_{i=1}^n c_{ni}^2 =$

$\sigma^2 \underline{c}_n^T \underline{c}_n$.

DasGupta p 64-5   Sen and Singer 119

c) (*) insures that no coefficient dominates the vector $\underline{c}_{ni}$. For ex,

if $\underline{c}_n = (1, \underline{0}^T)^T = (1, 0, \ldots, 0)^T$ then

(*) fails and the Hájek Šidák CLT fails.

13] p53-54 a) A sequence of RVs W_n is <u>tight</u> or <u>bounded in probability</u>

written $O_p(1)$, if for
"big O p"

every $\varepsilon > 0$, there exist positive constants D_ε and N_ε such that

$$P(|W_n| \le D_\varepsilon) \ge 1-\varepsilon \quad \forall n \ge N_\varepsilon.$$

Also $W_n = O_p(X_n)$ if $\left|\frac{W_n}{X_n}\right| = O_p(1)$.

b) $W_n = \underbrace{o_p}(n^{-\delta})$ if $n^\delta W_n = o_p(1)$

"little o p"

which means $n^\delta W_n \xrightarrow{P} 0$.

c) Wn has the same order as Xn in probability, written $W_n \asymp_p X_n$, if for every $\varepsilon > 0$, $\exists$ positive constants N_ε and $0 < d_\varepsilon < D_\varepsilon$ such that $P\left(d_\varepsilon \le \left|\frac{W_n}{X_n}\right| \le D_\varepsilon\right) =$

$$P\left(\frac{1}{D_\varepsilon} \le \left|\frac{x_n}{w_n}\right| \le \frac{1}{d_\varepsilon}\right) \ge 1-\varepsilon$$

$\forall\, n \ge N_\varepsilon$.

d) Similar notation is used for a $K \times r$ matrix $A_n = A = (a_{ij}(n))$ if each element $a_{ij}(n)$ has the desired property. For ex, $A = O_p(n^{-1/2})$ if each $a_{ij}(n) = O_p(n^{-\frac{1}{2}})$.

Note: $w_n = O_p(n^{-\frac{1}{2}})$ means

$|\sqrt{n}\, w_n| = O_p(1)$.

14) Let $w_n = \|\hat{\underline{\mu}}_n - \underline{\mu}\|$.

a) If $w_n \asymp_p n^{-\delta}$ for some $\delta > 0$, both w_n and $\hat{\underline{\mu}}_n$ have <u>rate n^δ</u>.

(want $0 < \delta \le 1$.)

b) If there exists a constant θ such that $n^\delta (W_n - \theta) \xrightarrow{D} X$ for some nondegenerate RV X, both W_n and $\hat{\mu}_n$ have convergence rate n^δ.

15) Th: Suppose $\exists$ constant $\theta \ni$

$$n^\delta (W_n - \theta) \xrightarrow{D} X.$$

a) Then $W_n = O_p(n^{-\delta})$.

b) If X is not degenerate, then

$$W_n \asymp_p n^{-\delta}$$

ex) If $W_n = O_p(n^{-\delta})$, then n^δ is a lower bound on the rate of W_n.

eg if the CLT holds, then

$$\bar{Y}_n = O_p(n^{-1/3}) \text{ but } \bar{Y}_n \asymp_p n^{-\frac{1}{2}}.$$

ex) W_n is a $\sqrt{n}$ consistent est. of θ if $(W_n - \theta) \asymp_p n^{-\frac{1}{2}}$

15) Th a) If $W_n \asymp_p X_n$ then $X_n \asymp_p W_n$.

b) If $W_n \asymp_p X_n$, then $W_n = O_p(X_n)$.

c) If $W_n \asymp_p X_n$, then $X_n = O_p(W_n)$.

d) $W_n \asymp_p X_n$ iff $W_n = O_p(X_n)$ and $X_n = O_p(W_n)$.

16) (Pratt) Th: Let $X_{1n}, \ldots, X_{kn}$ each be $O_p(1)$ where k is fixed (does not depend on n). Let $W_n = X_{i_n, n}$ for some $i_n \in \{1, \ldots, k\}$. Then $W_n = O_p(1)$.

17) Th Suppose $\|T_{jn} - \underline{\beta}\| = O_p(n^{-\delta})$ for $j = 1, \ldots, k$ where $0 < \delta \le 1$.

Let $T_n^* = T_{i_n, n}$ for some $i_n \in \{1, \ldots, k\}$.

Then $\|T_n^* - \underline{\beta}\| = O_p(n^{-\delta})$.

Note: So picking among the $\sqrt{n}$ consistent estimators gives a $\sqrt{n}$ consistent estimator, no matter how the estimator is picked.

183 Th Let W_n, X_n, Y_n and Z_n be sequences of RVs $\ni$ $Y_n > 0$ and $Z_n > 0$ (often Y_n and Z_n are deterministic, eg $Y_n = n^{-1/2}$.)

a) If $W_n = O_p(1)$ and $X_n = O_p(1)$, then

$W_n + X_n = O_p(1)$ and $W_n X_n = O_p(1)$.

Hence $O_p(1) + O_p(1) = O_p(1)$ and

$O_p(1)\, O_p(1) = O_p(1)$,

b) If $W_n = O_p(1)$ and $X_n = o_p(1)$,

then $W_n + X_n = O_p(1)$ and $W_n X_n = o_p(1)$.

So $O_p(1) + o_p(1) = O_p(1)$ while

$O_p(1)\, O_p(1) = O_p(1)$.

c) If $W_n = O_p(Y_n)$ and $X_n = O_p(Z_n)$, then $W_n + X_n = O_p(\max(Y_n, Z_n))$

and $W_n X_n = O_p(Y_n Z_n)$. So

$O_p(Y_n) + O_p(Z_n) = O_p(\max(Y_n, Z_n))$

and $O_p(Y_n)\, O_p(Z_n) = O_p(Y_n Z_n)$.

Prediction Intervals and Regions Och3.

1} We say $P(A_n)$ is "eventually bounded below" by $1-\delta$ if $P(A_n) > 1-\delta-\varepsilon$ for any $\varepsilon > 0$ if n is large enough. Here $0 < \delta < 1$. If $P(A_n) \to 1-\delta$ as $n \to \infty$ then $P(A_n)$ is eventually bounded below by δ.

2} Consider predicting a future <u>test</u> value Y_f given <u>training</u> data $Y_1, \ldots, Y_n$.

A <u>large sample</u> $100(1-\delta)\%$ <u>prediction interval</u> (PI) for Y_f has the form $[\hat{L}_n, \hat{U}_n]$ where $P[\hat{L}_n \le Y_f \le \hat{U}_n]$ is eventually bounded below by $1-\delta$ as the sample size $n \to \infty$. A large sample PI is <u>asymptotically optimal</u> if it has shortest asymptotic length: the length of $[\hat{L}_n, \hat{U}_n]$ converges to $U_s - L_s$ as $n \to \infty$ where $[L_s, U_s]$ is the <u>pop shorth</u>: the shortest interval covering at least $100(1-\delta)\%$ of the mass.

3] Let the <u>actual</u> <u>coverage</u> $1-\delta_n = P[\hat{L}_n \le Y_f \le \hat{U}_n]$ = prop times the PI contains (covers) Y_f. Then