Contiguity Theory – 1.

February 10, 2003

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 $P_n(A_n) \to 0$, we have $Q_n(A_n) \to 0$. Contiguity is also referred to as "asymptotic absolute continuity". We write $\{Q_n\} <<_{as} \{P_n\}$ (this is different from the symbol used in Wellner and in general. I can't recall the latex command for that symbol right now). 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 non-regular non-parametric problems, see Problem 4 of Homework 2.

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 $EV = 1$, 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 $Q_n(p_n = 0) \rightarrow 0$.

We will not bother with the proofs of these lemmas, which are rather technical. Instead 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 Q_n is contiguous with respect to P_n, the likelihood ratio of Q_n w.r.t P_n cannot escape to infinity in the limit, with positive probability under Q_n. On the other hand, if P_n is contiguous w.r.t Q_n, $P_n(L_n = 0)$ goes to 0; so, the likelihood ratio in the limit cannot concentrate at 0 with positive probability under P_n. This is compatible with the assertion made in (ii) of Proposition 0 (it better be!!) which says that P_n is contiguous w.r.t. Q_n, the limit (in law) of L_n under P_n is a strictly positive random variable. 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 $\log L_n \to_d \tilde{L}$ 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 $L_n \to_d \exp(\tilde{L})$ 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 $N(\tau^2/2, \tau^2)$.

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 $g_{ni}\equiv f_n$). Before we proceed any further, a few preliminaries.

By $L_2(\mu)$ we will denote the set of real–valued functions defined on $\mathcal{X}$ that are square–integrable. Thus,

$$L_2(\mu)=\left\{f:\int_{\mathcal{X}} f^2\,d\,\mu<\infty\right\}\,.$$

We will view this as a linear inner-product space, where the inner product between f and g, which we denote by the slightly unconventional $IP(f,g)$, is given by

$$IP(f,g)=\int fg\,d\mu\,.$$

The size or norm of an element f is measured by,

$$\|f\|_\mu=\sqrt{IP(f,f)}\equiv\sqrt{\int f^2\,d\,\mu}\,.$$

Often we will just denote the norm of f by $\|f\|$, dropping the subscript μ. Two elements f and g are said to be orthogonal if $IP(f,g)=0$. We will be using the identity,

$$\|f+g\|^2 = \|f\|^2 + \|g\|^2 + 2\,IP(f,g)\,.$$

The distance between two functions f and g is measured by

$$d(f,g) = \|f-g\|\,.$$

This is indeed a valid metric, since for a normed linear space (as $\mathcal{L}_2(\mu)$ is), we have the *triangle inequality* which says that,

$$\|x\| + \|y\| \leq \|x+y\|\,.$$

Another crucial inequality, which in particular, implies that the function $f \mapsto \|f\|$ is a continuous function is that

$$|\|h_1\| - \|h_2\|| \leq \|h_1 - h_2\|\,.$$

To talk about continuity and differentiability of a function taking values in μ we need a notion of convergence; this is defined in the usual way for metric spaces. A sequence $\{g_n\}$ in $\mathcal{L}_2(\mu)$ conveges to g in $\mathcal{L}_2(\mu)$ if $\|g_n - g\|_\mu \to 0$. Let ψ be a real-valued or vector-valued function defined on $\mathcal{L}_2(\mu)$. We say that ψ is continuous at the point g if for every sequence g_n converging to g, $\psi(g_n)$ converges to $\psi(g)$ (in the usual Euclidean metric).

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 $\psi : \Theta \to \mathcal{L}_2(\mu)$, 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.

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

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 V which is defined as:

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

for $\eta \in \mathcal{L}_2(\mu)$. 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

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

The following lemma (this is lemma 17.2.9. of Keener) gives sufficient conditions for this to be the case.

Lemma 0.1 *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*

$$\int \|\bigtriangledown_\theta \psi(x, \theta)\|^2 \, d\mu(x) < \infty \ ,$$

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,

$$\|f(\theta_0 + \epsilon)^{1/2} - f(\theta_0)^{1/2} - h^T \dot{s}(\cdot, \theta)\| = o(\|\epsilon\|) \,. \tag{0.1}$$

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

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

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: Here,

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

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, $E_f(d/f^{1/2}) = 0$. 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,

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

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,

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

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,

$$\begin{aligned}
E(K_1^2) &= E_f\,\left[\frac{f_n^{1/2}}{f^{1/2}}(X_1) - 1 - \frac{1}{\sqrt{n}} \frac{\delta}{f^{1/2}}(X_1) + \frac{\|\delta\|^2}{2\,n} \right]^2 \\
&= \int \left(f_n^{1/2} - f^{1/2} - \frac{\delta}{\sqrt{n}} + \frac{\|\delta\|^2}{2\,n}\,f^{1/2} \right)^2\,d\,\mu \\
&= \left\| r_n + \frac{\|\delta\|^2}{2\,n}\,f^{1/2} \right\|^2 \\
&\leq 2\,\left[\|r_n\|^2 + \frac{\|\delta\|^4}{4\,n^2}\,\|f^{1/2}\|^2 \right] \\
&= o(n^{-1})\,,
\end{aligned}$$

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

Step 3. Verification of the UAN condition. We have

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

since

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

and

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

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 (0.1). This implies that for a fixed vector h, we have,

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

equivalently

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

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

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

Consequently,

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

It remains to identify $\dot{s}(\cdot,\theta)$. For nice regular parametric models, the sufficient conditions in Lemma 0.1 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 0.1, the existence of $\bigtriangledown_\theta \dot{s}(x,\theta)$ for μ almost all x for every θ is generally guaranteed by the underlying regularity conditions and the finiteness of the integral is equivalent to the existence of the information matrix $I(\theta)$ as a well-defined finite quantity. For simplicity, if θ is 1–dimensional then,

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

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,

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