Statistics 581, Problem Set 4 Solutions

Wellner; 10/26/2005

1. Suppose that $\underline{N}_n \sim \text{Mult}_k(n, \underline{p})$ and $\widehat{p} = \underline{N}_n/n$. Suppose that the true $\underline{p}$ is $\underline{p}_n = \underline{p}_0 + n^{-1/2}\underline{c}$ where $\underline{1}^T\underline{c} = 0$. Use the Cramér - Wold device together with either the Liapunov or the Lindeberg-Feller CLT to show that
$$\underline{Z}_n = \left(\frac{N_{n,1} - np_{0,1}}{\sqrt{np_{0,1}}}, \ldots, \frac{N_{nk} - np_{0,k}}{\sqrt{np_{0,k}}} \right)$$
satisfies $\underline{Z}_n \rightarrow_d \text{diag}(1/\sqrt{p_0})\underline{c} + \underline{Z}$ where $\underline{Z} \sim N_k(0, I - \sqrt{p_0}\sqrt{p_0}^T)$. (It therefore follows, as outlined in class, that the chi-square statistic $Q_n \rightarrow_d \chi^2_{k-1}(\delta)$ with $\delta = \sum_{j=1}^k c_j^2/p_{0,j}$ under the local alternative $\underline{p}_n$.)

Solution: We argued heuristically in class that when the true $\underline{p} = \underline{p}_n = \underline{p}_0 + \underline{c}n^{-1/2}$, then
$$\underline{Z}_n \equiv \text{diag}(1/\sqrt{\underline{p}_0})n^{1/2}(\underline{\widehat{p}} - \underline{p}_0) \rightarrow \underline{Z} + \underline{d} \sim N_k(\underline{d}, \Sigma) \tag{1}$$
where $\underline{d} = \text{diag}(1/\sqrt{\underline{p}_0})\underline{c}$ and $\Sigma = I - \sqrt{\underline{p}_0}\sqrt{\underline{p}_0}'$. To prove that (1) holds, we can use the Cramér-Wold device and the Liapunov CLT. Fix $\underline{a} \in R^k$. Then we want to show that
$$\underline{a}^T\sqrt{n}(\underline{\widehat{p}}_n - \underline{p}_n) \rightarrow_d N(0, \underline{a}^T(\text{diag}(\underline{p}_0) - \underline{p}_0\underline{p}_0^T)\underline{a}) \,.$$
But since $\underline{N}_n = \sum_{i=1}^n \underline{V}_{ni}$ where $\underline{V}_{ni} \sim \text{Mult}_k(1, \underline{p}_n)$ are i.i.d. for each n, we can write
$$\begin{aligned} \underline{a}^T\sqrt{n}(\underline{\widehat{p}}_n - \underline{p}_n) &= \sum_{i=1}^n \sum_{j=1}^k a_j(V_{ni,j} - p_{nj})/\sqrt{n} \\ &\equiv \sum_{i=1}^n X_{ni} \end{aligned}$$
where the X_{ni}'s have $\mu_{ni} = E(X_{ni}) = 0$,
$$\sigma^2_{ni} = Var(X_{ni}) = \underline{a}^T(\text{diag}(\underline{p}_n) - \underline{p}_n\underline{p}_n^T)\underline{a}/n$$
and
$$\gamma_{ni} \;=\; E|X_{ni}|^3 = n^{-3/2}\sum_{j'=1}^k \left\{ \left| a_{j'}(1 - p_{nj'}) + \sum_{j \neq j', j=1}^k a_j(0 - p_{nj}) \right|^3 \right\} p_{nj'} \,,$$
so that
$$\sigma^2_n = \sum_1^n \sigma^2_{ni} = \underline{a}^T(\text{diag}(\underline{p}_n) - \underline{p}_n\underline{p}_n^T)\underline{a} \rightarrow \underline{a}^T\Sigma\underline{a}$$

while

$$
\begin{aligned}
\gamma_n &= \sum_1^n \gamma_{ni} \\
&= n^{-1/2} \sum_{j'=1}^k \left\{ \Big| \sum_{j=1}^k a_j(1-p_{nj}) + \sum_{j=1, j\neq j'}^k a_j(0-p_{nj}) \Big|^3 \right\} p_{nj'} \\
&\to 0 \cdot M(\underline{a}, \underline{p}_0) = 0
\end{aligned}
$$

where

$$
M(\underline{a}, \underline{p}_0) = \sum_{j'=1}^k \left\{ \Big| \sum_{j=1}^k a_j(1-p_{0j}) + \sum_{j=1}^k a_j(0-p_{0j}) \Big|^3 \right\} p_{0j'} .
$$

hence it follows that $\gamma_n/\sigma_n^{3/2} \to 0$, and

$$
\frac{\underline{a}^T \sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n)}{\sigma_n} = \frac{\sum_{i=1}^n X_{ni}}{\sigma_n} \to_d N(0,1) .
$$

This implies

$$
\underline{a}^T \sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n) \to_d N(0, \underline{a}^T \Sigma \underline{a}) ,
$$

and by Cramér - Wold, this implies

$$
\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_n) \to_d N_k(0, \Sigma) .
$$

2. Ferguson, ACILST, problem 5, page 50. (The Poisson dispersion test). A standard test of the hypothesis H_0 that a distribution is Poisson(λ) for some λ is to reject H_0 if the ratio of the sample variance to the sample mean, $S_n^2/\overline{X}_n$, is too large. This test is good against alternatives whose variance is greater than the mean, such as the negative binomial distribution or any other mixture of Poisson distributions.
(a) Find the asymptotic distribution of $S_n^2/\overline{X}_n$ for general distributions.
(b) Find the asymptotic distribution of $S_n^2/\overline{X}_n$ under H_0 and show that it is independent of λ.
Solution: (a) We can use the result of part (a) of the previous problem. We just need to proceed as in (b) of the previous problem with $g(u,v) = v/u$. Thus we find that $\nabla g(u,v) = (-v/u^2, 1/u) = (-v/u, 1)/u$. Hence $\nabla g(\mu, \sigma^2) = (-\sigma^2/\mu, 1)/\mu$, and the limiting variance is

$$
\begin{aligned}
\nabla g^T \Sigma \nabla g &= \frac{\sigma^4}{\mu^2} \left(\frac{\sigma^2}{\mu^2} - 2\frac{\mu_3}{\mu\sigma^2} + \frac{\mu_4}{\sigma^4} - 1 \right) \\
&= \frac{\sigma^4}{\mu^2} \left(\frac{\sigma^2}{\mu^2} - 2\frac{\sigma\gamma_1}{\mu} + 2 + \gamma_2 \right) .
\end{aligned}
$$

(b) When $X \sim$ Poisson(λ), $E(X) = \lambda$, $Var(X) = \lambda$, $\gamma_1 = 1/\sqrt{\lambda}$, and $\gamma_2 = 1/\lambda$. Thus we find that the asymptotic variance above is

$$
\frac{\lambda^2}{\lambda^2} \left\{ \frac{\lambda}{\lambda^2} - 2\frac{\lambda^{1/2}\lambda^{-1/2}}{\lambda} + 2 + \frac{1}{\lambda} \right\} = 2 .
$$

Thus it follows that under $X \sim \text{Poisson}(\lambda)$ we have

$$\sqrt{n}\left(\frac{S_n^2}{\overline{X}_n} - \frac{\sigma^2}{\mu}\right) = \sqrt{n}\left(\frac{S_n^2}{\overline{X}_n} - 1\right) \to_d N(0,2)\,.$$

3. The usual F test for the equality of variances of two independent normal populations is based on the ratio of the two sample variances S_X^2/S_Y^2. Show that this test is not asymptotically distribution-free within the class of distributions with finite fourth moments, by finding the asymptotic distribution of $\sqrt{n}(S_X^2/S_Y^2 - \sigma_X^2/\sigma_Y^2)$ within this class. Suppose both samples are of size n.

Solution: As we showed in class

$$\sqrt{n}(S_X^2 - \sigma_X^2) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mu_X)^2 - \sigma_X^2\right) + o_p(1),$$
$$\sqrt{n}(S_Y^2 - \sigma_Y^2) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n (Y_i - \mu_Y)^2 - \sigma_Y^2\right) + o_p(1).$$

Since the X's and Y's are independent, this implies (e.g. via Cramér- Wold) that

$$\begin{aligned}\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (X_i-\mu_X)^2 - \sigma_X^2 \\ \frac{1}{n}\sum_{i=1}^n (Y_i-\mu_Y)^2 - \sigma_Y^2 \end{pmatrix} &\to_d \underline{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \\ &\sim N_2(0, \begin{pmatrix} 2\sigma_X^4(1+\gamma_{2X}/2) & 0 \\ 0 & 2\sigma_Y^4(1+\gamma_{2Y}/2) \end{pmatrix})\end{aligned}$$

where $\gamma_{2X} = \mu_{4X}/\sigma_X^4 - 3$, $\gamma_{2Y} = \mu_{4Y}/\sigma_Y^4 - 3$. Since $S_X^2/S_Y^2 = g(S_X^2, S_Y^2)$ and $\sigma_X^2/\sigma_Y^2 = g(\sigma_X^2, \sigma_Y^2)$ with $g(u,v) = u/v$, the delta-method yields

$$\begin{aligned}\sqrt{n}\left(\frac{S_X^2}{S_Y^2} - \frac{\sigma_X^2}{\sigma_Y^2}\right) &\to_d g'(\sigma_X^2, \sigma_Y^2)\underline{Z} \\ &= \frac{1}{\sigma_Y^2}(1, -\sigma_X^2/\sigma_Y^2)\underline{Z} \\ &\sim N(0, 2\frac{\sigma_X^4}{\sigma_Y^4}(1+\frac{\gamma_{2X}}{2}) + 2\frac{\sigma_X^4}{\sigma_Y^4}(1+\frac{\gamma_{2Y}}{2})).\end{aligned}$$

Under the hypothesis $\sigma_X^2 = \sigma_Y^2$ this yields

$$\begin{aligned}\sqrt{n}\left(\frac{S_X^2}{S_Y^2} - 1\right) &\to_d N(0, 2(1+\frac{\gamma_{2X}}{2}) + 2(1+\frac{\gamma_{2Y}}{2})) \\ &= N(0,4) \qquad \text{under normality}\end{aligned}$$

since $\gamma_{2X} = \gamma_{2Y} = 0$ under normality. Since under normality $(n-1)S_X^2/\sigma_X^2 \sim \chi^2_{n-1}$ and $(n-1)S_Y^2/\sigma_Y^2 \sim \chi^2_{n-1}$ are independent, the usual normal theory F test of $H : \sigma_X^2 = \sigma_Y^2$ versus $K : \sigma_X^2 > \sigma_Y^2$ is "reject H if $S_X^2/S_Y^2 > F_{n-1,n-1,\alpha}$" where $P(F_{n-1,n-1} > F_{n-1,n-1,\alpha}) = \alpha$. Thus

$$\begin{aligned}\alpha &= P_{\sigma_X=\sigma_Y, normality}(S_X^2/S_Y^2 > F_{n-1,n-1,\alpha}) \\ &= P_{\sigma_X=\sigma_Y, normality}(\sqrt{n}(S_X^2/S_Y^2 - 1) > \sqrt{n}(F_{n-1,n-1,\alpha}) - 1)) \\ &\to P(2Z > 2z_\alpha) = \alpha.\end{aligned}$$

so we conclude that $\sqrt{n}(F_{n-1,n-1,\alpha} - 1) \to 2z_\alpha$. Thus when $\sigma_X^2 = \sigma_Y^2$ but the X's and Y's are not normal (but have $\mu_{4X} < \infty$ and $\mu_{4Y} < \infty$),

$$\begin{aligned} P_{\sigma_X=\sigma_Y}(S_X^2/S_Y^2 > F_{n-1,n-1,\alpha}) &= P_{\sigma_X=\sigma_Y}(\sqrt{n}(S_X^2/S_Y^2 - 1) > \sqrt{n}(F_{n-1,n-1,\alpha} - 1)) \\ &\to P(N(0, 4 + \gamma_{2X} + \gamma_{2Y}) > 2z_\alpha) \\ &= P(N(0, 1 + (\gamma_{2X} + \gamma_{2Y})/4) > z_\alpha) \neq \alpha \end{aligned}$$

if $\gamma_{2X} + \gamma_{2Y} \neq 0$.

4. Suppose that $X_1, X_2, \ldots$ are i.i.d. positive random variables, and define $\overline{X}_n \equiv n^{-1}\sum_{i=1}^n X_i$, $H_n \equiv 1/(n^{-1}\sum_{i=1}^n (1/X_i))$, and $G_n \equiv \{\prod_{i=1}^n X_i\}^{1/n}$ to be the *arithmetic, harmonic, and geometric* means respectively. We know that $\overline{X}_n \to_{a.s.} E(X_1) = \mu$ if and only if $E|X_1| < \infty$.
(a) Use the SLLN together with appropriate additional hypotheses to show that $H_n \to_{a.s.} 1/\{E(1/X_1)\} \equiv h$, and $G_n \to_{a.s.} \exp(E\{\log X_1\}) \equiv g$.
(c) Use the multivariate CLT and the delta method to find the joint limiting distribution of $\sqrt{n}(\overline{X}_n - \mu, H_n - h, G_n - g)$. You will need to impose or assume additional moment conditions to be able to prove this. Specify these additional assumptions carefully.
(d) Suppose that $X_i \sim \text{Gamma}(r, \lambda)$ with $r > 0$. For what values of r are the hypotheses you imposed in (c) satisfied? Compute the covariance of the limiting distribution in (c) as explicitly as you can in this case.
(e) Use the result in (c) to show that $\sqrt{n}(G_n/\overline{X}_n - g/\mu) \to_d N(0, V^2)$ and compute V^2 explicitly when $X_i \sim \text{Gamma}(r, \lambda)$ with r satisfying the conditions you found in (d).

Solution: (a) If $0 < E(1/X_1) < \infty$, then

$$\frac{1}{n}\sum_{i=1}^n (1/X_i) \to_{a.s.} E(1/X_1) > 0.$$

If $E|\log(X_1)| < \infty$, then

$$\log G_n = \frac{1}{n}\sum_{i=1}^n \log(X_i) \to_{a.s.} E\log X_1.$$

Thus by the continuous mapping theorem if both $E(1/X_1) < \infty$ and $E|\log X_1| < \infty$, it follows that

$$(H_n, G_n) \to_{a.s.} (1/E(1/X_1), \exp(E\log X_1)) \equiv (h, g).$$

(c) By the multivariate CLT, if $EX_1^2 < \infty$, $E(1/X_1)^2 < \infty$, and $E(\log X_1)^2 < \infty$, then

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ \overline{X^{-1}}_n - E(1/X_1) \\ \overline{\log X}_n - E\log X_1 \end{pmatrix} \to_d \underline{Z} \sim N_3(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, 1/X_1) & Cov(X_1, \log(X_1)) \\ Cov(X_1, 1/X_1) & Var(1/X_1) & Cov(1/X_1, \log X_1) \\ Cov(X_1, \log(X_1)) & Cov(1/X_1, \log X_1) & Var(\log(X_1)) \end{pmatrix}.$$

Hence by the delta method with $g(x,y,z) = (x, 1/y, \exp(z))$ so that $\nabla g(x,y,z) = \text{diag}(1, -y^{-2}, \exp(z))$ and $\nabla g(\mu, E(1/X_1), E(\log X_1)) = \text{diag}(1, -h^2, g)$, it follows that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ H_n - h \\ G_n - g \end{pmatrix} \to_d \nabla g \cdot \underline{Z} \sim N_3(0, \nabla g \Sigma \nabla g^T) \equiv N_3(0, \tilde{\Sigma}) .$$

(d) When $X \equiv X_1 \sim \text{Gamma}(r, \lambda)$, then $Y \equiv \lambda X \sim \text{Gamma}(r, 1)$, and it is easily seen that

$$\begin{aligned} &E(X^2) < \infty \qquad \text{if } \ r > 0 \\ &E((\lambda X)^{-2}) = \int_0^\infty y^{-2} \frac{y^{r-1}}{\Gamma(r)} e^{-y} dy = \frac{1}{\Gamma(r)} \int_0^\infty y^{r-3} e^{-y} dy < \infty \end{aligned}$$

if $r > 2$, and

$$E((\log \lambda X)^2) = \int_0^\infty (\log y)^2 y^{r-1} e^{-y} dy / \Gamma(r) < \infty$$

for all $r > 0$. Thus the covariance matrix Σ exists for $r > 2$; we now calculate it explicitly in this case. First,

$$\begin{aligned} E(X) &= \frac{r}{\lambda} \\ E(1/X) &= E(\lambda/(\lambda X) = \lambda E(1/Y) \\ &= \lambda \int_0^\infty y^{-1} y^{r-1} e^{-y} dy / \Gamma(r) \\ &= \lambda \int_0^\infty y^{r-2} e^{-y} dy / \Gamma(r) \\ &= \lambda \Gamma(r-1)/\Gamma(r) = \lambda/(r-1) \, , \\ E(\log X) &= E(\log(\lambda X/\lambda))) = E(\log Y) - \log \lambda \\ &= \int_0^\infty (\log y) y^{r-1} e^{-y} dy / \Gamma(r) - \log \lambda \\ &= \frac{\Gamma'(r)}{\Gamma(r)} - \log \lambda \equiv \psi(r) - \log \lambda \, . \end{aligned}$$

Next we calculate Σ:

$$\begin{aligned} Var(X) &= Var(\lambda X/\lambda) = Var(Y)/\lambda^2 = \frac{r}{\lambda} \, , \\ Var(1/X) &= \lambda^2 Var(1/Y) = \lambda^2 \{E(1/Y^2) - [E(1/Y)]^2\} \\ &= \lambda^2 \{\frac{1}{(r-1)(r-2)} - \frac{1}{(r-1)^2}\} \\ &= \lambda^2 \frac{1}{(r-1)^2(r-2)} \\ Var(\log X) &= Var(\log(\lambda X/\lambda)) = Var(\log(Y)) = \frac{\Gamma''(r)}{\Gamma(r)} - \left(\frac{\Gamma'(r)}{\Gamma(r)}\right)^2 \equiv C_r \\ Cov(X, 1/X) &= Cov(Y, 1/Y) = 1 - E(Y)E(1/Y) = 1 - \frac{r}{r-1} = \frac{-1}{r-1} \end{aligned}$$

$$
\begin{aligned}
Cov(X, \log X) &= \lambda^{-1} Cov(Y, \log Y) = \lambda^{-1}\{E(Y \log Y) - E(Y)E(\log Y)\} \\
&= \lambda^{-1}\{r\psi(r+1) - r\psi(r)\} = \frac{1}{\lambda}\,, \\
Cov(1/X, \log X) &= \lambda Cov(1/Y, \log Y) = \lambda\{E((1/Y)\log Y) - E(1/Y)E(\log Y)\} \\
&= \lambda\{\frac{\psi(r-1)}{r-1} - \frac{\psi(r)}{r-1}\} \\
&= \frac{\lambda}{r-1}\{\psi(r-1) - \psi(r)\} \equiv \frac{-\lambda}{(r-1)^2}\,,
\end{aligned}
$$

where we have used

$$
\begin{aligned}
A_r &\equiv \psi(r+1) - \psi(r) = (\log\Gamma(r+1) - \log\Gamma(r))' = (\log r)' = 1/r\,, \\
B_r &\equiv \psi(r-1) - \psi(r) = (\log\Gamma(r-1) - \log\Gamma(r))' = (-\log(r-1))' = \frac{-1}{r-1}\,.
\end{aligned}
$$

Hence

$$
\Sigma = \begin{pmatrix} r/\lambda^2 & -1/(r-1) & 1/\lambda \\ -1/(r-1) & \lambda^2/(r-1)^2(r-2) & -\lambda/(r-1)^2 \\ 1/\lambda & -\lambda/(r-1)^2 & C_r \end{pmatrix}.
$$

For the gradient ∇g we get $diag(1, -h^2, g) = diag(1, -(r-1)^2/\lambda^2, e^{\psi(r)}/\lambda)$, so it follows by direct calculation that

$$
\nabla g \Sigma (\nabla g)' = \lambda^{-2} \begin{pmatrix} r & r-1 & e^{\psi(r)} \\ r-1 & (r-1)^2/(r-2) & -e^{\psi(r)} \\ e^{\psi(r)} & -e^{\psi(r)} & e^{2\psi(r)} C_r \end{pmatrix}.
$$

(e) Let $g_2(x, y, z) \equiv z/x$. Then $\nabla g_2(x, y, z) = (-z/x^2, 0, 1/x) = (-z/x, 0, 1)/x$. when we evaluate this at $(\mu, h, g) = (r/\lambda, (r-1)/\lambda, e^{\psi(r)}/\lambda)$, we find that

$$
\nabla g_2 = (\lambda/r)(-r^{-1}e^{\psi(r)}, 0, 1)\,.
$$

Hence by the delta - method again we find that

$$
\sqrt{n}\left(\frac{G_n}{\overline{X}_n} - g/\mu\right) \to_d N(0, \nabla g_2 \tilde{\Sigma} (\nabla g_2)') = N(0, e^{2\psi(r)}\{C_r - (1/r)\})\,.
$$

Note that the scale factor λ washes out completely here, as we expect.

5. Suppose that $Y_i = \alpha + \theta'(x_i - \overline{x}) + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_i \sim (0, \sigma^2)$ are i.i.d. and the x_i's are known vectors in R^k. Equivalently, $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ where

$$
X^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 - \overline{x} & x_2 - \overline{x} & \cdots & x_n - \overline{x} \end{pmatrix}
$$

so that X is an $n \times (k+1)$ matrix. Let $\underline{\hat{\beta}}$ be the least squares estimator of $\underline{\beta} = (\alpha, \theta')'$; i.e. $\underline{\hat{\beta}} = (X^T X)^{-1} X^T \underline{Y}$. Suppose that $n^{-1}(X^T X) \to D$ where D is positive definite.

(a) What additional condition(s) do you need to impose to prove that

$$
\sqrt{n}(\hat{\beta}_n - \beta) \to_d N_{k+1}(0, \text{“}something\text{”})\,?
$$

(b) Find "something" in part (a).

Solution: (a) Let $\underline{a} \in R^{k+1}$. Now

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X)^{-1} X^T (X\beta + \epsilon) \\ &= \beta + (X^T X)^{-1} X^T \epsilon \,, \end{aligned}$$

so

$$\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n}(X^T X)^{-1} X^T \epsilon \equiv B_n \epsilon$$

where $B_n \equiv \sqrt{n}(X^T X)^{-1} X^T$ is a $(k+1) \times n$ matrix. Thus

$$\begin{aligned} a^T(\sqrt{n}(\hat{\beta} - \beta)) &= a^T B_n \epsilon \equiv b_n^T \epsilon \\ &= \sum_{i=1}^n b_{ni} \epsilon_i \equiv \sum_{i=1}^n X_{ni} \end{aligned}$$

where $b_n^T \equiv a^T B_n$ is an $1 \times n$ vector. Now we compute

$$\mu_{ni} = E(X_{ni}) = b_{ni} \cdot 0, \qquad \sigma_{ni}^2 = Var(X_{ni}) = b_{ni}^2 \sigma^2 \,,$$

and hence, using the hypothesized convergence of $n^{-1} X^T X \to D$ in the last line,

$$\begin{aligned} \sigma_n^2 &= \sigma^2 \sum_{i=1}^n b_{ni}^2 = \sigma^2 b_n^T b_n \\ &= \sigma^2 a^T B_n B_n^T a = n \sigma^2 a^T (X^T X)^{-1} (X^T X)(X^T X)^{-1} a \\ &= \sigma^2 a^T (n^{-1} X^T X)^{-1} a \to \sigma^2 a^T D^{-1} a \equiv V^2(a) > 0 \end{aligned}$$

since D is nonsingular. To establish asymptotic normality of $a^T(\sqrt{n}(\hat{\beta} - \beta))/\sigma_n$, it remains to verify the Lindeberg condition: namely

$$(2) \qquad \frac{1}{\sigma_n^2} \sum_{i=1}^n E\left\{ |X_{ni}|^2 1_{[|X_{ni}| > \delta \sigma_n]} \right\} \to 0$$

for every $\delta > 0$. But, as we have seen before, this holds if

$$(3) \qquad \max_{1 \le i \le n} |b_{ni}| \to 0 \qquad \text{as} \qquad n \to \infty \; :$$

the left side of (2) is bounded as follows:

$$\begin{aligned} & \frac{1}{\sigma_n^2} \sum_{i=1}^n b_{ni}^2 E\left\{ \epsilon_1^2 1_{[|\epsilon_1| > \delta \sigma_n / |b_{ni}|]} \right\} \\ & \quad \le \frac{1}{\sigma^2} E\left\{ \epsilon_1^2 1_{[|\epsilon_1| > \delta \sigma_n / \max_{1 \le i \le n} |b_{ni}|]} \right\} \\ & \quad \to \frac{1}{\sigma^2} \cdot 0 = 0 \end{aligned}$$

by (3), $E(\epsilon_1^2) < \infty$, and the dominated convergence theorem. Thus it follows from the Lindeberg-Feller CLT that

$$a^T(\sqrt{n}(\hat{\beta} - \beta))/\sigma_n \to_d N(0,1),$$

and since $\sigma_n^2 \to \sigma^2 a^T D^{-1} a$, this implies that

$$a^T(\sqrt{n}(\hat{\beta} - \beta)) \to_d N(0, a^T(\sigma^2 D^{-1})a)\,,$$

which in turn, via the Cramér-Wold device, implies

$$\sqrt{n}(\hat{\beta} - \beta) \to_d N_{k+1}(0, \sigma^2 D^{-1})\,.$$