6

Contiguity

"Contiguity" is another name for "asymptotic absolute continuity." Contiguity arguments are a technique to obtain the limit distribution of a sequence of statistics under underlying laws Q_n from a limiting distribution under laws P_n. Typically, the laws P_n describe a null distribution under investigation, and the laws Q_n correspond to an alternative hypothesis.

6.1 Likelihood Ratios

Let P and Q be measures on a measurable space $(\Omega, \mathcal{A})$. Then Q is *absolutely continuous* with respect to P if $P(A) = 0$ implies $Q(A) = 0$ for every measurable set A; this is denoted by $Q \ll P$. Furthermore, P and Q are *orthogonal* if Ω can be partitioned as $\Omega = \Omega_P \cup \Omega_Q$ with $\Omega_P \cap \Omega_Q = \emptyset$ and $P(\Omega_Q) = 0 = Q(\Omega_P)$. Thus P "charges" only Ω_P and Q "lives on" the set Ω_Q, which is disjoint with the "support" of P. Orthogonality is denoted by $P \perp Q$.

In general, two measures P and Q need be neither absolutely continuous nor orthogonal. The relationship between their supports can best be described in terms of densities. Suppose P and Q possess densities p and q with respect to a measure μ, and consider the sets

$$\Omega_P = \{p > 0\}, \qquad \Omega_Q = \{q > 0\}.$$

See Figure 6.1. Because $P(\Omega_P^c) = \int_{p=0} p \, d\mu = 0$, the measure P is supported on the set Ω_P. Similarly, Q is supported on Ω_Q. The intersection $\Omega_P \cap \Omega_Q$ receives positive measure from both P and Q provided its measure under μ is positive. The measure Q can be written as the sum $Q = Q^a + Q^\perp$ of the measures

$$Q^a(A) = Q\bigl(A \cap \{p > 0\}\bigr); \qquad Q^\perp(A) = Q\bigl(A \cap \{p = 0\}\bigr). \tag{6.1}$$

As proved in the next lemma, $Q^a \ll P$ and $Q^\perp \perp P$. Furthermore, for every measurable set A

$$Q^a(A) = \int_A \frac{q}{p} \, dP.$$

The decomposition $Q = Q^a + Q^\perp$ is called the *Lebesgue decomposition* of Q with respect to P. The measures Q^a and $Q^\perp$ are called the *absolutely continuous part* and the *orthogonal*

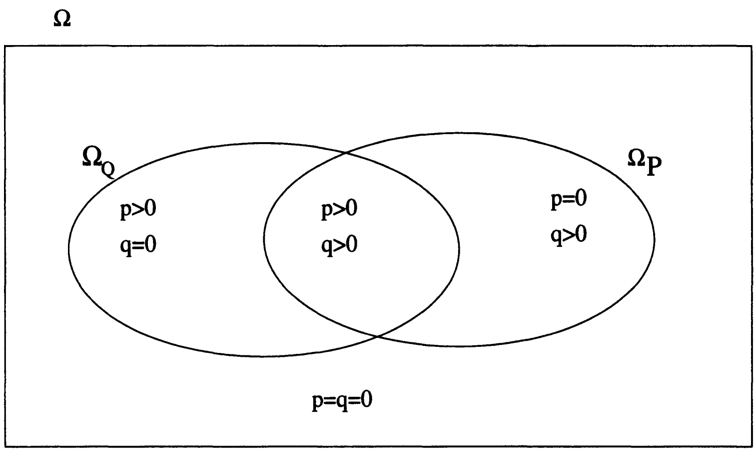

Figure 6.1. Supports of measures.

part (or *singular part*) of Q with respect to P, respectively. In view of the preceding display, the function q/p is a density of Q^a with respect to P. It is denoted dQ/dP (not: dQ^a/dP), so that

$$\frac{dQ}{dP} = \frac{q}{p}, \qquad P - \text{a.s.}$$

As long as we are only interested in the properties of the quotient q/p under P-probability, we may leave the quotient undefined for $p = 0$. The density dQ/dP is only P-almost surely unique by definition. Even though we have used densities to define them, dQ/dP and the Lebesgue decomposition are actually independent of the choice of densities and dominating measure.

In statistics a more common name for a Radon-Nikodym density is *likelihood ratio*. We shall think of it as a random variable $dQ/dP : \Omega \mapsto [0, \infty)$ and shall study its law under P.

6.2 ***Lemma.*** *Let P and Q be probability measures with densities p and q with respect to a measure μ. Then for the measures Q^a and $Q^\perp$ defined in (22.30)*

(i) $Q = Q^a + Q^\perp$, $Q^a \ll P$, $Q^\perp \perp P$.

(ii) $Q^a(A) = \int_A (q/p)\, dP$ for every measurable set A.

(iii) $Q \ll P$ if and only if $Q(p = 0) = 0$ if and only if $\int (q/p)\, dP = 1$.

Proof. The first statement of (i) is obvious from the definitions of Q^a and $Q^\perp$. For the second, we note that $P(A)$ can be zero only if $p(x) = 0$ for μ-almost all $x \in A$. In this case, $\mu\big(A \cap \{p > 0\}\big) = 0$, whence $Q^a(A) = Q\big(A \cap \{p > 0\}\big) = 0$ by the absolute continuity of Q with respect to μ. The third statement of (i) follows from $P(p = 0) = 0$ and $Q^\perp(p > 0) = Q(\emptyset) = 0$.

Statement (ii) follows from

$$Q^a(A) = \int_{A \cap \{p>0\}} q\, d\mu = \int_{A \cap \{p>0\}} \frac{q}{p} p\, d\mu = \int_A \frac{q}{p}\, dP.$$

For (iii) we note first that $Q \ll P$ if and only if $Q^\perp = 0$. By (22.30) the latter happens if and only if $Q(p = 0) = 0$. This yields the first "if and only if." For the second, we note that by (ii) the total mass of Q^a is equal to $Q^a(\Omega) = \int (q/p)\, dP$. This is 1 if and only if $Q^a = Q$. ■

It is not true in general that $\int f\, dQ = \int f\, (dQ/dP)\, dP$. For this to be true for every measurable function f, the measure Q must be absolutely continuous with respect to P. On the other hand, for any P and Q and nonnegative f,

$$\int f\, dQ \geq \int_{p>0} fq\, d\mu = \int_{p>0} f\frac{q}{p} p\, d\mu = \int f\frac{dQ}{dP}\, dP.$$

This inequality is used freely in the following. The inequality may be strict, because dividing by zero is not permitted.[†]

6.2 Contiguity

If a probability measure Q is absolutely continuous with respect to a probability measure P, then the Q-law of a random vector $X : \Omega \mapsto \mathbb{R}^k$ can be calculated from the P-law of the pair $(X, dQ/dP)$ through the formula

$$\mathrm{E}_Q f(X) = \mathrm{E}_P f(X)\frac{dQ}{dP}.$$

With $P^{X,V}$ equal to the law of the pair $(X, V) = (X, dQ/dP)$ under P, this relationship can also be expressed as

$$Q(X \in B) = \mathrm{E}_P 1_B(X)\frac{dQ}{dP} = \int_{B\times\mathbb{R}} v\, dP^{X,V}(x, v).$$

The validity of these formulas depends essentially on the absolute continuity of Q with respect to P, because a part of Q that is orthogonal with respect to P cannot be recovered from any P-law.

Consider an asymptotic version of the problem. Let $(\Omega_n, \mathcal{A}_n)$ be measurable spaces, each equipped with a pair of probability measures P_n and Q_n. Under what conditions can a Q_n-limit law of random vectors $X_n : \Omega_n \mapsto \mathbb{R}^k$ be obtained from suitable P_n-limit laws? In view of the above it is necessary that Q_n is "asymptotically absolutely continuous" with respect to P_n in a suitable sense. The right concept is contiguity.

6.3 Definition. The sequence Q_n is *contiguous* with respect to the sequence P_n if $P_n(A_n) \to 0$ implies $Q_n(A_n) \to 0$ for every sequence of measurable sets A_n. This is denoted $Q_n \triangleleft P_n$. The sequences P_n and Q_n are *mutually contiguous* if both $P_n \triangleleft Q_n$ and $Q_n \triangleleft P_n$. This is denoted $P_n \triangleleft\triangleright Q_n$.

The name "contiguous" is standard, but perhaps conveys a wrong image. "Contiguity" suggests sequences of probability measures living next to each other, but the correct image is "on top of each other" (in the limit).

[†] The algebraic identify $dQ = (dQ/dP)\, dP$ is false, because the notation dQ/dP is used as shorthand for dQ^a/dP: If we write dQ/dP, then we are not implicitly assuming that $Q \ll P$.

Before answering the question of interest, we give two characterizations of contiguity in terms of the asymptotic behavior of the likelihood ratios of P_n and Q_n. The likelihood ratios dQ_n/dP_n and dP_n/dQ_n are nonnegative and satisfy

$$\mathrm{E}_{P_n}\frac{dQ_n}{dP_n} \leq 1 \quad \text{and} \quad \mathrm{E}_{Q_n}\frac{dP_n}{dQ_n} \leq 1.$$

Thus, the sequences of likelihood ratios dQ_n/dP_n and dP_n/dQ_n are uniformly tight under P_n and Q_n, respectively. By Prohorov's theorem, every subsequence has a further weakly converging subsequence. The next lemma shows that the properties of the limit points determine contiguity. This can be understood in analogy with the nonasymptotic situation. For probability measures P and Q, the following three statements are equivalent by (iii) of Lemma 6.2:

$$Q \ll P, \qquad Q\left(\frac{dP}{dQ} = 0\right) = 0, \qquad \mathrm{E}_P\frac{dQ}{dP} = 1.$$

This equivalence persists if the three statements are replaced by their asymptotic counterparts: Sequences P_n and Q_n satisfy $Q_n \triangleleft P_n$, if and only if the weak limit points of dP_n/dQ_n under Q_n give mass 0 to 0, if and only if the weak limit points of dQ_n/dP_n under P_n have mean 1.

6.4 Lemma (Le Cam's first lemma). *Let P_n and Q_n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{A}_n)$. Then the following statements are equivalent:*

(i) $Q_n \triangleleft P_n$.
(ii) If $dP_n/dQ_n \overset{Q_n}{\rightsquigarrow} U$ *along a subsequence, then* $\mathrm{P}(U > 0) = 1$.
(iii) If $dQ_n/dP_n \overset{P_n}{\rightsquigarrow} V$ *along a subsequence, then* $\mathrm{E}V = 1$.
(iv) For any statistics $T_n : \Omega_n \mapsto \mathbb{R}^k$: *If* $T_n \overset{P_n}{\to} 0$, *then* $T_n \overset{Q_n}{\to} 0$.

Proof. The equivalence of (i) and (iv) follows directly from the definition of contiguity: Given statistics T_n, consider the sets $A_n = \{\|T_n\| > \varepsilon\}$; given sets A_n, consider the statistics $T_n = 1_{A_n}$.

(i) $\Rightarrow$ (ii). For simplicity of notation, we write just $\{n\}$ for the given subsequence along which $dP_n/dQ_n \overset{Q_n}{\rightsquigarrow} U$. For given n, we define the function $g_n(\varepsilon) = Q_n(dP_n/dQ_n < \varepsilon) - \mathrm{P}(U < \varepsilon)$. By the portmanteau lemma, $\liminf g_n(\varepsilon) \geq 0$ for every $\varepsilon > 0$. Then, for $\varepsilon_n \downarrow 0$ at a sufficiently slow rate, also $\liminf g_n(\varepsilon_n) \geq 0$. Thus,

$$\mathrm{P}(U = 0) = \lim \mathrm{P}(U < \varepsilon_n) \leq \liminf Q_n\left(\frac{dP_n}{dQ_n} < \varepsilon_n\right).$$

On the other hand,

$$P_n\left(\frac{dP_n}{dQ_n} \leq \varepsilon_n \wedge q_n > 0\right) = \int_{dP_n/dQ_n \leq \varepsilon_n} \frac{dP_n}{dQ_n}\, dQ_n \leq \int \varepsilon_n \, dQ_n \to 0.$$

If Q_n is contiguous with respect to P_n, then the Q_n-probability of the set on the left goes to zero also. But this is the probability on the right in the first display. Combination shows that $\mathrm{P}(U = 0) = 0$.

(iii) $\Rightarrow$ (i). If $P_n(A_n) \to 0$, then the sequence $1_{\Omega_n - A_n}$ converges to 1 in P_n-probability. By Prohorov's theorem, every subsequence of $\{n\}$ has a further subsequence along which

$(dQ_n/dP_n, 1_{\Omega_n - A_n}) \rightsquigarrow (V, 1)$ under P_n, for some weak limit V. The function $(v, t) \mapsto vt$ is continuous and nonnegative on the set $[0, \infty) \times \{0, 1\}$. By the portmanteau lemma

$$\liminf Q_n(\Omega_n - A_n) \geq \liminf \int 1_{\Omega_n - A_n} \frac{dQ_n}{dP_n} dP_n \geq \mathrm{E}1 \cdot V.$$

Under (iii) the right side equals $\mathrm{E}V = 1$. Then the left side is 1 as well and the sequence $Q_n(A_n) = 1 - Q_n(\Omega_n - A_n)$ converges to zero.

(ii) $\Rightarrow$ (iii). The probability measures $\mu_n = \frac{1}{2}(P_n + Q_n)$ dominate both P_n and Q_n, for every n. The sum of the densities of P_n and Q_n with respect to μ_n equals 2. Hence, each of the densities takes its values in the compact interval $[0, 2]$. By Prohorov's theorem every subsequence possesses a further subsequence along which

$$\frac{dP_n}{dQ_n} \overset{Q_n}{\rightsquigarrow} U, \qquad \frac{dQ_n}{dP_n} \overset{P_n}{\rightsquigarrow} V, \qquad W_n := \frac{dP_n}{d\mu_n} \overset{R_n}{\rightsquigarrow} W,$$

for certain random variables U, V and W. Every W_n has expectation 1 under μ_n. In view of the boundedness, the weak convergence of the sequence W_n implies convergence of moments, and the limit variable has mean $\mathrm{E}W = 1$ as well. For a given bounded, continuous function f, define a function $g : [0, 2] \mapsto \mathbb{R}$ by $g(w) = f\bigl(w/(2-w)\bigr)(2-w)$ for $0 \leq w < 2$ and $g(2) = 0$. Then g is bounded and continuous. Because $dP_n/dQ_n = W_n/(2 - W_n)$ and $dQ_n/d\mu_n = 2 - W_n$, the portmanteau lemma yields

$$\mathrm{E}_{Q_n} f\left(\frac{dP_n}{dQ_n}\right) = \mathrm{E}_{\mu_n} f\left(\frac{dP_n}{dQ_n}\right)\frac{dQ_n}{d\mu_n} = \mathrm{E}_{\mu_n} g(W_n) \to \mathrm{E}f\left(\frac{W}{2 - W}\right)(2 - W),$$

where the integrand in the right side is understood to be $g(2) = 0$ if $W = 2$. By assumption, the left side converges to $\mathrm{E}f(U)$. Thus $\mathrm{E}f(U)$ equals the right side of the display for every continuous and bounded function f. Take a sequence of such functions with $1 \geq f_m \downarrow 1_{\{0\}}$, and conclude by the dominated-convergence theorem that

$$\mathrm{P}(U = 0) = \mathrm{E}1_{\{0\}}(U) = \mathrm{E}1_{\{0\}}\left(\frac{W}{2 - W}\right)(2 - W) = 2\mathrm{P}(W = 0).$$

By a similar argument, $\mathrm{E}f(V) = \mathrm{E}f\bigl((2 - W)/W\bigr)W$ for every continuous and bounded function f, where the integrand on the right is understood to be zero if $W = 0$. Take a sequence $0 \leq f_m(x) \uparrow x$ and conclude by the monotone convergence theorem that

$$\mathrm{E}V = \mathrm{E}\left(\frac{2 - W}{W}\right)W = \mathrm{E}(2 - W)1_{W>0} = 2\mathrm{P}(W > 0) - 1.$$

Combination of the last two displays shows that $\mathrm{P}(U = 0) + \mathrm{E}V = 1$. ■

6.5 ***Example (Asymptotic log normality).*** The following special case plays an important role in the asymptotic theory of smooth parametric models. Let P_n and Q_n be probability measures on arbitrary measurable spaces such that

$$\frac{dP_n}{dQ_n} \overset{Q_n}{\rightsquigarrow} e^{N(\mu,\sigma^2)}.$$

Then $Q_n \triangleleft P_n$. Furthermore, $Q_n \triangleleft \triangleright P_n$ if and only if $\mu = -\frac{1}{2}\sigma^2$.

Because the (log normal) variable on the right is positive, the first assertion is immediate from (ii) of the theorem. The second follows from (iii) with the roles of P_n and Q_n switched, on noting that $\mathrm{E}\exp N(\mu, \sigma^2) = 1$ if and only if $\mu = -\frac{1}{2}\sigma^2$.

A mean equal to minus half times the variance looks peculiar, but we shall see that this situation arises naturally in the study of the asymptotic optimality of statistical procedures. □

The following theorem solves the problem of obtaining a Q_n-limit law from a P_n-limit law that we posed in the introduction. The result, a version of *Le Cam's third lemma*, is in perfect analogy with the nonasymptotic situation.

6.6 *Theorem.* *Let P_n and Q_n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{A}_n)$, and let $X_n : \Omega_n \mapsto \mathbb{R}^k$ be a sequence of random vectors. Suppose that $Q_n \triangleleft P_n$ and*

$$\left(X_n, \frac{dQ_n}{dP_n}\right) \overset{P_n}{\rightsquigarrow} (X, V).$$

Then $L(B) = \mathrm{E}1_B(X)\, V$ defines a probability measure, and $X_n \overset{Q_n}{\rightsquigarrow} L$.

Proof. Because $V \geq 0$, it follows with the help of the monotone convergence theorem that L defines a measure. By contiguity, $\mathrm{E}V = 1$ and hence L is a probability measure. It is immediate from the definition of L that $\int f\, dL = \mathrm{E}f(X)\, V$ for every measurable indicator function f. Conclude, in steps, that the same is true for every simple function f, any nonnegative measurable function, and every integrable function.

If f is continuous and nonnegative, then so is the function $(x, v) \mapsto f(x)\, v$ on $\mathbb{R}^k \times [0, \infty)$. Thus

$$\liminf \mathrm{E}_{Q_n} f(X_n) \geq \liminf \int f(X_n)\, \frac{dQ_n}{dP_n}\, dP_n \geq \mathrm{E}f(X)V,$$

by the portmanteau lemma. Apply the portmanteau lemma in the converse direction to conclude the proof that $X_n \overset{Q_n}{\rightsquigarrow} L$. ■

6.7 *Example (Le Cam's third lemma).* The name *Le Cam's third lemma* is often reserved for the following result. If

$$\left(X_n, \log\frac{dQ_n}{dP_n}\right) \overset{P_n}{\rightsquigarrow} N_{k+1}\left(\begin{pmatrix} \mu \\ -\frac{1}{2}\sigma^2 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^T & \sigma^2 \end{pmatrix}\right),$$

then

$$X_n \overset{Q_n}{\rightsquigarrow} N_k(\mu + \tau, \Sigma).$$

In this situation the asymptotic covariance matrices of the sequence X_n are the same under P_n and Q_n, but the mean vectors differ by the asymptotic covariance τ between X_n and the log likelihood ratios.†

The statement is a special case of the preceding theorem. Let (X, W) have the given $(k+1)$-dimensional normal distribution. By the continuous mapping theorem, the sequence $(X_n, dQ_n/dP_n)$ converges in distribution under P_n to (X, e^W). Because W is $N(-\frac{1}{2}\sigma^2, \sigma^2)$-distributed, the sequences P_n and Q_n are mutually contiguous. According to the abstract

† We set $\log 0 = -\infty$; because the normal distribution does not charge the point $-\infty$ the assumed asymptotic normality of $\log dQ_n/dP_n$ includes the assumption that $P_n(dQ_n/dP_n = 0) \to 0$.

version of Le Cam's third lemma, $X_n \overset{Q_n}{\rightsquigarrow} L$ with $L(B) = \mathrm{E}1_B(X)e^W$. The characteristic function of L is $\int e^{it^T x}\, dL(x) = \mathrm{E}e^{it^T X} e^W$. This is the characteristic function of the given normal distribution at the vector $(t, -i)$. Thus

$$\int e^{it^T x}\, dL(x) = e^{it^T\mu - \frac{1}{2}\sigma^2 - \frac{1}{2}(t^T, -i)\begin{pmatrix} \Sigma & \tau \\ \tau^T & \sigma^2 \end{pmatrix}\begin{pmatrix} t \\ -i \end{pmatrix}} = e^{it^T(\mu+\tau) - \frac{1}{2}t^T\Sigma t}.$$

The right side is the characteristic function of the $N_k(\mu + \tau, \Sigma)$ distribution. □

Notes

The concept and theory of contiguity was developed by Le Cam in [92]. In his paper the results that were later to become known as Le Cam's lemmas are listed as a single theorem. The names "first" and "third" appear to originate from [71]. (The second lemma is on product measures and the first lemma is actually only the implication (iii) ⇒ (i).)

PROBLEMS

1. 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 and only if the sequence μ_n is bounded.
2. Let P_n and Q_n be 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 $P_n \triangleleft\triangleright Q_n$ if and only if $\theta_n = O(1/\sqrt{n})$.
3. Let P_n and Q_n be the law of a sample of size n from the uniform distribution on $[0, 1]$ or $[0, 1+1/n]$, respectively. Show that $P_n \triangleleft Q_n$. Is it also true that $Q_n \triangleleft P_n$? Use Lemma 6.4 to derive your answers.
4. Suppose that $\|P_n - Q_n\| \to 0$, where $\|\cdot\|$ is the total variation distance $\|P - Q\| = \sup_A \big|P(A) - Q(A)\big|$. Show that $P_n \triangleleft\triangleright Q_n$.
5. Given $\varepsilon > 0$ find an example of sequences such that $P_n \triangleleft\triangleright Q_n$, but $\|P_n - Q_n\| \to 1 - \varepsilon$. (The maximum total variation distance between two probability measures is 1.) This exercise shows that it is wrong to think of contiguous sequences as being close. (Try measures that are supported on just two points.)
6. Give a simple example in which $P_n \triangleleft Q_n$, but it is not true that $Q_n \triangleleft P_n$.
7. Show that the constant sequences $\{P\}$ and $\{Q\}$ are contiguous if and only if P and Q are absolutely continuous.
8. If $P \ll Q$, then $Q(A_n) \to 0$ implies $P(A_n) \to 0$ for every sequence of measurable sets. How does this follow from Lemma 6.4?