Statistics 612: Superefficiency, Contiguity, LAN, Regularity, Convolution Theorems

Moulinath Banerjee

December 6, 2006

1 Some concepts from probability

Uniform integrability: A sequence of random variables $\{X_n\}$ (with X_n defined on $(\mathcal{X}_n, \mathcal{A}_n, P_n)$) is said to be *uniformly integrable* if:

$$\lim_{\lambda \to \infty} \sup_{n \geq 1} E_n(\mid X_n \mid 1(\mid X_n \mid \geq \lambda)) = 0 \,.$$

The following lemma gives a necessary and sufficient condition for uniform integrability.

Lemma 1.1 *$\{X_n\}$ is uniformly integrable if and only if the following two conditions are satisfied:*

(i) $sup_{n \geq 1} E_n \mid X_n \mid < \infty$.

(ii) For any sequence of sets $\{B_n\}$ with $B_n \in \mathcal{A}_n$, whenever $P_n(B_n)$ converges to 0, then $E_{P_n}(\mid X_n \mid 1_{B_n})$ converges to 0.

The following theorem illustrates the usefulness of uniform integrability; uniform integrability in conjunction with convergence in distribution implies convergence of moments.

Theorem 1.1 *Suppose that $X_n \in L_r(P)$ with $0 < r < \infty$ and $X_n \to_p X$. Then, the following are equivalent.*

(i) $\{\mid X_n \mid^r\}$ are uniformly integrable.

(ii) $X_n \to_r X$; in other words: $E(\mid X_n - X \mid^r)$ converges to 0.

(iii) $E \mid X_n \mid^r$ converges to $E \mid X \mid^r$.

2 Hodges' Superefficient Estimator

Let $X_1, X_2, \ldots, X_n$ be i.i.d P_θ from a one–dimensional regular parametric model $\mathcal{P}$, for which the conditions of the information inequality hold. If T_n is unbiased for estimating $q(\theta)$, then, from the information inequality:

$$\mathrm{Var}_\theta(T_n) \geq \frac{\dot{q}(\theta)^2}{n\,I(\theta)}\,.$$

Now, consider a general $\sqrt{n}$–consistent estimator S_n of $q(\theta)$, such that

$$\sqrt{n}\,(S_n - q(\theta)) \rightarrow_d N(0, V^2(\theta))\,.$$

If $\hat{\theta}_n$ is the MLE of θ, then

$$\sqrt{n}\,(q(\hat{\theta}_n) - q(\theta)) \rightarrow N(0, q'(\theta)^2/I(\theta))\,.$$

An application of the Skorokhod representation in conjunction with Fatou's lemma yields:

$$\mathrm{liminf}\,E\,[\sqrt{n}(S_n - q(\theta))]^2 \geq V^2(\theta)\,.$$

If S_n is unbiased, we have:

$$V^2(\theta) \leq \mathrm{liminf}\,n\,\mathrm{Var}_\theta(S_n)\,.$$

Now, suppose that $\lim n\,\mathrm{Var}_\theta(S_n)$ exists and equals $V^2(\theta)$. This is the case, for example, if the sequence $\{n(S_n - q(\theta))^2\}$ is uniformly integrable. Then, the information inequality would imply that:

$$V^2(\theta) \geq \frac{\dot{q}(\theta)^2}{I(\theta)}\,.$$

A question that then naturally arises is whether this inequality holds under the usual restrictions on the parametric model alone. So, for usual regular parametric models, is it possible to find $\sqrt{n}$ consistent estimators that are asymptotically normal with the variance of the limiting distribution strictly less than the bound from the information inequality? The following example due to Hodges shows that this is indeed the case.

Hodges' superefficient estimator: Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\theta, 1)$, so that $I(\theta) = 1$, identically. Let $\mid a \mid < 1$ and define:

$$T_n = \overline{X}_n\,1(\mid \overline{X}_n \mid > n^{-1/4}) + a\,\overline{X}_n\,1(\mid \overline{X}_n \mid \leq n^{-1/4})\,.$$

Then $\sqrt{n}\,(T_n - \theta) \rightarrow_d N(0, V^2(\theta))$ where $V^2(\theta) = 1 = 1/I(\theta)$ for $\theta \neq 0$ and $V^2(0) = a^2 < 1/I(0)$. Hence this beats the MLE $\overline{X}$ at 0, and is *superefficient* in this sense.

Proof: Note that $\sqrt{n}\,(\overline{X}_n - \theta) \equiv_d Z$ where Z follows $N(0,1)$, for all $n \geq 1$ and for all θ. Now,

$$\begin{aligned}
\sqrt{n}\,(T_n - \theta) &= \sqrt{n}\,(\overline{X}_n - \theta)\,1\{\mid \overline{X} \mid > n^{-1/4}\} + \sqrt{n}\,(a\,\overline{X}_n - \theta)\,1\{\mid \overline{X}_n \mid \leq n^{-1/4}\} \\
&= \sqrt{n}\,(\overline{X}_n - \theta)\,1\{\sqrt{n}\,\mid \overline{X} - \theta + \theta \mid > n^{1/4}\} \\
&\qquad + \{a\,\sqrt{n}(\overline{X}_n - \theta) + \sqrt{n}\,\theta\,(a-1)\}\,1\{\sqrt{n}\,\mid \overline{X} - \theta + \theta \mid \leq n^{1/4}\} \\
&\equiv_d Z\,1\{\mid Z + \sqrt{n}\,\theta \mid > n^{1/4}\} + \big(a\,Z + \sqrt{n}\,\theta\,(a-1)\big)\,1\{\mid Z + \sqrt{n}\,\theta \mid \leq n^{1/4}\}\,.
\end{aligned}$$

If $\theta = 0$, the above is easily seen to converge a.s. to $a\,Z$, whereas if $\theta \neq 0$ this converges a.s. to Z. This implies the result. Note that $V^2(\theta)$ is discontinuous in θ. If $\theta_n = c\,n^{-1/2}$, then under P_{θ_n} at stage n,

$$\sqrt{n}\,(T_n - \theta_n) \equiv_d Z\,1\,\{\mid Z + c \mid > n^{1/4}\} + \{aZ + c(a-1)\}\,1\{\mid Z + c \mid \leq n^{1/4}\}\,.$$

This converges to $N(c(a-1), a^2)$ as n goes to infinity. Thus, the limit here depends on c. In this sense, Hodges' superefficient estimator is *not locally regular* in the following sense.

Locally regular estimators: $\{T_n\}$ is called a sequence of locally regular estimators of θ at the point θ_0 if, for every sequence $\{\theta_n\}$ belonging to the parameter space with $\sqrt{n}\,(\theta_n - \theta_0) \rightarrow t \in \mathbb{R}^k$ (here k is the number of dimensions of the parameter θ), under P_{θ_n},

$$\sqrt{n}\,(T_n - \theta_n) \rightarrow_d \mathbb{Z}\,,$$

where the distribution of $\mathbb{Z}$ can depend on θ_0 but not on t.

The assumption of regularity is a key one for the convolution theorems we have in statistics; these are roughly results that provide lower bounds on the limiting variances (or other measures of precision) of broad classes of estimators, and in that sense provide natural thresholds for estimation. The most well known of these – Hajek's convolution theorem, to be discussed later, establishes the MLE as the most optimal estimator, among all regular ones, in the usual regular parametric models.

We next turn to the important concepts of contiguity and LAN and their consequences for inference in standard parametric models.

3 Contiguity and LAN

Consider a sequence of statistical problems with measure spaces $(\mathcal{X}_n, \mathcal{A}_n, \mu_n)$ (for the sake of concreteness and as is the case in many statistical applications, you can think of $\mathcal{X}_n$ as $\mathbb{R}^n$, $\mathcal{A}_n$ as the Borel sigma–field on $\mathbb{R}^n$ and μ_n as Lebesgue measure). Consider two sequences of probability measures $\{P_n\}$ and $\{Q_n\}$ with P_n and Q_n being defined on $\mathcal{A}_n$ and both being dominated by μ_n. Recall that this means that whenever $\mu_n(A_n) = 0$ for $A_n \in \mathcal{A}_n$ then $P_n(A_n) = Q_n(A_n) = 0$. Let p_n and q_n denote the densities of P_n and Q_n respectively with respect to μ_n (which exist by the Radon–Nikodym theroem). Define the sequence of likelihood ratios L_n where

$$L_n = \begin{cases} q_n/p_n\ ,\ p_n > 0 \\ 1\ ,\ q_n = p_n = 0 \\ n\ ,\ q_n > 0 = p_n \end{cases}.$$

Call the sequence $\{Q_n\}$ to be *contiguous with respect to* $\{P_n\}$ if, for every sequence $A_n \in \mathcal{A}_n$ for which $P_n(A_n) \rightarrow 0$, we have $Q_n(A_n) \rightarrow 0$. Contiguity is also referred to as "asymptotic absolute continuity". We write $\{Q_n\} <<_{as} \{P_n\}$. Of course contiguity of P_n with respect to Q_n is defined similarly. P_n and Q_n are mutually contiguous with respect to each other if Q_n is contiguous with respect to P_n and P_n is also contiguous with respect to Q_n.

Example 1: Contiguity is ubiquitous in parametric models. For any sufficiently regular parametric model $\{P_\theta : \theta \in \Theta\}$, the measures $\{P^n_{\theta_0 + n^{-1/2}\,h}\}$ ($P^n_{\theta_0+n^{-1/2}\,h}$ is the joint distribution of i.i.d. observations $X_1, X_2, \ldots, X_n$ drawn from $P_{\theta_0+n^{-1/2}\,h}$) and $\{P^n_{\theta_0}\}$ (defined similarly as before but with $h = 0$) are mutually contiguous. This will be established later in detail.

Example 2: Consider a regression model $Y_i = x_i\,\beta + \epsilon_i$ where the ϵ_i's are i.i.d. $N(0, \sigma^2)$ and $\sum_{i=1}^{\infty} x_i^2 < \infty$. Let P_n denote the joint distribution of $(Y_1, Y_2, \ldots, Y_n)$ under $\beta = \beta_0$ and Q_n denote the joint distribution of the data under $\beta = \beta_1$. Then, the sequences P_n and Q_n are mutually contiguous.

Example 3: For contiguous alternatives in nonparametric problems, see Problem (a) below.

Example 4: If P and Q are absolutely continuous with respect to each other, then the sequence $\{P_n\}$ where $P_n \equiv P$ and $\{Q_n\}$ where $Q_n \equiv Q$ are contiguous with respect to one another.

We will denote L_n often by dQ_n/dP_n. The following proposition describes various conditions (sufficient, necessary and sufficient) for contiguity.

Proposition 0. Contiguity and the behavior of likelihood ratios.

(i) If $L_n \rightarrow_d V$ under P_n where $E\,V = 1$, then $Q_n <<_{a.s}\, P_n$. (This proposition is known as Le Cam's first lemma and is one of the most important tools for establishing contiguity. We discuss a key corollary of this lemma very soon that we will use quite a lot subsequently).

(ii) If $L_n \rightarrow_d U$ under P_n where $P(U > 0) = 1$, then $P_n <<_{a.s}\, Q_n$.

(iii) $Q_n <<_{a.s}\, P_n$ if and only if L_n is uniformly integrable under P_n and $Q_n(p_n = 0) \rightarrow 0$.

The proposition will be (partially) established below. First, let's try to get a feel for contiguity from the following partially heuristic discussion. Let's split up the sample space $\mathcal{X}_n$ into 4 pieces – these are, (i) $A_n = \{p_n > 0\ ,\ q_n > 0\}$ (ii) $B_n = \{p_n = 0\ ,\ q_n > 0\}$ (iii) $C_n = \{q_n = 0\ ,\ p_n > 0\}$ (iv) $D_n = \{q_n = 0 = p_n\}$. On A_n, $0 < L_n < \infty$, on $B_n, L_n = n$, on $C_n, L_n = 0$(by definition), on $D_n, L_n = 1$. Now, note that the sets D_n do not really play a role in determining contiguity since they are ignorable under both P_n and Q_n, so we can forget that they exist and take B_n to be the set where p_n vanishes and C_n to be the set where q_n vanishes. For $Q_n <<_{a.s}\, P_n$, we must have $Q_n(B_n) \rightarrow 0$ (since $P_n(B_n) \equiv 0$) and for $P_n <<_{a.s.}\, Q_n$ we require that $P_n(C_n) \rightarrow 0$. (Thus, if P_n and Q_n are mutually contiguous, they must both asymptotically concentrate on A_n, the subregion of $\mathcal{X}_n$ where p_n and q_n are both positive.) If P_n is contiguous w.r.t Q_n, $P_n(L_n = 0)$ goes to 0. It is easy to deduce that $P_n(L_n = 0)$ must go to 0 under the hypothesis of (ii).

To show this: since L_n converges to U in distribution, by the Portmanteau Theorem (look up the characterization of distributional convergence in Billingsley, for example),

$$\liminf\, P_n(L_n \in (0,\infty)) \geq P(U \in (0,\infty))\,.$$

But the right-side of the display is 1 showing that the lim inf on the left side is at least as large as 1. But the sequence on the left side is a sequence of probabilities and hence bounded above by 1. Therefore the lim sup cannot exceed 1. It follows that both the lim sup and the lim inf coincide and are equal to 1 and hence $P_n(L_n \in (0,\infty))$ goes to 1. It follows that $P_n(L_n = 0)$ goes to 0.

On the other hand, it is easy to see that (i) implies that $Q_n(p_n = 0)$ must converge to 0, provided that L_n is uniformly integrable under P_n (as it must be if you look at (i) and (iii) in juxtaposition). Note that,

$$E_{P_n}(L_n) = \int_{p_n>0} \frac{q_n}{p_n}\, p_n\, d\,\mu = Q_n(p_n > 0) \leq 1\,.$$

If L_n is uniformly integrable under P_n, then $E_{P_n}(L_n)$ converges to $E\,V = 1$. Thus $Q_n(p_n > 0)$ converges to 1. It follows that $Q_n(p_n = 0) = 1 - Q_n(p_n > 0)$ converges to 0.

We now discuss a key corollary of Le Cam's first lemma.

Corollary to Le Cam's first lemma: Suppose that $\log\, L_n \rightarrow_d \tilde{L}$ under P_n where $\tilde{L}$ follows $N(-\sigma^2/2, \sigma^2)$. Then the sequences of probability measures P_n and Q_n are mutually contiguous.
Proof: We have $L_n \rightarrow_d \exp(\tilde{L})$ under P_n. Since $\exp(\tilde{L}) \equiv U$ is positive with probability 1, by (ii) of Proposition 0, P_n is contiguous with respect to Q_n. To show the converse, note that $E(\exp(\tilde{L}) = \phi_L(1)$ where ϕ_L is the moment–generating function of L. Now,

$$\phi_L(t) = \exp\left(\mu\, t + \frac{1}{2}\,\sigma^2\, t^2\right)\,,$$

where μ is the mean of L and σ^2 the variance. But $\mu = -\sigma^2/2$, so

$$\phi_L(1) = \exp\left(-\frac{1}{2}\,\sigma^2 + \frac{1}{2}\,\sigma^2\right) = 1\,.$$

The desired conclusion now follows from (i) of Proposition 0.

Let us illustrate the above corollary on Example 2. We can write,

$$\begin{aligned}
\log\, L_n &= \log\, \frac{\Pi_{i=1}^n \left(\frac{1}{\sqrt{2\,\pi\,\sigma^2}}\right)\, \exp\,\left[-\frac{1}{2\,\sigma^2}\, \sum_{i=1}^n (Y_i - x_i\,\beta_1)^2\right]}{\Pi_{i=1}^n \left(\frac{1}{\sqrt{2\,\pi\,\sigma^2}}\right)\, \exp\,\left[-\frac{1}{2\,\sigma^2}\, \sum_{i=1}^n (Y_i - x_i\,\beta_0)^2\right]} \\
&= -\frac{1}{2\,\sigma^2}\, \sum_{i=1}^n \left[(Y_i - x_i\,\beta_1)^2 - (Y_i - x_i\,\beta_0)^2\right] \\
&= \sum_{i=1}^n \frac{Y_i\, x_i(\beta_1 - \beta_0)}{\sigma^2} - \frac{1}{2}\, \sum_{i=1}^n \frac{x_i^2\, (\beta_1^2 - \beta_0^2)}{2\,\sigma^2}\,.
\end{aligned}$$

Now, under P_n, $Y_i\, x_i \sim N(x_i^2\, \beta_0, x_i^2\, \sigma^2)$ and are independent and simple algebra shows that

$$\sum_{i=1}^{n} \frac{Y_i\, x_i(\beta_1 - \beta_0)}{\sigma^2} \sim N\left(\frac{\beta_0\,(\beta_1 - \beta_0)}{\sigma^2} \sum_{i=1}^{n} x_i^2,\ \ \frac{(\beta_1 - \beta_0)^2}{\sigma^2} \sum_{i=1}^{n} x_i^2\right).$$

Thus,

$$\begin{aligned}
\log\, L_n &\sim N\left(\left[\beta_0(\beta_1 - \beta_0) - \frac{\beta_1^2 - \beta_0^2}{\sigma^2}\right] \frac{\sum_{i=1}^n x_i^2}{\sigma^2},\ \ \frac{(\beta_1 - \beta_0)^2}{\sigma^2} \sum_{i=1}^{n} x_i^2\right) \\
&\equiv N\left(-\frac{\sum_{i=1}^n x_i^2}{2\,\sigma^2}\,(\beta_1 - \beta_0)^2,\ \ \frac{\sum_{i=1}^n x_i^2}{\sigma^2}\,(\beta_1 - \beta_0)^2\right) \\
&\equiv N\left(-\frac{\tau_n^2}{2},\ \tau_n^2\right)
\end{aligned}$$

where

$$\tau_n^2 = \frac{\sum_{i=1}^n x_i^2}{\sigma^2}\,(\beta_1 - \beta_0)^2\,.$$

Thus,

$$\log\, L_n \rightarrow_d N\left(\lim_n -\frac{\tau_n^2}{2},\ \lim_n \tau_n^2\right) \equiv N\left(-\frac{\tau^2}{2},\ \tau^2\right),$$

under P_n, where

$$\tau^2 = \frac{\sum_{i=1}^\infty x_i^2}{\sigma^2}\,(\beta_1 - \beta_0)^2 < \infty\,.$$

It follows from a direct application of the corollary above that P_n and Q_n are mutually contiguous. One can similarly work out the limit distribution of log L_n under Q_n. Check for yourselves that under Q_n, log L_n converges to $N(\tau^2/2, \tau^2)$.

Example 4 discussed: To show here that Q_n is contiguous with respect to P_n, consider the likelihood ratio L_n; this is $(q/p)\,1(p > 0) + n\,1(p = 0)$, where q and p are densities with respect to a common dominating measure μ. Since the set $\{p = 0\}$ has 0 probability under $P_n \equiv P$, L_n is q/p a.s. P and is independent of n. Checking uniform integrability boils down to checking integrability of L_n; this is simply $\int_{p>0} q\, d\,\mu \leq 1$. Also $Q_n(p_n = 0) = Q(p = 0) = 0$ since Q is dominated by P and $P(p = 0) = 0$. Part (iii) of Proposition 0 now says that Q_n is contiguous w.r.t P_n. Similarly, the fact that P_n is contiguous with respect to Q_n may be established.

The converse, i.e. P and Q being mutually contiguous implies that they are mutually absolutely continuous, follows on taking a sequence of sets $\{A_n\}$ with $A_n = A$ and using the basic definition of contiguity.

Example 5: Here is a scenario, where P_n is contiguous with respect to Q_n but not the other way around. Let P_n be the distribution of the maximum of n i.i.d. $U(0, 1)$ random variables, and Q_n be the distribution of the maximum of n i.i.d $U(0, 1 + 1/n)$ random variables. To show that P_n is contiguous with respect to Q_n, we can write $\tilde{L}_n$, the likelihood ratio of P_n with respect to Q_n. We have:

$$\tilde{L}_n(x_1, x_2, \ldots, x_n) = (1 + 1/n)^{-n} \, 1(\max_i x_i < 1) + n \, 1(\max_i x_i > 1 + 1/n) \,.$$

For all n, $\tilde{L}_n$ is bounded by 1 a.s. with respect to Q_n; hence $\tilde{L}_n$ is uniformly integrable with respect to Q_n. Also $P_n(q_n = 0) = 0$, since the support of q_n contains the support of p_n. It follows that P_n is contiguous with respect to Q_n, by Part (iii) of Proposition 0. To show that the converse is not true, consider the set $A_n = \{\max_i x_i > 1\}$. This set A_n has 0 probability under P_n; the probability of this set under Q_n is simply the probability that $\max_i X_i$ lies between 1 and $1 + 1/n$, where X_i's are i.i.d. $U(0, 1 + 1/n)$. This is simply $1 - (1 + 1/n)^{-n}$ and converges to $1 - 1/e > 0$. Hence $Q_n(A_n)$ does not go to 0.

Proof of Proposition 0, (i): Consider $B_n \in \mathcal{A}_n$ with $P_n(B_n) \to 0$. Consider now, the problem of testing $H_0 : f = p_n$ versus $H_1 : f = q_n$ based on an observation x_n from the sample space $\mathcal{X}_n$. We seek to obtain the most powerful test of level $\alpha_n = P_n(B_n)$ for this problem. By the NP–lemma, the most powerful test ϕ_n must be of the form

$$\phi_n \equiv 1\,\{L_n > k_n\} + \gamma_n \, 1\{L_n = k_n\} \,,$$

and $E_{P_n} \, \phi_n = \alpha_n = P_n(B_n)$. Consider the test function $\tilde{\phi}_n = 1(x_n \in B_n)$. Then $E_{P_n} \, \tilde{\phi}_n$ has level α_n and hence $E_{Q_n} \, \phi_n \geq E_{Q_n} \, \tilde{\phi}_n = Q_n(B_n)$. So, to show that $Q_n(B_n) \to 0$ (the defining property of contiguity), it suffices to show that $E_{Q_n} \, \phi_n$ goes to 0.

Now, for any fixed $0 < y < \infty$,

$$\begin{aligned}
E_{Q_n} \, \phi_n &= Q_n(\phi_n \, 1\{L_n \leq y\}) + Q_n(\phi_n \, 1\{L_n > y\}) \\
&= P_n\,(\phi_n \, L_n \, 1\{L_n \leq y\}) + Q_n(\phi_n \, 1\{L_n > y\}) \\
&\leq y \, P_n(\phi_n) + Q_n(1\,\{L_n > y\}) \\
&= y \, P_n(\phi_n) + 1 - Q_n(1\,\{L_n \leq y\}) \\
&= y \, P_n(\phi_n) + 1 - P_n(L_n \, 1\,\{L_n \leq y\}) \,.
\end{aligned}$$

We want to show that given any $\epsilon > 0$, the quantity in the last step of the above display can be made smaller than ϵ for all sufficiently large n. Now, choose y to be a continuity point of the distribution of L, so large that $E\,(L\,1\{L \leq y\})$ is greater than $1 - \epsilon/2$; since $E\,L = 1$ and the continuity points of L are dense in $\mathbb{R}$ this can be arranged. Now $L_n \, 1(L_n \leq y)$ converges to $L\,1\{L \leq y\}$ by Slutsky's theorem (note that $1\,\{L_n \leq y\}$ converges in probability to $1\,\{L \leq y\}$). Since the random variables involved are all bounded (and therefore uniformly integrable), expectations converge as well, and therefore, for sufficiently large n, $1 - P_n(L_n \, 1\,\{L_n \leq y\})$ can be made less than $\epsilon/2$. The term $y \, P_n(\phi_n)$ can be made eventually smaller than $\epsilon/2$ as well, since $\alpha_n = P_n(\phi_n)$ converges to 0 by hypothesis. This finishes the proof. $\Box$.

Proof of Proposition 0, (iii): First, note that for $B_n \in \mathcal{A}_n$, we have:

$$\begin{aligned} Q_n(B_n) &= \int 1_{B_n}\, q_n\, d\,\mu_n \\ &= \int 1_{B_n \cap \{p_n > 0\}}\, q_n\, d\,\mu_n + \int 1_{B_n \cap \{p_n = 0\}}\, q_n\, d\,\mu_n \\ &= \int 1_{B_n}\, L_n\, d\,P_n + \int 1_{B_n \cap \{p_n = 0\}}\, q_n\, d\,\mu_n\,. \end{aligned}$$

Thus, on one hand, we have:

$$Q_n(B_n) \leq \int 1_{B_n}\, L_n\, d\,P_n + Q_n(p_n = 0) \quad (\star),$$

and on the other,

$$Q_n(B_n) \geq \int 1_{B_n}\, L_n\, d\,P_n \quad (\star\star).$$

First suppose that L_n is uniformly integrable and $Q_n(p_n = 0)$ converges to 0. Now, take a sequence of sets B_n, such that $P(B_n)$ converges to 0. From Lemma (1.1), it follows that $\int 1_{B_n}\, L_n\, d\,P_n$ converges to 0 . Then $(\star)$ implies that $Q_n(B_n)$ goes to 0. It follows that Q_n is contiguous with respect to P_n.

Conversely, suppose that Q_n is contiguous with respect to P_n. Then $Q_n(p_n = 0)$ must go to 0, since $P_n(p_n = 0) = 0$. Take any sequence of sets B_n such that $P_n(B_n)$ goes to 0. Then $Q_n(B_n)$ goes to 0. From $(\star\star)$ it follows that $\int 1_{B_n}\, L_n\, d\,P_n$ converges to 0. Also $E_{P_n}(L_n)$ is bounded by 1. Hence, uniform integrability of L_n follows by Lemma (1.1). $\square$.

Problems: Here are a set of problems that illustrate contiguity in various different ways.

(1) (a) Let $P_n = N(0,1)$ and $Q_n = N(\mu_n, 1)$. Show that the sequences P_n and Q_n are mutually contiguous if the sequence $\{\mu_n\}$ is bounded.

(b) Let P_n and Q_n denote the distribution of the mean of a sample of size n from the $N(0,1)$ and the $N(\theta_n, 1)$ distribution, respectively. Show that P_n and Q_n are mutually contiguous if $\theta_n = O(n^{-1/2})$.

(c) This part does not need to be turned in, but what about the converse? Suppose that it is given in (a) and (b) that the sequences P_n and Q_n are mutually contiguous. Can you conclude that the sequence μ_n is bounded?

(2) (a) Suppose that $\|P_n - Q_n\| \to 0$ where $\| \;\; \|$ is the total variation distance: i.e. $\|P_n - Q_n\| = \sup_{A \in \mathcal{A}} \mid P(A) - Q(A) \mid$. Show that P_n and Q_n are mutually contiguous with respect to each other.

(b) Given $\epsilon > 0$ find an example of mutually contiguous sequences but for which the total variation distance is at least $1-\epsilon$. This exercise shows that it is wrong to think of contiguous sequences as being close. (Try measures supported on two points).

(3) **Contiguity in nonparametric problems:** Consider i.i.d. data $\{X_i, Z_i\}_{i=1}^n$ where (X_1, Z_1) has the following distribution: Z_1 follows some Lebesgue density p_Z, and $X_1 \mid Z_1 = z$ has density $p(\cdot; \psi(z))$ where $p(x, \theta)$ is a regular parametric family of models. Here ψ is some unknown (dependence) function that needs to be estimtated.

Consider a fixed point z_0 such that both p_Z and ψ are continuously differentiable in a neighborhood of z_0. Let B_n be a sequence of continuous functions that vanish outside the interval $(-c, c)$ and suppose that B_n converges uniformly to a (continuous) function B (that also vanishes outside of $(-c, c)$). Define a sequence of functions ψ_n in the following manner: $\psi_n(z) = \psi_0(z) + n^{-\alpha} B_n(n^{1-2\,\alpha}(z - z_0))$, with $0 < \alpha < 1/2$. Let Q_n denote the joint distribution of $\{(X_i, Z_i)\}_{i=1}^n$, when the dependence function is ψ_n, and P_n the joint distribution of $\{(X_i, Z_i)\}_{i=1}^n$, when the dependence function is ψ_0. Show that the sequences $\{P_n\}$ and $\{Q_n\}$ are mutually contiguous.

We will now discuss Le Cam' second lemma and its applications which will involve establishing a LAN (local asymptotic normality) expansion of the log–likelihood ratios in a regular parametric model. (Example 1).

The Set-Up of Le Cam's second lemma: Consider a measure space $(\mathcal{X}, \mathcal{A}, \mu)$ and let $\underline{X}_n = (X_1, X_2, \ldots, X_n) \in \mathcal{X}_n \equiv \mathcal{X}^n$ with the product sigma–field $\mathcal{A}^n$ and probability measure $\mu_n \equiv \mu^n$ defined on it. Consider two sequences of measures $\{P_n\}$ and $\{Q_n\}$ where $P_n = \Pi_{i=1}^n P_{ni}$, P_{ni} being some measure on $(\mathcal{X}, \mathcal{A})$ that is dominated by μ and has density f_{ni} and where $Q_n = \Pi_{i=1}^n Q_{ni}$, Q_{ni} being some measure on $(\mathcal{X}, \mathcal{A})$ that is dominated by μ and has density g_{ni}. The density of P_n with respect to μ^n is

$$p_n(x_1, x_2, \ldots, x_n) = \Pi_{i=1}^n f_{ni}(x_i)$$

and the density of Q_n with respect to μ^n is

$$q_n(x_1, x_2, \ldots, x_n) = \Pi_{i=1}^n g_{ni}(x_i)\,.$$

Statistically, you can think of n independent observations from an underlying sample space with two possible candidates for the distribution of the i'th observation at stage n – either P_{ni} (think of this as the null disitribution at stage n) or Q_{ni} (think of this as the alternative at stage n) and P_n and Q_n denote the joint distributions of the observed random vector $\underline{X}_n$ under the null and the alternative respectively (at stage n). Fundamental to a study of the contiguity of these two sequences of probability measures is an understanding of the likelihood ratio or equivalently its logarithm,

$$\log\, L_n = \sum_{i=1}^n \log\, \left(\frac{g_{ni}}{f_{ni}}(X_i)\right)\,.$$

A way to determine contiguity is to study the limiting behavior of L_n (or equivalently $\log L_n$). Le Cam's second lemma gives a way of doing this by analysing quantities of the type,

$$W_n \equiv 2 \sum_{i=1}^{n} \left\{ \frac{g_{ni}^{1/2}}{f_{ni}^{1/2}}(X_i) - 1 \right\} \equiv \sum_{i=1}^{n} T_{ni} \,.$$

This is the sum of independent random variables and provided variances do not blow up, there is hope of invoking Central Limit Theorems. Note that each T_{ni} has finite variance under P_n , since

$$E_{P_n} \, (T_{ni} + 1)^2 = E \, \left(g_{ni}(X_i)/f_{ni}(X_i) \right) = \int_{f_{ni}>0} g_{ni} \, d\,\mu \leq 1 \,.$$

Le Cam's second lemma reduces the proof of asymptotic normality of $\log L_n$ to the problem of establishing asymptotic normality of the sequence W_n.

Le Cam's second lemma: Suppose that the following condition (the UAN (uniform asymptotic negligibility) condition) holds :

$$\max_{1 \leq i \leq n} P_n \left(\left| \frac{g_{ni}}{f_{ni}}(X_i) - 1 \right| > \epsilon \right) \rightarrow 0 \,.$$

Suppose also that W_n converges in distribution to $N(-\sigma^2/4, \sigma^2)$ for some $\sigma^2 > 0$. Then,

$$\log L_n - (W_n - \sigma^2/4) = o_{P_n}(1)$$

and hence

$$\log L_n \rightarrow_d N(-\sigma^2/2, \sigma^2)$$

under P_n showing thereby that Q_n and P_n are mutually contiguous.

We will not prove this lemma here. The proof is long and provided in Wellner's notes. However, we will study an important consequence of this lemma. We specialise to the case of i.i.d. observations. Thus, at stage n, $X_1, X_2, \ldots, X_n$ are i.i.d f under P_n (thus $f_{ni} \equiv f$) and under Q_n, they are i.i.d. f_n (thus $g_{ni} \equiv f_n$). But before, we proceed further, we need some preliminary notions. These are discussed below.

The notion of the derivative of a function taking values in $\mathcal{L}_2(\mu)$ is defined in a way analogous to that in multivariate calculus. Formally, a map $\psi : \Theta \rightarrow \mathcal{L}_2(\mu)$, where Θ is an open subset of $\mathbb{R}^p$, is said to be differentiable in quadratic mean (QMD) at θ_0 with derivative vector V in $\mathcal{L}_2(\mu)^p$ if

$$\psi(\theta_0 + \epsilon) - \psi(\theta_0) - \epsilon^T \, V = o(\|\epsilon\|) \,,$$

i.e.

$$\frac{\|\psi(\theta_0 + \epsilon) - \psi(\theta_0) - \epsilon^T \, V\|_\mu}{\|\epsilon\|} \rightarrow 0 \,.$$

The vector V is the *total derivative* of the map ψ at the point θ_0 and can be viewed as a linear map $D_\psi(\theta_0)$ from $\mathbb{R}^p$ to $\mathcal{L}_2(\mu)$ which is defined as:

$$D_\psi(\theta_0)(\eta) = \eta^T V \in \mathcal{L}_2(\mu)\,,$$

for $\eta \in \mathbb{R}^p$. Since the vector V has the interpretation of a derivative (the derivative of ψ at the point θ_0) in an extended sense, a natural question arises as to whether V is actually the pointwise derivative of $\psi(\cdot,\theta)$ with respect to θ at the point θ_0 (assuming adequate smoothness of ψ in θ). In other words, is

$$D_\psi(\theta_0) = \frac{\partial}{\partial\,\theta}\,\psi(\cdot,\theta)\,|_{\theta=\theta_0}\ \ ?$$

The following lemma gives sufficient conditions for this to be the case.

Lemma 3.1 *Let $\theta \mapsto \psi(\cdot,\theta)$ be a map from $\mathbb{R}^p$ to $\mathcal{L}_2(\mu)$. If $\bigtriangledown_\theta\,\psi(\cdot,\theta)$ exists for almost all x (w.r.t μ), for θ in some neighborhood of θ_0 and the function*

$$\theta \mapsto \int \|\bigtriangledown_\theta\,\psi(x,\theta)\|^2\; d\,\mu(x)\ \ ,$$

is continuous at θ_0, then ψ is QMD at θ_0 with derivative vector $V \equiv V_{\theta_0} \equiv \bigtriangledown_\theta\,\psi(x,\theta_0)$.

Note that if h is a density function on $\mathcal{X}$, then $h^{1/2}$ is in $\mathcal{L}_2(\mu)$. For our current purpose, we will be specially interested in studying the function $s(\theta) = f(\cdot,\theta)^{1/2}$ where $\{f(\cdot,\theta) : \theta \in \Theta\}$ is a regular parametric model. Now suppose that the function $\theta \mapsto s(\theta)$ is QMD with derivative vector $\dot{s}(\theta) \in \mathcal{R}^p$, p being the dimension of θ. Thus,

$$\|f(\theta_0+\epsilon)^{1/2} - f(\theta_0)^{1/2} - \epsilon^T\,\dot{s}(\cdot,\theta_0)\| = o(\|\epsilon\|)\,. \tag{3.1}$$

This is often referred to as *Hellinger differentiability* of the model at the point θ. (Formally, the Hellinger distance between two probability densities p and q on $\mathcal{X}$ is defined as $H(p,q) = \|p^{1/2} - q^{1/2}\|_\mu$.) This will be seen to have an important bearing on the local log–likelihood ratios of the model. To that end, we require the following proposition.

Proposition 1: Suppose that we have a sequence of densities $\{f_n\}$ and a fixed density f such that,

$$\left\|\sqrt{n}(f_n^{1/2} - f^{1/2}) - \delta\right\|_2 \to 0$$

as $n \to \infty$ for some $\delta \in L_2(\mu)$. Thus the sequence

$$\frac{f_n^{1/2} - f^{1/2}}{1/\sqrt{n}} \to_{n\to\infty} \delta$$

in the $\mathcal{L}_2(\mu)$ metric. Then,

$$E_f\left(\frac{\delta}{f^{1/2}}\right) = 0$$

and

$$\log L_n - \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{2\,\delta}{f^{1/2}}(X_i) - \frac{1}{2}\, \|2\,\delta\|^2 \right) = o_{P_n}(1)\,.$$

Consequently,

$$\log L_n \to_d N\left(-\frac{1}{2}\, \|2\,\delta\|^2\,, \|2\,\delta\|^2 \right)\,.$$

It follows that the sequence of probability measures Q_n and P_n are mutually contiguous.

Proof: In the following proof $IP(f, g)$ will denote the usual inner product between functions f and g in $L_2(\mu)$.

We have,

$$W_n = \sum_{i=1}^{n} T_{ni}$$

where $T_{n1}, T_{n2}, \ldots, T_{nn}$ are independent and $T_{ni} = 2\,(f_n^{1/2}(X_i)/f^{1/2}(X_i) - 1)$. We will first show that as claimed, $E_f(\delta/f^{1/2}) = 0$. Then, by the CLT it will follow that,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{2\,\delta}{f^{1/2}}(X_i) \to N\left(0, \mathrm{Var}\left(\frac{2\,\delta}{f^{1/2}}\right)\right) \equiv N(0, \|2\,\delta\|^2)\,,$$

since,

$$\mathrm{Var}\left(\frac{2\,\delta}{f^{1/2}}\right) = E_f\,\left(\frac{4\,\delta^2}{f}\right) = \int 4\,\delta^2/f\ f\,d\mu = \|2\,\delta\|^2\,.$$

Next, we will show that,

$$W_n - \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta}{f^{1/2}}(X_i) + \|\delta\|^2 = o_{P_n}(1)\,.\ \ (\star)$$

This will imply that

$$W_n \to_d N(-\|\delta\|^2, \|2\,\delta\|^2) \equiv N(-\sigma^2/4, \sigma^2) \ \text{ under }\ P_n\,,$$

where $\sigma^2 = \|2\,\delta\|^2$. By Le Cam's second lemma,

$$\log L_n - (W_n - \frac{\sigma^2}{4}) = o_{P_n}(1)\,.$$

Now $(\star)$ readily implies that,

$$\log L_n - \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{2\,\delta}{f^{1/2}}(X_i) - \frac{1}{2}\, \|2\,\delta\|^2 \right) = o_{P_n}(1)\,.$$

What we have omitted above is the verification of the UAN condition. This will be done last.
Step 1. Show that, $E_f(\delta/f^{1/2})(X_1) = 0$. Now, using

$$\|\sqrt{n}(f_n^{1/2} - f^{1/2}) - \delta\|^2 \to 0 \ \ (0)$$

we easily conclude that,

$$\|\sqrt{n}(f_n^{1/2} - f^{1/2})\|^2 \equiv n\,\|(f_n^{1/2} - f^{1/2})\|^2 \to \|\delta\|^2 . \ \ (1)$$

Thus,

$$\|(f_n^{1/2} - f^{1/2})\|^2 \to 0 \ \ (2) .$$

Now,

$$1 = \int f_n \, d\,\mu = \int \left(f^{1/2} + \frac{\delta}{\sqrt{n}} + r_n \right)^2 d\,\mu \, ,$$

where

$$r_n \equiv f_n^{1/2} - f^{1/2} - \frac{\delta}{\sqrt{n}} = o(n^{-1/2})$$

by (0). Thus,

$$\begin{aligned}
1 &= \left\| f^{1/2} + \frac{\delta}{\sqrt{n}} + r_n \right\|^2 \\
&= \left\| f^{1/2} + \frac{\delta}{\sqrt{n}} \right\|^2 + \|r_n\|^2 + 2\,IP\left(f^{1/2} + \frac{\delta}{\sqrt{n}}, r_n \right) \\
&= \|f^{1/2}\|^2 + 2\,IP\left(f^{1/2}, \frac{\delta}{\sqrt{n}} \right) + \frac{\|\delta\|^2}{n} + \|r_n\|^2 + 2\,IP\left(f^{1/2} + \frac{\delta}{\sqrt{n}}, r_n \right) \\
&= 1 + 2\,\frac{1}{\sqrt{n}} \int \delta\, f^{1/2}\, d\,\mu + o(n^{-1/2}) \, ,
\end{aligned}$$

since $\|\delta\|^2$ and $\|r_n\|^2$ are $O(n^{-1})$ and $IP\left(f^{1/2} + \frac{\delta}{\sqrt{n}}, r_n \right)$ is $o(n^{-1/2})$. It follows that

$$0 = 2\,\frac{1}{\sqrt{n}} \int \delta\, f^{1/2}\, d\,\mu + o(n^{-1/2})$$

or equivalently

$$0 = 2 \int \delta\, f^{1/2}\, d\,\mu + n^{1/2}\, o(n^{-1/2}) .$$

But $n^{-1/2}\, o(n^{-1/2})$ is $o(1)$ showing that,

$$\int \delta\, f^{1/2}\, d\,\mu = E_f \left(\frac{\delta}{f^{1/2}} \right) = 0 .$$

Step 2. To show that

$$W_n - \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta}{f^{1/2}}(X_i) + \|\delta\|^2 = o_{P_n}(1)$$

it suffices to prove, by Markov's inequality, that

$$V_n^2 \equiv E_{P_n}\left[W_n - \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta}{f^{1/2}}(X_i) + \|\delta\|^2\right]^2 = o_{P_n}(1)\,.$$

Now,

$$\begin{aligned} V_n^2 &= E\left[\sum_{i=1}^{n} 2\left(\frac{f_n^{1/2}}{f^{1/2}}(X_i) - 1\right) - \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta}{f^{1/2}}(X_i) + \|\delta\|^2\right]^2 \\ &= 4\,E\left(K_1 + K_2 + \ldots + K_n\right)^2, \end{aligned}$$

where $K_1, K_2, \ldots, K_n$ are i.i.d. random variables and

$$K_i = \left(\frac{f_n^{1/2}}{f^{1/2}}(X_i) - 1\right) - \frac{1}{\sqrt{n}}\,\frac{\delta}{f^{1/2}}(X_i) + \frac{\|\delta\|^2}{2\,n}\,.$$

Thus,

$$V_n^2 = n\,E(K_1^2) + n(n-1)\left(E(K_1)\right)^2.$$

To show that V_n^2 goes to 0 it suffices to show that both $E(K_1)$ and $E(K_1^2)$ are $o(n^{-1})$. Now,

$$\begin{aligned} E(K_1) &= E\left(\frac{f_n^{1/2}}{f^{1/2}}(X_1) - 1 - \frac{1}{\sqrt{n}}\,\frac{\delta}{f^{1/2}}(X_1) + \frac{\|\delta\|^2}{2\,n}\right) \\ &= E\left(\frac{f_n^{1/2}}{f^{1/2}}(X_1)\right) - 1 + \frac{\|\delta\|^2}{2\,n} \\ &= \int \sqrt{f_n(x)\,f(x)}\,d\,\mu(x) - 1 + \frac{\|\delta\|^2}{2\,n} \\ &= -\frac{1}{2}\left[2 - 2\int f_n^{1/2}\,f^{1/2}\,d\,\mu\right] + \frac{\|\delta\|^2}{2\,n} \\ &= -\frac{1}{2}\,\|f_n^{1/2} - f^{1/2}\|^2 + \frac{\|\delta\|^2}{2\,n}\,. \end{aligned}$$

Thus,

$$n\,E(K_1) = \frac{1}{2}\left(-n\,\|f_n^{1/2} - f^{1/2}\|^2 + \|\delta\|^2\right) \to 0\,,$$

by (1). Next,

$$\begin{aligned} E(K_1^2) &= E_f\left[\frac{f_n^{1/2}}{f^{1/2}}(X_1) - 1 - \frac{1}{\sqrt{n}}\,\frac{\delta}{f^{1/2}}(X_1) + \frac{\|\delta\|^2}{2\,n}\right]^2 \\ &= \int\left(f_n^{1/2} - f^{1/2} - \frac{\delta}{\sqrt{n}} + \frac{\|\delta\|^2}{2\,n}\,f^{1/2}\right)^2 d\,\mu \end{aligned}$$

$$
\begin{aligned}
&= \left\| r_n + \frac{\|\delta\|^2}{2\,n}\, f^{1/2} \right\|^2 \\
&\leq 2\left[\|r_n\|^2 + \frac{\|\delta\|^4}{4\,n^2}\, \|f^{1/2}\|^2 \right] \\
&= o(n^{-1})\,,
\end{aligned}
$$

since $\|r_n\|^2$ is $o(n^{-1})$. This completes the proof of Step 2.

Step 3. Verification of the UAN condition. We have

$$
\begin{aligned}
\max_{1\leq i\leq n} P_n\left(\left|\frac{g_{ni}}{f_{ni}}(X_i) - 1\right| > \epsilon\right) &= P_f\left(\left|\frac{f_n}{f}(X_1) - 1\right| > \epsilon\right) \\
&\leq \frac{1}{\epsilon}\, E_f\left(\left|\frac{f_n}{f}(X_1) - 1\right|\right) \\
&= \frac{1}{\epsilon} E_f\left(\left|\frac{f_n^{1/2}}{f^{1/2}}(X_1) - 1\right|\,\left|\frac{f_n^{1/2}}{f^{1/2}}(X_1) + 1\right|\right) \\
&\leq \frac{1}{\epsilon}\left(E_f\left[\frac{f_n^{1/2}}{f^{1/2}}(X_1) - 1\right]^2 E_f\left[\frac{f_n^{1/2}}{f^{1/2}}(X_1) + 1\right]^2\right)^{1/2} \\
&= \frac{1}{\epsilon}\left[\int \left(f_n^{1/2} - f^{1/2}\right)^2 d\,\mu \int \left(f_n^{1/2} + f^{1/2}\right)^2 d\,\mu\right]^{1/2} \\
&\rightarrow 0\,,
\end{aligned}
$$

since

$$
\int (f_n^{1/2} + f^{1/2})^2\, d\,\mu \equiv \|f_n^{1/2} + f^{1/2}\|^2 \leq 2\,(\|f_n^{1/2}\|^2 + \|f_n^{1/2}\|^2) = 4
$$

and

$$
\int \left(f_n^{1/2} - f^{1/2}\right)^2 = \|f_n^{1/2} - f^{1/2}\|^2 \rightarrow 0
$$

by (2). This proves the UAN condition.

LAN in a Hellinger–differentiable parametric model: Recall the definition of Hellinger differentiability of a regular parametric model. This is illustrated in display (3.1). This implies that for a fixed vector h, we have,

$$
\frac{\left\|p(\cdot,\theta_0 + n^{-1/2}\,h)^{1/2} - p(\cdot,\theta_0)^{1/2} - n^{-1/2}\,h^T\,\dot{s}(\cdot,\theta_0)\right\|_\mu}{n^{-1/2}\,\|h\|} \rightarrow 0\,;
$$

equivalently

$$
\left\|\sqrt{n}\left(p\left(\cdot,\theta_0 + \frac{h}{\sqrt{n}}\right)^{1/2} - p(\cdot,\theta_0)^{1/2}\right) - h^T\,\dot{s}(\cdot,\theta_0)\right\|_\mu \rightarrow 0\,.
$$

Thus, we are in the set-up of Proposition 1 with $f_n \equiv p(\cdot, \theta_0 + h/\sqrt{n})$, $f \equiv p(\cdot, \theta_0)$ and $\delta = h^T s(\cdot, \theta_0)$. Hence, using Proposition 1 we obtain,

$$\log L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{2\, h^T \dot{s}(X_i, \theta_0)}{f(X_i, \theta_0)^{1/2}} - \frac{1}{2} \|2\, h^T \dot{s}(\cdot, \theta_0)\|^2 + o_{P_n}(1)\,.$$

Consequently,

$$\log L_n \equiv \frac{d\, P^n_{\theta_0 + h/\sqrt{n}}}{d\, P^n_{\theta_0}} (X_1, X_2, \ldots, X_n) \to_d N\left(-\frac{1}{2} \|2\, h^T \dot{s}(\cdot, \theta_0)\|^2, \|2\, h^T \dot{s}(\cdot, \theta_0)\|^2\right)\,.$$

It remains to identify $\dot{s}(\cdot, \theta)$. For nice regular parametric models, the sufficient conditions in Lemma 3.1 hold good and $\dot{s}(x, \theta)$ is simply the partial derivative of $f^{1/2}(x, \theta)$ with respect to θ. Thus,

$$\dot{s}(x, \theta) = \frac{\partial}{\partial\, \theta} f(x, \theta)^{1/2} = f(x, \theta)^{1/2}\, \frac{1}{2} \frac{\partial}{\partial\, \theta} \log f(x, \theta)^{1/2} \equiv f(x, \theta)^{1/2}\, \frac{1}{2}\, \dot{l}(x, \theta)\,.$$

Going back to the conditions of Lemma 3.1, the existence of $\bigtriangledown_\theta\, \dot{s}(x, \theta)$ for μ almost all x for every θ is guaranteed by the underlying regularity conditions and the continuity hypothesis is easy to verify. For simplicity, if θ is 1–dimensional then,

$$\int \| \bigtriangledown_\theta\ \dot{s}(x, \theta)\|^2\, d\,\mu = \int \frac{1}{4}\, \dot{l}(x, \theta)^2\, f(x\, \theta)\, d\,\mu(x) = \frac{1}{4}\, E_\theta(\dot{l}(X_1, \theta)^2) = \frac{I(\theta)}{4} < \infty\,,$$

and continuity of $I(\theta)$ at θ_0 guarantees Hellinger differentiability of the model at θ_0.

Thus, on plugging in the expression for $\dot{s}$ obtained above, $\log L_n$ has the asymptotic linear representation (known as **the LAN expansion**) given by,

$$\log L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h^T\, \dot{l}(X_i, \theta) - \frac{1}{2}\, h^T\, I(\theta)\, h + o_{P_n}(1)\,.$$

By Proposition 0, it follows that the sequence of measures $\{P^n_{\theta_0 + h\, n^{-1/2}}\}$ and $\{P^n_{\theta_0}\}$ are mutually contiguous, as claimed in Example 1. We next come to LeCam's third lemma.

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

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

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,

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

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. We first provide a proof of (a simplified version of) this result.

The simpler version: Suppose that a statistic T_n satisfies:

$$\mathcal{L}((T_n, \log\, L_n)^T \mid P_n) \to \mathcal{L}((T, \log\, L)^T)\,,$$

where

$$(T, \log\, L)^T \sim N_2\left(\left(\begin{array}{c} \mu \\ -\sigma^2/2 \end{array}\right), \left(\begin{array}{cc} \tau^2 & c \\ c & \sigma^2 \end{array}\right)\right).$$

Then, $\mathcal{L}((T_n, \log\, L_n)^T \mid Q_n) \to \mathcal{L}((T + c, \log\, L + \sigma^2)^T)$ which has distribution:

$$N_2\left(\left(\begin{array}{c} \mu + c \\ \sigma^2/2 \end{array}\right), \left(\begin{array}{cc} \tau^2 & c \\ c & \sigma^2 \end{array}\right)\right).$$

Proof: Since $\mathcal{L}(\log\, L_n \mid P_n) \to \mathcal{L}(\log\, L) = N(-\sigma^2/2, \sigma^2)$, it follows that $E\, L = 1$, and hence Q_n is contiguous w.r.t. P_n by Proposition 0 (part (i)). Hence, by part (iii) of Proposition 0, L_n is uniformly integrable and $Q_n(p_n = 0) \to 0$.

To show that under Q_n, the distribution of $(T_n, \log\, L_n)$ converges to the claimed limit, we proceed as follows. Take any bounded continuous function f from $\mathbb{R}^2 \to \mathbb{R}$. We want to show that $E_{Q_n}\, f(T_n, \log\, L_n) \to E\, f(T + c, \log\, L + \sigma^2)$. Now, we have:

$$\begin{aligned} E_{Q_n}\, f(T_n, \log\, L_n) &= E_{Q_n}\left[f(T_n, \log\, L_n)\, 1\{p_n > 0\}\right] + E_{Q_n}\left[f(T_n, \log\, L_n)\, 1\{p_n = 0\}\right] \\ &= E_{P_n}\left[f(T_n, \log\, L_n)\, L_n\right] + E_{Q_n}\left[f(T_n, \log\, L_n)\, 1\{p_n = 0\}\right] \\ &\to E\left[f(T, \log\, L) L\right]. \end{aligned}$$

The last step here follows on noting that $E_{Q_n}\left[f(T_n, \log\, L_n)\, 1\{p_n = 0\}\right]$ is bounded up to a constant by $Q_n(p_n = 0)$ which goes to 0 (established above). Since f is bounded and L_n is uniformly integrable, so is the sequence $f(T_n, \log\, L_n)\, L_n$. Furthermore, under P_n the sequence $f(T_n, \log\, L_n)\, L_n$ converges in distribution to $f(T, \log\, L)\, L$ by the continuous mapping theorem. It follows from Theorem 1.1 that $E_{P_n}\left[f(T_n, \log\, L_n)\, L_n\right] \to E\left[f(T, \log\, L) L\right]$.

It remains to show that $E\left[f(T, \log\, L) L\right] = E\left[f(T + c, \log\, L + \sigma^2)\right]$, whence it will follow that the limit distribution of $(T_n, \log\, L_n)$ under Q_n is precisely the distribution of $(T + c, \log\, L + \sigma^2)$. To this end, first observe that:

$$\mathcal{L}\left(T \mid \log\, L\right) = \mathcal{L}\left(\frac{c}{\sigma^2}\left(\log\, L + \frac{\sigma^2}{2}\right) + \tilde{Z}\right)$$

where $\mathcal{L}(\tilde{Z}) = N(\mu, \sigma^2(1-\rho^2))$ where $\rho = c/\sigma\,\tau$ and is independent of $\log L$. This is a direct consequence of the fact that for a normal random vector (X, Y), we can write $Y = E(Y \mid X) + (Y - E(Y \mid X))$, where $Y - E(Y \mid X)$ is independent of X and $E(Y \mid X)$ is an explicitly computable linear function of X. It follows that:

$$\mathcal{L}(T + c \mid \log\, L) = \mathcal{L}\left(\frac{c}{\sigma^2}\left(\log\, L + \sigma^2 + \frac{\sigma^2}{2}\right) + \tilde{Z}\right).$$

Now, consider

$$\begin{aligned} E\, f(T, \log\, L) L &= E\left[E\left(f(T, \log\, L)\, L \mid \log\, L\right]\right. \\ &= E\left\{L\, E\left[f\left(\frac{c}{\sigma^2}\left(\log\, L + \frac{\sigma^2}{2}\right) + \tilde{Z}, \log\, L\right) \mid \log\, L\right]\right\} \\ &\equiv E\left[L\, \xi(\log\, L)\right], \end{aligned}$$

say. Let g denote the $N(-\sigma^2/2, \sigma^2)$ density and $\tilde{g}$ denote the $N(\sigma^2/2, \sigma^2)$ density. Now,

$$E\left[L\, \xi(\log\, L)\right] = \int e^w\, \xi(w)\, g(w)\, dw = \int \xi(w)\, \tilde{g}(w)\, dw\,.$$

But since $\tilde{g}$ is the distribution of $\log\, L + \sigma^2$, it follows that

$$E\left[L\, \xi(\log\, L)\right] = E\left[\xi(\log\, L + \sigma^2)\right].$$

Hence, we can write:

$$\begin{aligned} E\, f(T, \log\, L)\, L &= E\left\{E\left[f\left(c + \frac{c}{\sigma^2}\left(\log\, L + \frac{\sigma^2}{2}\right) + \tilde{Z}, \log\, L + \sigma^2\right) \mid \log\, L\right]\right\} \\ &= E\, f(T + c, \log\, L + \sigma^2)\,. \end{aligned}$$

This finishes the proof.

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,

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

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,

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

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 $\text{Cov}(W_1) = \Gamma$. Then,

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

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

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

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,

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

under P_n. Consequently, under P_n,

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

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,

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

Thus,

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

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.

$$2 \log \lambda_n - W_n \rightarrow_p 0 \;\; \text{under} \;\; P^n_{\theta_0}\,;$$

in fact,

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

We seek to compute the limit distributions of these three statistics under the sequence $P^n_{\theta_n}$ where $\theta_n = \theta_0 + h\, n^{-1/2}$. By contiguity, $o_{P^n_{\theta_0}}(1)$ is also $o_{P^n_{\theta_n}}(1)$; thus, under $P^n_{\theta_n}$ 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 $P^n_{\theta_n}$, $\sqrt{n}\, (\hat{\theta} - \theta_0) \rightarrow_d Z_h$ where $Z_h \sim N(h, I(\theta_0)^{-1})$. It follows that

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

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

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

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

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

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

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

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

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

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

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 $\hat{\theta}^0_n = (\nu_0, \hat{\eta}^0_n)$ 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)

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

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,

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

where

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

In the above we have used the facts that

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

and that

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

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 $2\,\log\,\lambda_n$ 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.

4 Convolution Theorem

See Sections 4 and 5 of Chapter 3 of Wellner's notes.