Contiguity Theory – 2.

March 1, 2003

In the previous notes, we looked at LeCam's first and second lemmas. We continue our study of conitguity here with LeCam's third and fourth lemmas. All notation is as introduced previously.

LeCam's third lemma: Suppose that $T_n = (T_{n,1}, T_{n,2}, \ldots, T_{n,p})$ satisfies,

$$\begin{pmatrix} T_n \\ \log \frac{d\,Q_n}{d\,P_n} \end{pmatrix} \rightarrow_d N_{p+1}\left((\mu_{p\times 1}^T, -\sigma^2/2)^T, \Gamma\right)$$

under P_n with

$$\Gamma = \begin{pmatrix} \Sigma & \gamma_{p\times 1} \\ \gamma^T & \sigma^2 \end{pmatrix}.$$

Then $\{Q_n\}$ and $\{P_n\}$ are mutually contiguous and furthermore,

$$\begin{pmatrix} T_n \\ \log \frac{d\,Q_n}{d\,P_n} \end{pmatrix} \rightarrow_d N_{p+1}\left(((\mu+\gamma)_{p\times 1}^T, \sigma^2/2)^T, \Gamma\right),$$

under the sequence $\{Q_n\}$.

This proposition is extremely useful as it allows us to deduce the limit distribution of T_n under Q_n from its joint distribution with the log–likelihood ratio under P_n. We will use this lemma to obtain the limit distribution of the likelihood ratio statistic, the Wald statistic and the score statistic under contiguous alternatives in regular parametric problems and finally obtain approximations to the power of these tests at alternatives close to the null hypothesis.

Let us deduce the limit distribution of $\hat{\theta}_n$, the MLE under a sequence of contiguous alternatives of the form $\{P_{\theta_0+h/\sqrt{n}}\}$ in a regular parametric model. Under $\{P_{\theta_0}^n\}$ we have,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}(X_i, \theta_0) + o_p(1).$$

Also, by the LAN expansion established in the previous notes, the local log–likelihood ratio is,

$$\log\, L_n \equiv \log \frac{\Pi_{i=1}^n f(X_i, \theta_0 + h/\sqrt{n})}{\Pi_{i=1}^n f(X_i, \theta_0)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T\, \dot{l}(X_i, \theta_0) - \frac{1}{2}\, h^T\, I(\theta_0)\, h + o_p(1).$$

Thus, under $\{P_{\theta_0}^n\}$ we have the representation,

$$\begin{pmatrix} \sqrt{n}\,(\hat{\theta}_n - \theta_0) \\ \log L_n \end{pmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}(X_i, \theta_0) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \dot{l}(X_i, \theta_0) \end{bmatrix} + \begin{pmatrix} o_p(1) \\ -\frac{1}{2} h^T I(\theta_0)\, h \end{pmatrix}.$$

To handle the first term on the right-side of the above display we need a multivariate central limit theorem. The one that is suitable for our purposes can be stated as follows.

Multivariate CLT: Let $W_1, W_2, \ldots,$ be a sequence of i.i.d. random vectors with $E(W_1) = \eta$ and $\mathrm{Cov}(W_1) = \Gamma$. Then,

$$\sqrt{n}\,(\overline{W}_n - \eta) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i - \eta) \rightarrow_d N_p(0, \Gamma).$$

Here p is the dimensionality of the W_i's.

We apply this theorem with $W_i = (\dot{l}(X_i, \theta_0)^T, h^T \dot{l}(X_i, \theta_0))^T$. Check that $\eta = 0$. Also, check that

$$\Gamma = \begin{pmatrix} I(\theta_0)^{-1} & h_{p \times 1} \\ h^T & h^T I(\theta_0)\, h \end{pmatrix},$$

where

$$h = \mathrm{Cov}\left[I(\theta_0)^{-1} \dot{l}(X_1, \theta_0), h^T \dot{l}(X_1, \theta_0)\right]$$

since

$$E\left[I(\theta_0)^{-1} \dot{l}(X_1, \theta_0)\, \dot{l}(X_1, \theta_0)^T h\right] = I(\theta_0)^{-1} I(\theta_0)\, h = h.$$

Thus,

$$\begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}(X_i, \theta_0) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \dot{l}(X_i, \theta_0) \end{bmatrix} \rightarrow_d N_{(p+1) \times 1}(0_{(p+1) \times 1}, \Gamma)$$

under P_n. Consequently, under P_n,

$$\begin{pmatrix} \sqrt{n}\,(\hat{\theta}_n - \theta_0) \\ \log L_n \end{pmatrix} \rightarrow N_{(p+1) \times 1}\left(\left(0_{p \times 1}^T, -\frac{1}{2} h^T I(\theta_0)\, h\right)^T, \Gamma\right).$$

So, the hypotheses of Le Cam's third lemma are satisfied with $\sigma^2 = h^T I(\theta_0)\, h$ and $\gamma = h$ and we conclude that under the sequence Q_n,

$$\begin{pmatrix} \sqrt{n}\,(\hat{\theta}_n - \theta_0) \\ \log L_n \end{pmatrix} \rightarrow N_{(p+1) \times 1}\left(\left(h_{p \times 1}^T, -\frac{1}{2} h^T I(\theta_0)\, h\right)^T, \Gamma\right).$$

Thus,

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \rightarrow_d N_p(h, I(\theta_0)^{-1})$$

under Q_n.

Now, recall the three different statistics for testing the null hypothesis $H_0 : \theta = \theta_0$. These are, (i) The likelihood ratio statistic, $LRS = 2\,\log\,\lambda_n$ (ii) The Wald statistic, $W_n = n\,(\hat{\theta}_n - \theta_0)^T\,\hat{I}_n\,(\hat{\theta}_n - \theta_0)$ and (iii) The Score statistic,

$$R_n = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{l}(X_i,\theta_0)\right]^T\,\hat{I}_n^{-1}\,\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{l}(X_i,\theta_0)\right].$$

In the above expressions, $\hat{I}_n$ is an estimate of the information matrix based on $X_1, X_2, \ldots, X_n$. If we explicitly know the form of the information matrix $I(\theta)$ and $I(\theta)$ is continuous in θ we can take $\hat{I}_n$ to be $I(\hat{\theta}_n)$ (or even $I(\theta_0)$!!); otherwise, we can also estimate $I(\theta_0)$ by

$$\frac{1}{n}\sum_{i=1}^{n}\dot{l}(X_i,\theta_0)\,\dot{l}^T(X_i,\theta_0) \;\text{ or }\; -\frac{1}{n}\sum_{i=1}^{n}\ddot{l}(X_i,\theta_0)\,.$$

We know that W_n and LRS are asymptotically equivalent; i.e.

$$2\,\log\,\lambda_n - W_n \to_p 0 \;\text{ under }\; P_{\theta_0}^n\,;$$

in fact,

$$2\,\log\,\lambda_n = n\,(\hat{\theta} - \theta_0)^T\,I(\theta_0)\,(\hat{\theta} - \theta_0) + o_{P_{\theta_0}^n}(1)\,.$$

We seek to compute the limit distributions of these three statistics under the sequence $P_{\theta_n}^n$ where $\theta_n = \theta_0 + h\,n^{-1/2}$. By contiguity, $o_{P_{\theta_0}^n}(1)$ is also $o_{P_{\theta_n}^n}(1)$; thus, under $P_{\theta n}{}^n$ we still have the representation,

$$2\,\log\,\lambda_n = \sqrt{n}(\hat{\theta} - \theta_0)^T\,I(\theta_0)\,\sqrt{n}\,(\hat{\theta} - \theta_0) + o_p(1)\,.$$

Under $P_{\theta_n}^n$, $\sqrt{n}\,(\hat{\theta} - \theta_0) \to_d Z_h$ where $Z_h \sim N(h, I(\theta_0)^{-1})$. It follows that

$$\sqrt{n}(\hat{\theta} - \theta_0)^T\,I(\theta_0)\,\sqrt{n}\,(\hat{\theta} - \theta_0) \to_d Z_h^T\,I(\theta_0)\,Z_h\,,$$

and consequently that, $2\,\log\,\lambda_n \to_d Z_h^T\,I(\theta_0)\,Z_h$. Noting that,

$$W_n = \sqrt{n}(\hat{\theta} - \theta_0)^T\,I(\theta_0)\,\sqrt{n}\,(\hat{\theta} - \theta_0) + o_p(1)$$

under $P_{\theta_0}^n$ and arguing as above, we conclude that under $P_{\theta_n}^n$, $W_n \to_d Z_h^T\,I(\theta_0)\,Z_h$ as well.

The limit distribution of R_n can be similarly derived. To this end, first establish that under $P_{\theta_0}^n$,

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{l}(X_i,\theta_0) \\ \log\,L_n \end{pmatrix} \to_d N_{p+1}\left[\begin{pmatrix} 0_{p\times 1} \\ -\frac{1}{2}\,h^T\,I(\theta_0)\,h \end{pmatrix}, \begin{pmatrix} I(\theta_0) & I(\theta_0)\,h \\ h^T\,I(\theta_0)^T & h^T\,I(\theta_0)\,h \end{pmatrix}\right].$$

This is left as an exercise (and follows on using the multivariate CLT as before). Thus, under $P_{\theta_n}^n$,

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{l}(X_i,\theta_0) \\ \log\,L_n \end{pmatrix} \to_d N_{p+1}\left[\begin{pmatrix} I(\theta_0)\,h \\ \frac{1}{2}\,h^T\,I(\theta_0)\,h \end{pmatrix}, \begin{pmatrix} I(\theta_0) & I(\theta_0)\,h \\ h^T\,I(\theta_0)^T & h^T\,I(\theta_0)\,h \end{pmatrix}\right].$$

Thus, under $P^n_{\theta_n}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{l}(X_i, \theta_0) \to_d N_p(I(\theta_0)\, h, I(\theta_0))\,. \quad (\star\star)$$

Also, under $P^n_{\theta_n}$,

$$R_n = \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{l}(X_i, \theta_0)\right]^T I(\theta_0)^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{l}(X_i, \theta_0)\right] + o_p(1)\,.$$

As before, using contiguity and the fact $(\star\star)$ deduce that

$$R_n \to_d \tilde{Z}_h^T\, I(\theta_0)^{-1}\, \tilde{Z}_h \equiv_d Z_h^T\, I(\theta_0)\, Z_h\,,$$

where $\tilde{Z}_h \sim N_p(I(\theta_0)\, h, I(\theta_0))$. The last equality in distribution needs justification. To this end, we briefly discuss the non-central χ^2 distribution.

Non-central χ^2: Let $Z_1, Z_2, \ldots, Z_p$ be independent normal random variables each with unit variance and Z_i having mean μ_i. So, $Z = (Z_1, Z_2, \ldots, Z_p)^T \sim N_p(\mu, I_p)$ where $\mu = (\mu_1, \mu_2, \ldots, \mu_p)$ and I_p is the identity matrix. Then, the distribution of $\|Z\|^2 = Z^T Z = Z_1^2 + Z_2^2 + \ldots + Z_p^2$ depends on μ only through $\|\mu\|^2$ and $\|Z\|^2$ is said to follow a *non–central* χ^2_p distribution with non–centrality parameter $\Delta = \|\mu\|^2$. We write this distribution as $\chi^2_p(\|\mu\|^2)$.

Now, consider the expression

$$Z_h^T\, I(\theta_0)\, Z_h = (I(\theta_0)^{1/2}\, Z_h)^T\, I(\theta_0)^{1/2}\, Z_h$$

where $I(\theta_0)^{1/2}$ is the symmetric square root of $I(\theta_0)$. Check that $I(\theta_0)^{1/2}\, Z_h \sim N_p(I(\theta_0)^{1/2}\, h, I_p)$; hence,

$$Z_h^T\, I(\theta_0)\, Z_h \sim \chi^2_p(\Delta = \|I(\theta_0)^{1/2}\, h\|^2 = h^T\, I(\theta_0)\, h)\,.$$

Check that $\tilde{Z}_h^T\, I(\theta_0)^{-1}\, \tilde{Z}_h$ is also distributed as $\chi^2_p(h^T\, I(\theta_0)\, h)$.

We next turn our attention to composite hypothesis of the form $H_{0,\nu} : \nu = \nu_0$ where ν is a k–dimensional sub–parameter of θ with $k < p$; thus we write $\theta = (\nu, \eta)$. Let $\hat{\theta}_n^0 = (\nu_0, \hat{\eta}_n^0)$ denote the MLE of θ computed under H_0 and let $(\hat{\nu}_n, \hat{\eta}_n)$ denote the unrestricted MLE of θ. Let,

$$Z_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{l}(X_i, \theta) = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{l}_\nu(X_i, \nu, \eta) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{l}_\eta(X_i, \nu, \eta) \end{pmatrix}.$$

For testing $H_{0,\nu}$ we can once again consider three different statistics:

(i)

$$2\, \log\, \lambda_n = 2\, \log\, \frac{\sup_{(\nu,\eta)\in\Theta} \Pi_{i=1}^n f(X_i, \nu, \eta)}{\sup_{(\nu_0,\eta)\in\Theta} \Pi_{i=1}^n f(X_i, \nu_0, \eta)}\,;$$

this is the likelihood ratio statistic.

(ii)

$$W_n = \sqrt{n}\,(\hat{\nu}_n - \nu_0)^T\,\widehat{I_{11.2}}\,\sqrt{n}\,(\hat{\nu}_n - \nu_0)\,;$$

this is the Wald statistic. In the above expression $\widehat{I_{11.2}}$ is an (consistent) estimate of $I_{11.2} = I_{11} - I_{12}\,I_{22}^{-1}\,I_{21}$; these symbols carrying their usual meanings.

(iii)

$$R_n = Z_n(\hat{\theta}_n^0)^T\,\hat{I}^{-1}\,Z_n(\theta_0)\,;$$

this being the score statistic. In the above expression $\hat{I}$ is an estimate of the information matrix I; if $I(\theta)$ is known explicitly as a function of θ, one can use $I(\hat{\theta}_n^0)$ as an estimate; otherwise one can prescribe $n^{-1}\sum_{i=1}^n \ddot{l}(X_i, \nu_0, \hat{\eta}_n^0)$ as an estimate.

We have the following proposition.

Proposition: Under the sequence $P_{\theta_0}^n$, where $\theta_0 = (\nu_0, \eta_0) \in H_{\nu,0}$, each of the above three statistics converges in distribution to χ_k^2. Under the sequence of contiguous alternatives $P_{\theta_n}^n$ where $\theta_n = (\nu_n, \eta_n) = (\nu_0 + n^{-1/2}\,h_1, \eta_0 + n^{-1/2}\,h_2)$, each of the above statistics converges in distribution to $\chi_k^2(h_1^T\,I_{11.2}\,h_1)$.

We will not give a complete proof of this proposition but will sketch a derivation (the details of which can be filled in) for the likelihood ratio and the Wald statistics. In Homework 2, we've established that under $\{P_{\theta_0}^n\}$,

$$2\,\log\,\hat{\lambda}_n = \sqrt{n}\,(\hat{\nu}_n - \nu_0)^T\,I_{11.2}\,\sqrt{n}\,(\hat{\nu}_n - \nu_0) + o_p(1)\,.$$

Also,

$$W_n = \sqrt{n}\,(\hat{\nu}_n - \nu_0)^T\,\widehat{I_{11.2}}\,\sqrt{n}\,(\hat{\nu}_n - \nu_0) + o_p(1)\,.$$

To deduce the limit distributions of the above two statistics under $P_{\theta_n}^n$, it suffices to find the limit distribution of $\sqrt{n}(\hat{\nu}_n - \nu_0)$. Recall that under $P_{\theta_0}^n$,

$$\sqrt{n}\,(\hat{\nu}_n - \nu_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{11.2}^{-1}\,\dot{l}^\star(X_i, \theta_0) + o_p(1)\,,$$

with

$$\dot{l}^\star(x, \theta_0) = \dot{l}_\nu(x, \theta_0) - I_{12}\,I_{22}^{-1}\,\dot{l}_\eta(x, \theta_0)$$

being the efficient score function for the estimation of ν when the true parameter is θ_0. Apply the multivariate CLT as before, to conclude that,

$$\begin{pmatrix} \sqrt{n}\,(\hat{\nu}_n - \nu_0) \\ \log\,L_n \end{pmatrix} \to_d N_{p+1}\left[\begin{pmatrix} 0_{p\times 1} \\ -\frac{1}{2}\,h^T\,I(\theta_0)\,h \end{pmatrix}, \begin{pmatrix} I_{11.2}^{-1} & \gamma \\ \gamma^T & h^T\,I(\theta_0)\,h \end{pmatrix}\right],$$

where

$$\begin{aligned}
\gamma &= \mathrm{Cov}\left[I_{11.2}^{-1}\,\dot{l}^{\star}(X_1,\theta_0), h^T\,\dot{l}(X_1,\theta_0)\right] \\
&= I_{11.2}^{-1}\,\mathrm{Cov}\left[\dot{l}^{\star}(X_1,\theta_0), \dot{l}(X_1,\theta_0)\right]\, h \\
&= I_{11.2}^{-1}\left\{\mathrm{Cov}\left[\dot{l}^{\star}(X_1,\theta_0), \dot{l}_{\nu}(X_1,\theta_0)\right]\ ,\ \mathrm{Cov}\left[\dot{l}^{\star}(X_1,\theta_0), \dot{l}_{\eta}(X_1,\theta_0)\right]\right\}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\
&= I_{11.2}^{-1}\ \left[I_{11.2_{k\times k}}\ ,\ ,0_{k\times(p-k)}\right]\begin{pmatrix} h_{1_{k\times 1}} \\ h_{2\,(p-k)\times 1} \end{pmatrix} \\
&= h_1\,.
\end{aligned}$$

In the above we have used the facts that

$$\mathrm{Cov}\,[\dot{l}^{\star}(X_1,\theta_0), \dot{l}_{\eta}(X_1,\theta_0)] = 0$$

and that

$$\begin{aligned}
\mathrm{Cov}\left[\dot{l}^{\star}(X_1,\theta_0), \dot{l}_{\nu}(X_1,\theta_0)\right] &= \mathrm{Cov}\left[\dot{l}^{\star}(X_1,\theta_0), \dot{l}^{\star}(X_1,\theta_0)\right] + \mathrm{Cov}\left[\dot{l}^{\star}(X_1,\theta_0), \dot{l}_{\nu}(X_1,\theta_0) - \dot{l}^{\star}(X_1,\theta_0)\right] \\
&= I_{11.2} + 0\,.
\end{aligned}$$

These can be obtained through direct computations or by using the fact that $\dot{l}^{\star}(x,\theta_0)$ is the (orthogonal) projection of $\dot{l}_{\nu}(x,\theta_0)$ into the orthocomplement of the linear span of $\dot{l}_{\eta}(x,\theta_0)$ in the (Hilbert/inner product) space of all square integrable functions with respect to P_{θ_0}.

It follows that under $P_{\theta_n}^n$,

$$\sqrt{n}\,(\hat{\nu}_n - \nu_0) \rightarrow N(h_1, I_{11.2}^{-1})$$

as a direct application of LeCam's third lemma. Since any estimate $\widehat{I_{11.2}}$ that is consistent under $P_{\theta_0}^n$ is also consistent under $P_{\theta_n}^n$, conclued that 2 log λ_n and W_n both converge in distribution to $S_h^T\, I_{11.2}\, S_h$ with $S_h \sim N_k(h_1, I_{11.2}^{-1})$. But this has the $\chi_k^2(h_1^T\, I_{11.2}\, h)$ distribution.