Chapter 2

Contiguity

SECTION 1 defines contiguity as a property of two sequences of probability measures, $\{\mathbb{P}_n\}$ and $\{\mathbb{Q}_n\}$, that transfers $o_p(\cdot)$ assertions under $\{\mathbb{P}_n\}$ to $o_p(\cdot)$ under $\{\mathbb{Q}_n\}$. Various equivalent forms of the definition, which are useful for establishing contiguity, are derived.

SECTION 2 establishes conditions for contiguity of sequences of product measures for a sequence of parametric alternatives.

SECTION 3 derives useful consequences of contiguity, which transfer convergence in distribution under $\{\mathbb{P}_n\}$ to analogous convergence under $\{\mathbb{Q}_n\}$. In a special asymptotic normal case (the so-called Third Lemma of Le Cam) the change in limiting distribution involves only a shift in the vector of means.

1. Definition and equivalences

In many asymptotic problems one needs to study estimators under various sequences of probability models. For example, in Chapter 1, we saw that the Hodges estimator θ_n^* behaves badly under a sequence of alternatives $\theta_n := \theta_0 + \delta/\sqrt{n}$. For a careful analysis we would have to consider behavior of $\theta_n^*(x_1, \ldots, x_n)$ under the product measure $\mathbb{P}_{n,\theta_n} := P_{\theta_n}^n$ on $\mathcal{X}^n$. Ignoring the limitations of the heuristic arguments, we already know a lot about the behavior under the product measure $\mathbb{P}_{n,\theta_0} := P_{\theta_0}^n$. We could repeat the arguments with θ_n taking over the role played by θ_0, following closely the steps used for the θ_0 analysis, to derive the asymptotics under the alternative. There is, however, a more elegant approach, whereby the analysis is concentrated into a study of the density $dP_{\theta_n}/dP_{\theta_0}$. The underlying magic is called ***contiguity***, a subtle (see the Notes in Section 5) invention of Le Cam (1960).

As you will learn in the next few Chapters, contiguity lies at the root of a number of well known asymptotic facts. Rather than following the tradition of presenting one monolithic theorem collecting together all the interesting equivalences and consequences, I will split the ideas into a sequence of small lemmas, each focussing on one key idea.

The contiguity idea is not restricted to independent sampling. It makes sense—and has interesting consequences—for any two sequences $\{\mathbb{P}_n\}$ and $\{\mathbb{Q}_n\}$ of probability measures. For each n, both $\mathbb{P}_n$ and $\mathbb{Q}_n$ should live on the same space $(\Omega_n, \mathcal{F}_n)$, but there need be no constraint on how the spaces change with n.

For example, $\mathbb{P}_n$ and $\mathbb{Q}_n$ might be the joint distributions of random vectors with dimension k_n, corresponding to parametric models whose dimensions change with sample size.

As a convenient abreviation, I will write $o(1;\mathbb{P}_n)$ instead of "$o_p(1)$ under the sequence of models $\{\mathbb{P}_n\}$", with analogous interpretations for $O_p(1;\mathbb{P}_n)$ and other stochastic order symbols.

<1> **Definition.** *A sequence $\{\mathbb{Q}_n\}$ is said to be contiguous to $\{\mathbb{P}_n\}$ if every sequence of random variables $\{Y_n\}$ of order $o_p(1;\mathbb{P}_n)$ is also of order $o_p(1;\mathbb{Q}_n)$. That is, if $\mathbb{P}_n\{|Y_n| > \eta\} \to 0$ for each $\eta > 0$, then $\mathbb{Q}_n\{|Y_n| > \eta\} \to 0$ for each $\eta > 0$. Write $\{\mathbb{Q}_n\} \lhd \{\mathbb{P}_n\}$, or just $\mathbb{Q}_n \lhd \mathbb{P}_n$, to denote contiguity.*

Rewriting the limiting requirements of the definition as explicit δ, ϵ inequalities, we get a more cumbersome (but more versatile) characterization.

<2> **Lemma.** *The contiguity $\mathbb{Q}_n \lhd \mathbb{P}_n$ is equivalent to the assertion: for each $\epsilon > 0$ there exists an n_0 and a $\delta > 0$, both depending on ϵ, such that*

$$\sup\{\mathbb{Q}_n F : \mathbb{P}_n F < \delta \text{ and } n \geq n_0\} \leq \epsilon,$$

with the supremum ranging over all $n \geq n_0$ and all sets F in $\mathcal{F}_n$ for which $\mathbb{P}_n F < \delta$.

Proof. The δ, ϵ condition implies contiguity: if $\mathbb{P}_n\{|Y_n| > \eta\} \to 0$ then $\{|Y_n| > \eta\}$ is eventually one of the F sets over which the supremum is taken.

If the δ, ϵ condition is violated then, for some $\epsilon > 0$, there exists a subsequence $\mathbb{N}_1$ and sets F_n in $\mathcal{F}_n$ for $n \in \mathbb{N}_1$ such that $\mathbb{Q}_n F_n > \epsilon$ and $\mathbb{P}_n F_n \to 0$ along $\mathbb{N}_1$. Take Y_n as the indicator of F_n for $n \in \mathbb{N}_1$, and put $Y_n := 0$ for other values of n, to define a sequence $\{Y_n\}$ that violates the contiguity property. $\square$

REMARK. Most authors use a sequential analog of Lemma <2> as the definition of contiguity. That is, they define $\mathbb{Q}_n \lhd \mathbb{P}_n$ to mean that $\mathbb{Q}_n F_n \to 0$ for each sequence $\{F_n\}$ for which $\mathbb{P}_n F_n \to 0$.

<3> **Example.** Let $\mathbb{P}_n$ denote the $N(\alpha_n, 1)$ distribution and $\mathbb{Q}_n$ denote the $N(\beta_n, 1)$ distribution, both on the real line. Under what conditions on the sequences of constants $\{\alpha_n\}$ and $\{\beta_n\}$ do we have $\mathbb{Q}_n \lhd \mathbb{P}_n$?

If the sequence $\delta_n := \beta_n - \alpha_n$ is not bounded then contiguity fails. For example, suppose $\delta_n \to \infty$ along some subsequence $\mathbb{N}_1$. Define $F_n := [\beta_n, \infty)$ if $n \in \mathbb{N}_1$ and $F_n := \emptyset$ otherwise. Then $\mathbb{P}_n F_n \to 0$ but $\mathbb{Q}_n F_n = 1/2$ along the subsequence.

We will soon have elegant ways to show that $\mathbb{Q}_n \lhd \mathbb{P}_n$ if δ_n is bounded in absolute value by some finite constant C. For the moment, brute force will suffice. Then

$$\begin{aligned}
\mathbb{Q}_n F &= (2\pi)^{-1/2} \int \{x \in F\} \exp\left(-(x-\beta_n)^2/2\right) dx \\
&= (2\pi)^{-1/2} \int \{x \in F\} \exp\left(\delta_n(x-\alpha_n) - \delta_n^2/2 - (x-\alpha_n)^2/2\right) dx \\
&\leq \mathbb{P}_n^x \left(\{x \in F\} \exp\left(C|x-\alpha_n|\right)\right).
\end{aligned}$$

If we split the last integrand according to whether $|x - \alpha_n| \le M$ or not, for some constant M, then make the change of variable $z = x - \alpha_n$ in the second contribution, we get a bound for the expectation:

$$\exp(CM)\mathbb{P}_n F + (2\pi)^{-1/2} \int \{|z| > M\} \exp\Big(C|z| - z^2/2\Big)\, dz.$$

If M is large enough, the second contribution is smaller than $\epsilon/2$. The first contribution is also smaller than $\epsilon/2$ if $\mathbb{P}_n F < \epsilon \exp(-CM)/2$. □

The δ, ϵ formulation of contiguity broadens its applicability to cover sequences of events that are eventually small for $\mathbb{P}_n$, not just those sequences with $\mathbb{P}_n$ probabilities tending to zero. The fine difference is of the type that distinguishes between $o_p(\cdot)$ and $O_p(\cdot)$ assertions.

<4> **Lemma.** *The contiguity $\mathbb{Q}_n \lhd \mathbb{P}_n$ is equivalent to the assertion: every sequence of random variables $\{Y_n\}$ of order $O_p(1; \mathbb{P}_n)$ is also of order $O_p(1; \mathbb{Q}_n)$.*

Proof. Under contiguity, if M is chosen so that $\mathbb{P}_n\{|Y_n| > M\} < \delta$ eventually then $\mathbb{Q}_n\{|Y_n| > M\} \le \epsilon$ eventually, by virtue of Lemma <2>.

For the converse, suppose $Y_n = o_p(1; \mathbb{P}_n)$. Then (see Problem [3]) there exists a sequence $\{\delta_n\}$ of positive numbers converging to zero for which $\mathbb{P}_n\{|Y_n| > \delta_n\} \to 0$. The sequence $\{Y_n/\delta_n\}$ is of order $O_p(1; \mathbb{P}_n)$, and hence also of order $O_p(1; \mathbb{Q}_n)$. That is, $Y_n = O_p(\delta_n; \mathbb{Q}_n) = o_p(1; \mathbb{Q}_n)$, as required for contiguity. □

REMARK. A sequence of real random variables $\{Y_n\}$ of order $O_p(1; \mathbb{P}_n)$ is sometimes said to be stochastically bounded (under $\{\mathbb{P}_n\}$), or uniformly tight. Such a sequence must have a subsequence that converges in distribution to a probability measure concentrated on $\mathbb{R}$. For real-valued random variables the proof is easy: a Cantor diagonalization argument applied to the sequence of distribution functions evaluated on a countable dense subset of $\mathbb{R}$. The analog for more general spaces is often called the Prohorov/Le Cam theorem (UGMTP §7.5).

The preceding Lemma shows that contiguity is a matter of inheritance of a $O_p(1)$: to verify contiguity we could check the $O_p(1; \mathbb{Q}_n)$ property for all $O_p(1; \mathbb{P}_n)$ sequences. The next characterization simplifies the task by allowing us to check the inheritance for just one particular case, the sequence of ***likelihood ratios***, which is automatically $O_p(1; \mathbb{P}_n)$ but is $O_p(1; \mathbb{Q}_n)$ only when contiguity holds.

It pays to be quite precise in the definition of a likelihood ratio, to avoid later ambiguities concerning singular parts. Suppose both $\mathbb{P}$ and $\mathbb{Q}$ are probability measures defined on the same space $(\Omega, \mathcal{F})$. There is a unique decomposition of $\mathbb{Q}$ into a sum $\mathbb{Q}_a + \mathbb{Q}_s$, where $\mathbb{Q}_a$ is absolutely continuous with respect to $\mathbb{P}$ and $\mathbb{Q}_s$ is singular with respect to $\mathbb{P}$, that is, $\mathbb{Q}_s$ concentrates on a set $\mathcal{N}_{\mathbb{P}}$ with zero $\mathbb{P}$ measure. At the slight risk of misleading you into thinking that $\mathbb{Q}$ equals $\mathbb{Q}_a$, I will follow conventional usage by writing $d\mathbb{Q}/d\mathbb{P}$ for the density of $\mathbb{Q}_a$ with respect to $\mathbb{P}$. At least for nonnegative measurable functions f,

<5>
$$\mathbb{Q} f = \mathbb{Q}_a f + \mathbb{Q}_s f = \mathbb{P}\left(f \frac{d\mathbb{Q}}{d\mathbb{P}} \mathcal{N}_{\mathbb{P}}^c \right) + \mathbb{Q}\left(f \mathcal{N}_{\mathbb{P}} \right)$$

Of course the $\mathcal{N}_{\mathbb{P}}^c$ is irrelevant for the $\mathbb{P}$ contribution, but it sometimes helps to be reminded indirectly that the density applies only to the contribution from $\mathbb{Q}_a$.

If both $\mathbb{P}$ and $\mathbb{Q}$ are absolutely continuous with respect to a measure λ, with densities p and q, then we can take

$$\frac{d\mathbb{Q}}{d\mathbb{P}} := (q/p)\{p \neq 0\} \qquad \text{and} \qquad \mathcal{N}_{\mathbb{P}} := \{p = 0\}.$$

In the Statistics literature, the density $d\mathbb{Q}/d\mathbb{P}$ is usually called the ***likelihood ratio*** and is often denoted by a letter like L or $\mathcal{L}$. The definition of the likelihood ratio on the set $\mathcal{N}_{\mathbb{P}}$ has no effect on the equality <5>. We could even define it as $+\infty$, taking $L := (q/p)\{p \neq 0\} + \infty\{p = 0\}$. This definition would lead to some economy of notation. For example, with likelihood ratios $\{L_n\}$ for sequences $\{\mathbb{P}_n\}$ and $\{\mathbb{Q}_n\}$, a statement like $L_n = O_p(1; \mathbb{Q}_n)$ would imply both

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} = O_p(1; \mathbb{Q}_n) \qquad \text{and} \qquad \mathbb{Q}_n \mathcal{N}_{\mathbb{P}_n} \to 0.$$

The set $\mathcal{N}_{\mathbb{P}_n} = \{L_n = \infty\}$ would get absorbed into the set $\{L_n > M\}$ for each finite constant M.

REMARK. After some experimentation on live audiences, I have decided that the possibilities for confusion outweigh the notational disadvantages of the more explicit treatment of singular parts of the $\{\mathbb{Q}_n\}$. I will always regard the likelihood ratio as a real-valued random variable.

<6> **Definition.** *For probability measures $\mathbb{Q}$ and $\mathbb{P}$ defined on the same space, the likelihood ratio is defined as $L := d\mathbb{Q}/d\mathbb{P}$, the density of the absolutely continuous part of $\mathbb{Q}$ with respect to $\mathbb{P}$. The value of L on the singularity set $\mathcal{N}_{\mathbb{P}}$ can be defined arbitrarily.*

In what follows, L_n will always denote the likelihood ratio $d\mathbb{Q}_n/d\mathbb{P}_n$, and $\mathcal{N}_n$ will denote the singularity set $\mathcal{N}_{\mathbb{P}_n}$. Note that $\mathbb{P}_n L_n = \mathbb{Q}\mathcal{N}_n^c$, so that the sequence $\{L_n\}$ is always $O_p(1; \mathbb{P}_n)$.

<7> **Lemma.** *$\mathbb{Q}_n \lhd \mathbb{P}_n$ if and only if both $L_n = O_p(1; \mathbb{Q}_n)$ and $\mathbb{Q}_n(\mathcal{N}_n) \to 0$.*

Proof. From Lemma <4>, contiguity and the automatic $O_p(1; \mathbb{P}_n)$ property for L_n, deduce that $L_n = O_p(1; \mathbb{Q}_n)$. And $\mathbb{Q}_n \mathcal{N}_n \to 0$ because $\mathbb{P}_n \mathcal{N}_n \equiv 0$.

Conversely, for a fixed finite M, and an F in $\mathcal{F}_n$,

$$\begin{aligned} \mathbb{Q}_n F &= \mathbb{P}_n \left(L_n F \mathcal{N}_n^c \{L_n \leq M\}\right) + \mathbb{Q}_n \left(F \mathcal{N}_n^c \{L_n > M\}\right) + \mathbb{Q}_n \left(F \mathcal{N}_n\right) \\ &\leq M \mathbb{P}_n F + \mathbb{Q}_n \{L_n > M\} + \mathbb{Q}_n \mathcal{N}_n. \end{aligned}$$

If $L_n = O_p(1; \mathbb{Q}_n)$, we can find M to make $\mathbb{Q}_n\{L_n > M\} < \epsilon/2$ eventually. Then the choice $\delta = \epsilon/(2M)$ leads to the characterization of contiguity in Lemma <2>. □

REMARK. If I had adopted the convention that $L_n = \infty$ on $\mathcal{N}_n$, the proof would have been slightly shorter. The case where $\mathbb{Q}_n \mathcal{N}_n = 1$, with $L_n \equiv 0$, shows that the condition $L_n = O_p(1; \mathbb{Q}_n)$ by itself would not suffice for contiguity.

<8> **Example.** For the $\mathbb{P}_n$ and $\mathbb{Q}_n$ from Example <3>,

$$L_n = \exp\left(\delta_n(x - \alpha_n) - \delta_n^2/2\right) \qquad \text{where } \delta_n := \beta_n - \alpha_n.$$

Under $\mathbb{Q}_n$ the random variable $x - \alpha_n$ has a $N(\delta_n, 1)$ distribution. If $\{\delta_n\}$ is bounded then $\delta_n(x - \alpha_n)$, and hence L_n, is of order $O_p(1; \mathbb{Q}_n)$. □

The automatic $O_p(1;\mathbb{P}_n)$ property of $\{L_n\}$ implies existence of subsequences that converge in distribution. Suppose L, on some probability space $(\Omega,\mathcal{A},\mathbb{P})$, represents the limit distribution along some such subsequence $\{L_n : n \in \mathbb{N}_1\}$. Be careful: $\mathbb{P}$ need not be a limit of the $\mathbb{P}_n$ in any sense; the probability $\mathbb{P}$ exists only to give L a distribution. The image of $\mathbb{P}_n$ under L_n converges, along the subsequence, to the image of $\mathbb{P}$ under L, that is, $L_n(\mathbb{P}_n) \rightsquigarrow L(\mathbb{P})$.

For each finite constant M,

$$\mathbb{P}(L \wedge M) = \lim_{n\in\mathbb{N}_1} \mathbb{P}_n(L_n \wedge M) \le \liminf_{n\in\mathbb{N}_1} \mathbb{P}_n L_n \le 1.$$

Let M increase to infinity to deduce that $\mathbb{P}L \le 1$. Equality here will translate into a $O_p(1;\mathbb{Q}_n)$ property of $\{L_n : n \in \mathbb{N}_1\}$. Equality for all such subsequences will translate into contiguity.

<9> **Lemma.** *The contiguity $\mathbb{Q}_n \lhd \mathbb{P}_n$ is equivalent to the equality $\mathbb{P}L = 1$ for every L that is a limit in distribution of a subsequence of the likelihood ratios $\{L_n\}$ under $\{\mathbb{P}_n\}$.*

Proof. Problems [1] and [2] show (via subsequencing arguments) that there is no loss of generality in considering only the case where L_n itself converges in distribution to some random variable L on a probability space $(\Omega,\mathcal{A},\mathbb{P})$.

Fix a finite M with $\mathbb{P}\{L = M\} = 0$. Fix $\epsilon > 0$. From the definition of L_n,

<10> $$\mathbb{Q}_n\{L_n \le M\} = \mathbb{P}_n L_n\{L_n \le M\} + \mathbb{Q}_n\mathcal{N}_n\{L_n \le M\}$$

If $\mathbb{Q}_n \lhd \mathbb{P}_n$, then by Lemma <7> we can choose M so large that the left-hand side of <10> is eventually greater that $1-\epsilon$ and the second term on the right-hand side is less than ϵ. In the limit, via the Continuous Mapping Theorem (UGMTP §7.1) we get $1 \ge \mathbb{P}L \ge \mathbb{P}L\{L \le M\} \ge 1-2\epsilon$, whence $\mathbb{P}L = 1$.

Conversely, if $\mathbb{P}L = 1$ we may choose M so that $\mathbb{P}L\{L \le M\} \ge 1-\epsilon$, which implies that $\mathbb{P}_n L_n\{L_n \le M\} > 1-\epsilon$ eventually. When this inequality holds we have both $\mathbb{Q}\{L_n \le M\} > 1-\epsilon$ and $\mathbb{Q}_n\mathcal{N}_n < \epsilon$. □

The last Lemma has an interesting interpretation, which lends support to the idea that contiguity is a form of asymptotic absolute continuity. For simplicity, suppose L_n converges in distribution under $\mathbb{P}_n$ to an L on $(\Omega,\mathcal{F},\mathbb{P})$. Contiguity requires $\mathbb{P}L = 1$, a condition that begs for interpretation of L as the density of another probability measure $\mathbb{Q}$ with respect to $\mathbb{P}$. The limit assertion then becomes

<11> $$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} \text{ (under } \{\mathbb{P}_n\}\text{)} \rightsquigarrow \frac{d\mathbb{Q}}{d\mathbb{P}} \text{ (under } \mathbb{P}\text{)},$$

with $\mathbb{Q}$ a probability measure absolutely continuous with respect to $\mathbb{P}$.

Contiguity is also closely related to convergence in Le Cam's sense. In fact, under regularity assumptions ensuring existence of conditional distributions, Problem [6] shows that the convergence <11> implies existence of Markov kernels K_n for which $K_n\mathbb{P} = \mathbb{P}_n$ and $\|K_n\mathbb{Q} - \mathbb{Q}_n\|_1 \to 0$. In fact, it can easily be shown (Le Cam & Yang 2000, Section 3.1) that contiguity is equivalent to absolute continuity of $\mathbb{Q}$ with respect to $\mathbb{P}$, for every $(\mathbb{P},\mathbb{Q})$ that is a limit of a subsequence of $(\mathbb{P}_n,\mathbb{Q}_n)$ in Le Cam's sense.

<12> **Example.** Once again consider the $\mathbb{P}_n$ and $\mathbb{Q}_n$ from Example <3>, with likelihood ratio $L_n = \exp\big(\delta_n(x-\alpha_n) - \delta_n^2/2\big)$, where $\delta_n := \beta_n - \alpha_n$. The difference $x - \alpha_n$ has a $N(0,1)$ distribution, and thus $\log L_n$ is distributed as $N(-\delta_n^2/2, \delta_n^2)$, under $\mathbb{P}_n$. For L_n to converge in $\mathbb{P}_n$-distribution we must have $\delta_n^2 \to \delta^2 < \infty$ (compare with Problem [4]). The limit distribution is that of $L := \exp(\delta x - \delta^2/2)$ under the $N(0,1)$ distribution $\mathbb{P}$ on the real line. By direct calculation, $\mathbb{P}L = 1$. (Compare with the moment generating function of the normal distribution.) The corresponding $\mathbb{Q}$ is the $N(\delta, 1)$ distribution. □

The form of the limit distribution in the previous Example is not coincidental.

<13> **Example.** In many classical situations, $\log L_n$ has a limiting normal distribution, or, more precisely, $L_n \rightsquigarrow \exp(X)$, with X defined on some $(\Omega, \mathcal{A}, \mathbb{P})$, with distribution $N(\mu, \sigma^2)$. For contiguity we must have $1 = \mathbb{P}\exp(X) = \exp(\mu + \frac{1}{2}\sigma^2)$. That is, $\mu = -\frac{1}{2}\sigma^2$ is equivalent to contiguity in this setting. □

2. Contiguity for product measures

For the study of asymptotic behavior under sequences of alternatives, we often need to consider sequences of probability measures $\mathbb{Q}_n := P_{\theta_n}^n$ and $\mathbb{P}_n := P_{\theta_0}^n$, where θ_n is a sequence converging to θ_0 at a $1/\sqrt{n}$ rate. For simplicity suppose θ is a real parameter, and P_θ has a smooth density f_θ with respect to a dominating measure λ.

Classical approximation arguments can be used to establish contiguity, $\mathbb{Q}_n \lhd \mathbb{P}_n$, when the density is twice continuously differentiable. The arguments become a little subtle when the densities do not all have the same support. The difficulties are avoided when $\{f_\theta > 0\}$ does not change with θ. For this case, by restricting λ to the common support set, we may even suppose $f_\theta(x) > 0$ for all θ and x, which ensures that there are no $\log 0$ problems when defining $\ell_\theta(x) := \log f_\theta(x)$.

<14> **Theorem.** *Suppose the density f_θ is everywhere strictly positive, and that $\ell_\theta(x)$ is twice differentiable in some neighborhood U of θ_0, with*

(i) $J_0 := P_{\theta_0}\dot\ell_{\theta_0}^2 < \infty$

(ii) $\theta \mapsto \ddot\ell_\theta(x)$ is continuous at θ_0

(iii) there exists a P_0-integrable function $M(x)$ for which $\sup_{\theta\in U}|\ddot\ell_\theta(x)| \le M(x)$.

Then $P_{\theta_0}^x\dot\ell(x) = 0$, and

$$Z_n := \sum\nolimits_{i\le n}\dot\ell_{\theta_0}(x_i)/\sqrt{n} \rightsquigarrow N(0, J_0) \qquad \text{under } \mathbb{P}_n.$$

If $\theta_n = \theta_0 + \delta_n/\sqrt{n}$, with $\{\delta_n\}$ bounded, then

$$\log\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} = \delta_n Z_n - \tfrac{1}{2}\delta_n^2 J_1 + o_p(1; \mathbb{P}_n),$$

where $J_1 := -P_{\theta_0}^x\ddot\ell_{\theta_0}(x)$. If $J_1 = J_0$ then $\mathbb{Q}_n \lhd \mathbb{P}_n$.

Proof. For simplicity of notation, suppose $\theta_0 = 0$. We may also suppose that $\theta_n \in U$. By Taylor's theorem,

<15> $$\ell_\theta(x) = \ell_0(x) + \theta\dot\ell_0(x) + \tfrac{1}{2}\theta^2\ddot\ell_0(x) + \tfrac{1}{2}\theta^2 r(x,\theta),$$

where, for some t (depending on x and θ) with $|t| \le |\theta|$,

$$2M(x) \ge |\ddot{\ell}_t(x) - \ddot{\ell}_0(x)| = |r(x,\theta)| \to 0 \qquad \text{as } \theta \to 0.$$

By Dominated Convergence, $P_0|r(x,\theta)| \to 0$ as $\theta \to 0$, and hence

<16> $$n^{-1}\sum\nolimits_{i\le n} r_\theta(x_i) = o_p(1 : \mathbb{P}_n) \qquad \text{as } \theta \to 0.$$

Also, by integrating both sides of <15> we get

$$-D(P_0\|P_\theta) = P_0^x\left(\ell_\theta(x) - \ell_0(x)\right) = -\theta P_0^x \dot{\ell}_0(x) - \tfrac{1}{2}\theta^2 P_0^x \ddot{\ell}_0(x) + o(\theta^2).$$

For $D(P_0\|P_\theta)$ to achieves its minimum of zero at $\theta = 0$ we must have the coefficient $P_0^x \dot{\ell}_0(x)$ of the linear term equal to zero.

REMARK. The argument about the linear term at the minimum tacitly assumes that 0 is an interior point of the parameter set.

The logarithm of the likelihhod ratio $L_n := d\mathbb{Q}_n/d\mathbb{P}_n$ equals

$$\begin{aligned}\sum\nolimits_{i\le n}\left(\ell_{\theta_n}(x_i) - \ell_0(x_i)\right) &= \theta_n \sum\nolimits_{i\le n}\dot{\ell}_0(x_i) + \tfrac{1}{2}\theta_n^2 \sum\nolimits_{i\le n}\left(\ddot{\ell}_0(x_i) + r(x_i,\theta_n)\right) \\ &= \delta_n Z_n + \tfrac{1}{2}\delta_n^2\left(n^{-1}\sum\nolimits_{i\le n}\ddot{\ell}_0(x_i) + n^{-1}\sum\nolimits_{i\le n} r(x_i,\theta_n)\right).\end{aligned}$$

The Law of Large Numbers and <20> let us replace the coefficient of δ_n^2 by $P_{\theta_0}^x \ddot{\ell}_{\theta_0}(x) + o_p(1;\mathbb{P}_n)$.

For contiguity, according to Lemma <9> we need to prove that if $L_n \rightsquigarrow L$ along a subsequence then $\mathbb{P}L = 1$. By a further subsequencing we may also assume that $\delta_n \to \delta$, a finite limit. Along the sub-subsequence we then have

$$\log L_n \rightsquigarrow \delta N(0, J_0) - \tfrac{1}{2}\delta^2 J_1.$$

□ Example <13> then shows why $J_1 = J_0$ is equivalent to contiguity.

The equality $J_1 = P_{\theta_0}^x \ddot{\ell}_{\theta_0}(x) = -\text{var}_{\theta_0}\left(\dot{\ell}_{\theta_0}(x)\right) = J_0$ is the classical dual representation for the information function $\mathbb{I}_{\theta_0}$ at θ_0. As Le Cam & Yang (2000, page 41) commented,

> The equality ... is the classical one. One finds it for instance in the standard treatment of maximum likelihood estimation under Cramér's conditions. There it is derived from conditions of differentiability under the integral sign.

The classical equality is nothing more than contiguity in disguise.

The statement of the Theorem left unresolved the conditions on the densities under which we must have $-P_{\theta_0}^x \ddot{\ell}_{\theta_0}(x) = \mathbb{I}_{\theta_0}$. The usual argument starts from the identity $\lambda^x\left(f_\theta \dot{\ell}_\theta(x)\right) = 0$, then justifies differentiation under the integral by a domination condition, to deduce that $\lambda\left(\dot{f}_\theta(x)\dot{\ell}_\theta(x) + f_\theta(x)\ddot{\ell}_\theta(x)\right) = 0$. Many authors just assume, even more directly, that differentiation under the integral is justified, without imposing explicit conditions. There are more elegant, indirect, ways to derive the identity. The next Theorem will provide an example.

The analysis becomes more complicated if the sets $\{f_\theta > 0\}$ are not all the same. We then need to impose a condition regarding the mass of the part of P_θ that is singular with respect to P_{θ_0}.

For simplicity of notation, again suppose $\theta_0 = 0$. Write N_0 for the set $\{x : f_0(x) = 0\}$, and $\alpha(\theta)$ for $P_\theta N_0$, the total mass of the part of P_θ that is not absolutely continuous with respect to P_0. The $\mathcal{F}_n$-measurable set $F_n := \bigcup_{i\le n}\{x_i \in N_0\}$ has zero P_0^n probability, but

$$P_\theta^n F_0^c = \prod\nolimits_{i\le n} P_\theta N_0^c = \left(1-\alpha(\theta)\right)^n .$$

If $\alpha(\theta)$ were not of order $o(\theta^2)$ we could find a sequence $\{\theta_n\}$ of order $O(n^{-1/2})$ and an $\epsilon > 0$ for which $\alpha(\theta_n) \ge \epsilon/n$ infinitely often. We would then have a sequence for which $\liminf_n P_{\theta_n}^n F_n \ge 1 - e^{-\epsilon} > 0$ but $P_0^n F_n \equiv 0$, ruling out contiguity. Thus a necessary condition for contiguity, $P_{\theta_n}^n \lhd P_0^n$ whenever $\theta_n = O(n^{-1/2})$ is

<17> $$P_\theta\{x : f_0(x) = 0\} = o(\theta^2) \qquad \text{as } \theta \to 0.$$

Assumption <17> takes care of one difficulty in the the case when the sets $\{f_\theta > 0\}$ are not the same as $\{f_0 > 0\}$. Another, more subtle, problem arises with the defintion of $\log f_\theta$. If $f_0(x) > 0$ then, by continuity, we know that $f_\theta(x) > 0$ for $|\theta| \le \delta(x)$. There might be no fixed δ, not depending on x, for which $f_\theta(x) > 0$ when $|\theta| \le \delta$. We might have $P_0 \log f_\theta(x) = -\infty$ for all $\theta \ne 0$, which would cast doubt on some of the calculations used to prove Theorem <14>. For example, how could assumption (iii) hold? The function $\ell_\theta(x) := \log f_\theta(\theta)$ might only be defined on an interval of θ values that depend on x. It still makes sense to work with the pointwise derivative $\dot\ell_0(x)$, but we might encounter the value $-\infty$ with positive P_0 probability when studying $\ell_\theta(x)$ for a fixed $\theta \ne 0$. It appears that we have to impose the regularity conditions directly on $f_\theta(x)$, and not on $\log f_\theta(x)$.

<18> **Theorem.** *Suppose the map $\theta \mapsto f_\theta$ is twice differentiable in a neighborhood U of 0 with:*

(i) $\theta \mapsto \ddot f_\theta(x)$ *is continuous at* 0;

(ii) there exists a measurable function $M(x)$ *with* $P_0^x\left(M(x)/f_0(x)\right) < \infty$ *for which* $\sup_{\theta\in U} |\ddot f_\theta(x)| \le M(x)$;

(iii) $P_0^x\left(\dot f_\theta(x)/f_0(x)\right)^2 \to P_0^x\left(\dot f_0(x)/f_0(x)\right)^2 < \infty$ *as* $\theta \to 0$;

(iv) $P_\theta\{f_0 = 0\} = o(\theta^2)$ *as* $\theta \to 0$.

Then $P_0\dot\ell_0(x) = 0 = P_0\left(\ddot f(x)/f_0(x)\right)$.

Define $\mathbb{P}_n := P_0^n$ *and* $\mathbb{Q}_n := P_{\theta_n}^n$. *If* $\theta_n := \theta_0 + \delta_n/\sqrt{n}$, *with* $\{\delta_n\}$ *bounded, then*

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} = \left(1 + o_p(1;\mathbb{P}_n)\right)\exp\left(\delta_n Z_n - \tfrac{1}{2}\delta_n^2\mathbb{I}_0\right),$$

where $\mathbb{I}_0 := \mathrm{var}_0(\dot\ell_0)$ *and* $Z_n := \sum_{i\le n}\dot\ell_0(x_i)/\sqrt{n} \rightsquigarrow N(0,\mathbb{I}_0)$ *under* $\mathbb{P}_n$. *Consequently,* $\mathbb{Q}_n \lhd \mathbb{P}_n$.

Proof. There are several useful ways to write the Taylor expansion of f_θ around 0. When $\theta \in U$,

<19> $$f_\theta(x) = f_0(x) + \theta\dot f_0(x) + \tfrac{1}{2}\theta^2\ddot f_0(x) + \tfrac{1}{2}\theta^2 r(x,\theta),$$

where, for some t (depending on x and θ) with $|t| \le |\theta|$,

$$2M(x) \ge |\ddot f_t(x) - \ddot f_0(x)| = |r(x,\theta)| \to 0 \qquad \text{as } \theta \to 0.$$

Dominated Convergence and (ii) then gives

<20> $$P_0|r(x,\theta)/f_0(x))| \to 0 \qquad \text{as } \theta \to 0.$$

From <19> we also have

$$\{f_0(x) > 0\}\frac{f_\theta(x) - f_0(x) - \theta \dot f_0(x)}{\theta^2 f_0(x)} \to \frac{\ddot f_0(x)}{f_0(x)}\{f_0(x) > 0\} \qquad \text{as } \theta \to 0.$$

Moreover, the ratio is bounded in absolute value by the P_0-integrable function $\{f_0(x) > 0\}M(x)/f_0(x)$. By Dominated Convergence for P_0, followed by a cancellation of the $f_0(x)$ factor in the first term, we have

$$\theta^{-2}\left(P_\theta\{f_0 > 0\} - P_0 1 - \theta P_0 \dot\ell_0\right) \to P_0^x\left(\ddot f_0/f_0\right).$$

Assumption (iv) simplifies the assertion to $o(1) - \theta^{-1}P_0\dot\ell_0 \to P_0^x\left(\ddot f_0/f_0\right)$, from which it follows that $P_0\dot\ell_0 = 0$ (because $P_0^x|\ddot f_0/f_0| \le P_0(M/f_0) < \infty$), and hence $P_0^x\left(\ddot f_0(x)/f_0(x)\right) = 0$.

It will also be helpful to have the Taylor expansion with the remainder written in the Lagrange style,

<21> $$f_\theta(x) = f_0(x) + \theta\int_0^1 \dot f_{\theta t}(x)\,dt,$$

a form that will be useful because it does not involve the second derivative.

The likehood ratio $L_n := d\mathbb{Q}_n/d\mathbb{P}_n$ can be written as

$$\prod_{i\le n}\frac{f_{\theta_n}(x_i)}{f_0(x_i)} = \prod_{i\le n}\left(1+\epsilon_{n,i}\right) \qquad \text{where } \epsilon_{n,i} := \{f_0(x_i) > 0\}\frac{f_{\theta_n}(x_i) - f_0(x_i)}{f_0(x_i)}.$$

The indicator functions are not really need if we consider only behavior under $\mathbb{P}_n$, but they will prevent inadvertent appeals to $0/0 \overset{?}{=} 1$. Until further notice, all calculuations are carried out under $\mathbb{P}_n$, so I will temporarily dispense with the indicators, and write $o_p(\cdot)$ instead of $o_p(\cdot\,;\mathbb{P}_n)$.

By <19>,

<22> $$\epsilon_{n,i} = \theta_n\dot\ell_0(x_i) + \tfrac12\theta_n^2\left(\ddot f_0(x_i) + r(x_i,\theta_n)\right)/f_0(x_i),$$

whence

$$|\epsilon_{n,i}| \le |\theta_n\dot\ell_0(x_i)| + \tfrac12\theta_n^2 Z_i \qquad \text{where } Y_i := \left(|\ddot f_0(x_i)| + 2M(x_i)\right)/f_0(x_i).$$

Under $\mathbb{P}_n$, the random variables $\dot\ell_0(x_i)$ are identically distributed, with finite second moments, and the random variables Y_i are identically distributed, with finite first moments. Problem [7] shows that

$$\max_{i\le n}|\dot\ell_0(x_i)| = o_p(n^{-1/2}) \qquad \text{and} \qquad \max_{i\le n}|Y_i| = o_p(n^{-1}).$$

from which it follows that

<23> $$\max_{i\le n}|\epsilon_{n,i}| = o_p(1) \qquad \text{when } \theta_n = \delta_n/\sqrt{n} = O(n^{-1/2}).$$

Expansion <22> also gives

$$\sum_{i\le n}\epsilon_{n,i} = \delta_n Z_n + \tfrac12\delta_n^2\left(n^{-1}\sum_{i\le n}\ddot f(x_i)/f_0(x_i) + n^{-1}\sum_{i\le n} r(x_i,\theta_n)/f_0(x_i)\right)$$

<24> $$= \delta_n Z_n + o_p(1),$$

with the Law of Large Numbers and the fact that $P_0\left(\ddot f(x)/f_0(x)\right) = 0$ disposing of the first average in parentheses, and <20> disposing of the second.

Assumption (iii) will lead to a neat asymptotic form for the sum of squares of the ϵ's. Define $W_\theta := \{f_0 > 0\}\dot{f}_\theta/f_0$. By Fatou's Lemma (along a sequence of θ values, if you prefer),

find the source of the Fatou trick; Blyth? I&H?

$$\begin{aligned}4P_0W_0^2 - \limsup_{\theta\to 0} P_0|W_\theta - W_0|^2 \\ &= \liminf_{\theta\to 0} P_0\left(2W_\theta^2 + 2W_0^2 - |W_\theta - W_0|^2\right) \\ &\geq P_0 \liminf_{\theta\to 0}\left(2W_\theta^2 + 2W_0^2 - |W_\theta - W_0|^2\right) = 4P_0W_0^2.\end{aligned}$$

That is,

<25> $$\gamma(\theta)^2 := P_0|W_\theta - W_0|^2 \to 0 \qquad \text{as } \theta \to 0, \text{ where } W_\theta := \{f_0 > 0\}\dot{f}_\theta/f_0.$$

From <21>, we also have the representation $\epsilon_{n,i} = \theta_n \int_0^1 W_{\theta_n t}(x_i)\,dt$. Hence

$$\begin{aligned}\mathbb{P}_n|\epsilon_{n,i}^2 - \theta_n^2 W_0(x_i)^2| &= \theta_n^2 P_0^x\left|\int_0^1\int_0^1 W_{\theta t}(x)W_{\theta s}(x) - W_0(x)^2\,ds\,dt\right| \\ &\leq \theta_n^2\int_0^1\int_0^1\left(C\gamma(\theta_n t) + C\gamma(\theta_n t) + \gamma(\theta_n t)\gamma(\theta_n s)\right)\,ds\,dt,\end{aligned}$$

where $C^2 := P_0W_0^2 = \mathbb{I}_0$. It follows that

$$\sum\nolimits_{i\leq n}\mathbb{P}_n|\epsilon_{n,i}^2 - \theta_n^2W_0(x_i)^2| \to 0,$$

implying

<26> $$\sum\nolimits_{i\leq n}\epsilon_{n,i}^2 = \delta_n^2 n^{-1}\sum\nolimits_{i\leq n}W_0(x_i)^2 + o_p(1) = \delta_n^2\mathbb{I}_0 + o_p(1).$$

The results <23>, <24>, and <26> lead rapidly to the desired approximation for L_n, via the inequality

$$|\log(1+t) - t + \tfrac{1}{2}t^2| \leq |t|^3 \qquad \text{for } |t| \leq 1/2..$$

When $\max_{i\leq n}|\epsilon_{n,i}| \leq 1/2$ we have

$$|\log(L_n) - \sum_{i\leq n}\epsilon_{n,i} + \tfrac{1}{2}\sum_{i\leq n}\epsilon_{n,i}^2| \leq \sum_{i\leq n}|\epsilon_{n,i}|^3 \leq \max_{i\leq n}|\epsilon_{n,i}|\sum_{i\leq n}\epsilon_{n,i}^2 = o_p(1),$$

that is,

$$L_n\{\max_{i\leq n}|\epsilon_{n,i}| \leq 1/2\} = \{\max_{i\leq n}|\epsilon_{n,i}| \leq 1/2\}\exp\left(\delta_n Z_n - \tfrac{1}{2}\delta_n^2\mathbb{I}_0 + o_p(1)\right).$$

The $1+o_p(1)$ factor in the statement of the Theorem absorbs the $o_p(1)$ in the exponent, as well as allowing for arbitrarily bad behavior of L_n when $\max_{i\leq n}|\epsilon_{n,i}| > 1/2$.

□ Example <13> gives contiguity.

3. Limit distributions under contiguous alternatives

Contiguity was introduced in Section 1 as a way to transfer either $o_p(\cdot)$ or $O_p(\cdot)$ assertions from $\{\mathbb{P}_n\}$ to $\{\mathbb{Q}_n\}$. It can also be used to transfer assertions of convergence in distribution for sequences of random vectors $\{Y_n\}$, if we control the joint behaviour of Y_n and the likelihood ratio. The idea behind the proof is straightforward if we ignore complications such as unbounded likelihoods: for bounded, uniformly continuous g,

$$\mathbb{Q}_n g(Y_n) \stackrel{?}{=} \mathbb{P}_n L_n g(Y_n) \xrightarrow{?} \mathbb{P}Lg(Y).$$

In a rigorous proof, contiguity controls the contributions from regions of large L_n, and from the singularity region $\mathcal{N}_n$, and then convergence in distribution of (L_n, Y_n) takes care of the convergence assertion. The limit expression becomes $\mathbb{Q}g(Y)$, where $\mathbb{Q}$ is the probability measure defined to have density L with respect to $\mathbb{P}$. That is, the limit distribution of Y_n under $\mathbb{Q}_n$ is given by Y, as a random vector on $(\Omega, \mathcal{A}, \mathbb{Q})$.

We will need the result only for random vectors Y_n, but the proof actually works for random elements more general spaces.

<27> **Lemma.** *Suppose (Y_n, L_n) converges in distribution under $\{\mathbb{P}_n\}$ to a limit represented by a pair (Y, L) on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with $\mathbb{P}L = 1$. Then $\{Y_n\}$ converges in distribution under $\{\mathbb{Q}_n\}$ to the limit represented by Y as a random element on the probability space $(\Omega, \mathcal{A}, \mathbb{Q})$, where $\mathbb{Q}$ has density L with respect to $\mathbb{P}$. That is, $\mathbb{Q}_n g(Y_n) \to \mathbb{Q}g(Y) := \mathbb{P}Lg(Y)$, at least for bounded, continuous g.*

Proof. The condition $\mathbb{P}L = 1$ ensures that $\mathbb{Q}_n \lhd \mathbb{P}_n$. Fix $\epsilon > 0$ and let g be a bounded, continuous function. For convenience suppose $0 \le g \le 1$. Invoke contiguity to find a finite M such that $\mathbb{P}\{L = M\} = 0$ and $\mathbb{Q}\{L > M\} < \epsilon$ and $\mathbb{Q}_n\{L_n > M\} < \epsilon$ eventually. Then from the definition of L_n,

$$|\mathbb{Q}_n g(Y_n) - \mathbb{P}_n L_n g(Y_n)\{L_n \le M\}| \le \mathbb{Q}_n\{L_n > M\} + \mathbb{Q}\mathcal{N}_n < 2\epsilon \qquad \text{eventually.}$$

By the Continuous Mapping Theorem,

$$\mathbb{P}_n L_n g(Y_n)\{L_n \le M\} \to \mathbb{P}Lg(Y)\{L \le M\},$$

□ which differs from $\mathbb{P}Lg(Y) = \mathbb{Q}g(Y)$ by at most ϵ.

REMARK. By the same argument (or just by substitution of (Y_n, L_n) for Y_n in the conclusion of the Lemma), the pair (Y, L) under $\mathbb{Q}$ also represents the limit distribution for the pairs (Y_n, L_n) under $\{\mathbb{Q}_n\}$.

Convergence in distribution of (Y_n, L_n) is equivalent to convergence in distribution of $(Y_n, \log L_n)$. When the joint limit is normal, the assertion of the preceding Lemma takes a particularly simple form. The result is known as ***Le Cam's Third Lemma***.

<28> **Example.** Suppose $(Y_n, L_n) \rightsquigarrow (Y, e^Z)$ under $\{\mathbb{P}_n\}$, where the pair (Y, Z), defined on $(\Omega, \mathcal{A}, \mathbb{P})$, has a joint normal distribution. To ensure contiguity, the marginal Z distribution must be $N(-1/2\sigma^2, \sigma^2)$ for some $\sigma^2 > 0$. Let the marginal Y distribution be $N(\mu, V)$, and let γ denote the vector of covariances between Y and Z. Under $\mathbb{P}$ the pair (Y, Z) has moment generating function

$$M(s,t) := \mathbb{P}\exp(s'Y + tZ) = \exp\left(s'\mu + \tfrac{1}{2}s'Vs + s'\gamma t - \tfrac{1}{2}\sigma^2 t + \tfrac{1}{2}\sigma^2 t^2\right).$$

The limiting distribution under $\{\mathbb{Q}_n\}$ has moment generating function

$$\begin{aligned}\mathbb{Q}\exp(s'Y + tZ) &= \mathbb{P}\exp(Z)\exp(s'Y + tZ)\\ &= M(s, t+1)\\ &= \exp\left(s'(\mu+\gamma) + \tfrac{1}{2}s'Vs + s'\gamma t + \tfrac{1}{2}\sigma^2 t + \tfrac{1}{2}\sigma^2 t^2\right).\end{aligned}$$

That is, the variances and covariances stay the same, but the mean of Y is shifted to $\mu + \gamma$. □

<29> **Example.** In Chapter 1, a heuristic argument gave the asymptotic behavior of the estimator $\widehat{\theta}_n$ defined to minimize $\sum_{i\le n} g(x_i, \theta)$. Assuming $\theta = \text{argmin}_t \, P_\theta g(x,t)$ for each θ, I argued that $\widehat{\theta}_n$ should converge in P_θ^n probability to θ, and also

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) = n^{-1/2} m_\theta(x_i) + o_p(1),$$

where $m_\theta(x) = -\dot{g}(x,\theta)/J_g(\theta)$, with $J_g(\theta) := P_\theta \ddot{g}(x,\theta)$, and the $o_p(1)$ is an abbreviation for $o_p(1; P_\theta^n)$.

For a fixed θ and δ, let $\theta_n := \theta + \delta/\sqrt{n}$, and $\mathbb{P}_n := P_\theta^n$, and $\mathbb{Q}_n := P_{\theta_n}^n$. Assume that the conditions of Theorem <18> are satisfies, so that

$$L_n = \left(1 + o_p(1;\mathbb{P}_n)\right)\exp(\delta Z_n - \tfrac{1}{2}\delta^2 \mathbb{I}_\theta),$$

with

$$Z_n = n^{-1/2}\sum\nolimits_{i\le n} \dot{\ell}_\theta(x_i) \rightsquigarrow N(0, \mathbb{I}_\theta) \qquad \text{under } \mathbb{P}_n.$$

Write Y_n for $\sqrt{n}\left(\widehat{\theta}_n - \theta\right)$. Under $\mathbb{P}_n$, the pair (Y_n, Z_n) is approximated by a standardized sum of random vectors,

$$(Y_n, Z_n) = o_p(1) + n^{-1/2}\sum\nolimits_{i\le n}\left(m(x_i), \dot{\ell}_\theta(x_i)\right),$$

which has a limiting bivariate normal distribution (Y, Z) with Z distributed $N(-\delta^2\mathbb{I}_\theta/2, \delta^2\mathbb{I}_\theta)$, and Y distributed $N(0, v_\theta)$ for $v_\theta := P_\theta \dot{g}(x,\theta)^2/J_g(\theta)^2$, and $\text{cov}(Y,Z) = \gamma_\theta := -\delta P_\theta\left(\dot{g}(x,\theta)\dot{\ell}_\theta(x)\right)/J_g(\theta)$.

Under $\mathbb{Q}_n$ the Y_n has a $N(\gamma_\theta, v_\theta)$ limit distribution, by Example <28>. Thus

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_n\right) = Y_n - \delta \rightsquigarrow N(\gamma_\theta - \delta, v_\theta) \qquad \text{under } \mathbb{Q}_n.$$

The limit distribution for $\sqrt{n}\left(\widehat{\theta}_n - \theta_n\right)$ is the same under $\mathbb{Q}_n$ as under $\mathbb{P}_n$ if $\gamma_\theta = \delta$, that is, if $J_g(\theta) = -P_\theta\left(\dot{g}(x,\theta)\dot{\ell}_\theta(x)\right)$. This equality is precisely the condition derived in Chapter 1 from the assumption that $P_\theta g(x,t)$ is minimized at $t = \theta$.

Thus, insofar as the heuristics can be believed, we have the limiting distribution of $\sqrt{n}\left(\widehat{\theta}_n - \theta_n\right)$ under $P_{\theta_n}^n$ the same as the limiting distribution of $\sqrt{n}\left(\widehat{\theta}_n - \theta\right)$ under P_θ^n. Estimators with this property are usually said to be ***Hájek regular***, a property that we will later meet as one of the assumptions for the Hájek-Le Cam Convolution Theorem. □

4. Problems

[1] Suppose $\{\mathbb{P}_n\}$ and $\{\mathbb{Q}_n\}$ are sequences of probability measures with the following property: for each subsequence $\mathbb{N}_1 \subseteq \mathbb{N}$ there exists a subsubsequence $\mathbb{N}_2 \subseteq \mathbb{N}_1$ for which $\{\mathbb{Q}_n : n \in \mathbb{N}_2\} \lhd \{\mathbb{P}_n : n \in \mathbb{N}_2\}$. Show that $\{\mathbb{Q}_n : n \in \mathbb{N}\} \lhd \{\mathbb{P}_n : n \in \mathbb{N}\}$. Hint: If contiguity fails, there is subsequence for which there are sets with $\mathbb{P}_n F_n \to 0$ but $\mathbb{Q}_n F_n > \epsilon$, for some $\epsilon > 0$.

[2] Suppose $\{X_n\}$ is a sequence of random variables with the following property: for each subsequence $\mathbb{N}_1 \subseteq \mathbb{N}$ there exists a subsubsequence $\mathbb{N}_2 \subseteq \mathbb{N}_1$ for which $\{X_n : n \in \mathbb{N}_2\} = O_p(1)$. Show that $\{X_n : n \in \mathbb{N}\} = O_p(1)$.

[3] If $\{X_n\} = o_p(1)$, show that there exists a sequence $\{\epsilon_n\}$ that converges to zero slowly enough to ensure $\mathbb{P}\{|X_n| > \epsilon_n\} \to 0$. Hint: Build ϵ_n using an increasing sequence $n(k)$ such that $\mathbb{P}\{|X_n| > 1/k\} < 1/k$ for $n \geq n(k)$.

[4] Suppose $Z_n \rightsquigarrow N(0, I_k)$ and that $\alpha_n Z_n + \beta_n$ has a nondegenerate limit distribution, for a pair of deterministic sequences $\{\alpha_n\}$ and $\{\beta_n\}$. Show that both $|\alpha_n|$ and β_n must converge to finite limits.

[5] Let $\mathbb{P}_n$ denote the $N(\alpha_n, 1)$ distribution and $\mathbb{Q}_n$ denote the $N(\beta_n, 1)$ distribution, both on the real line. Under what conditions on the sequences of constants $\{\alpha_n\}$ and $\{\beta_n\}$ do we have $\{\mathbb{Q}_n\} \lhd \{\mathbb{P}_n\}$?

[6] Suppose $\mathbb{P}_n$ and $\mathbb{Q}_n$ are probability measures on $(\Omega_n, \mathcal{F}_n)$, for $n \in \overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$, with $\mathbb{Q}_n \ll \mathbb{P}_n$. Write L_n for the corresponding densities. Define $F_n(x) := \mathbb{P}_n\{L_n \leq x\}$, and write F_n^{-1} for the corresponding quantile function. Suppose the conditional distributions $\tau_{n,t}(\cdot) := \mathbb{P}_n\,(\cdot \mid L_n = t)$ exists, as Markov kernels from $\mathbb{R}$ to Ω_n. For each $n \in \mathbb{N}$ define a Markov kernel $K_{n,\omega}$ from Ω_∞ to Ω_n, as follows.

Given $\omega_\infty \in \Omega_\infty$, define $T_\infty := L_\infty(\omega_\infty)$; then generate $U \in (0,1)$ with $U \mid T_\infty = t \sim \text{Unif}[F_\infty(t-), F_\infty(t)]$; then define $T_n = F_n^{-1}(U)$; then generate $\omega_n \mid T_n = t \sim \tau_{n,t}$.

(i) Show that $K_n \mathbb{P}_\infty = \mathbb{P}_n$. That is, the probability measure $\mathbb{M}_n := \mathbb{P}_\infty \otimes K_n$ on $\Omega_\infty \times \Omega_n$ has marginals $\mathbb{P}_\infty$ and $\mathbb{P}_n$.

(ii) For each measurable function with $|f| \leq 1$ on Ω_n, show that

$$\begin{aligned} |K_n \mathbb{Q}_\infty f - \mathbb{Q}_n f| &= |\mathbb{P}_\infty^\omega \left(L_\infty(\omega) K_{n,\omega}^x f(x)\right) - \mathbb{P}_n^x \left(L_n(x) f(x)\right)| \\ &\leq \mathbb{M}^{\omega,x} |L_\infty(\omega) - L_n(x)| \\ &= \int_0^1 |F_\infty^{-1}(u) - F_n^{-1}(u)|\, du. \end{aligned}$$

(iii) Deduce that if$L_n(\mathbb{P}_n) \rightsquigarrow L_\infty(\mathbb{P}_\infty)$ then $\|K_n \mathbb{Q}_\infty - \mathbb{Q}_n\|_1 \to 0$.

(iv) Extend the result (iii) to the case where $\mathbb{Q}_n \lhd \mathbb{P}_n$, for $n \in \mathbb{N}$, with $\mathbb{Q}_\infty \ll \mathbb{P}_\infty$.

[7] Let $Z_1, Z_2, \ldots$ be a sequence of independent, identically distributed random variables with $\mathbb{P}|Z_i|^r < \infty$ for a constant $r \geq 1$. Prove that $\max_{i \leq n} |Z_i| = o_p(n^{1/r})$. Hint: Show that $\mathbb{P}\{\max_{i \leq n} |Z_i| > \epsilon n^{1/r}\}$ is smaller than $\epsilon^{-r} \mathbb{P}|Z_1|^r \{|Z_1| > \epsilon n^{1/r}\}$, then invoke Dominated Convergence.

5. Notes

Le Cam (1960) defined contiguity and derived its most important properties, in a few pages. The name “Le Cam’s Third Lemma” seems due to Hájek & Šidák (1967, Chapter VI). It was the third of the lemmas in their chapter describing contiguity. The numbering now should have little significance.

Lucien Le Cam himself felt that describing contiguity as a subtle invention was an exaggeration. In a private letter to me he wrote “Really, contiguity is a very trivial affair. I just gave it a name that pleased people.” Maybe the only subtlety lies in the recognition that something so trivial is worth noticing. To my chagrin, I ignored the concept for many years, because it seemed hardly worth bothering about. Moreover, I have found that I was not alone in my oversight. Maybe subtlety lies in the eye of the beholder.

References

Hájek, J. & Šidák, Z. (1967), *Theory of Rank Tests*, Academic Press. Also published by Academia, the Publishing House of the Czechoslavak Academy of Sciences, Prague.

Le Cam, L. (1960), 'Locally asymptotically normal families of distributions', *University of California Publications in Statistics* **3**, 37–98.

Le Cam, L. & Yang, G. L. (2000), *Asymptotics in Statistics: Some Basic Concepts*, 2nd edn, Springer-Verlag.