Stats 300b: Theory of Statistics **Winter 2017**

Lecture 18 – March 9

Lecturer: John Duchi *Scribe: Michael Celentano*

Warning: *these notes may contain factual errors*

Reading: Van der Vaart, *Asymptotic Statistics*, Chapter 6.

Outline:

1. Absolute continuity of measures.
2. Contiguity and Asymptotics; LeCam's Lemmas.
3. Exact Calculations with Testing.

1 Absolute Continuity

We begin with the notion that allows us to change measures in a non-asymptotic setting.

Definition 1.1. *Probability measure Q is* absolutely continuous *with respect to probability measure P, written $Q \ll P$, if $P(A) = 0$ implies $Q(A) = 0$ for any set A.*

If $Q \ll P$, then the Radon-Nikodym theorem says that there exists a nonnegative measurable function g, denoted $\frac{dQ}{dP}$, such that $\mathbb{E}_Q[f] = \mathbb{E}_P[fg] = \int fg dP = \int f \frac{dQ}{dP} dP$ for all f integrable with respect to Q. The following corollary is another formulation of some of these ideas.

Corollary 1. *Let M be the joint measure (law) of the pair $(X, V) := \left(X, \frac{dQ}{dP}\right)$ under distribution P. (So M is defined on $\mathcal{X} \times \mathbb{R}_+$). Then $V \geq 0$, $\mathbb{E}_M[V] = 1$, and $Q(B) = \mathbb{E}_P\left[\mathbf{1}\{B\}(X)\frac{dQ}{dP}\right] = \mathbb{E}_P[\mathbf{1}\{B\}(X)V] = \int_{B \times \mathbb{R}_+} V dM(x, v)$.*

Idea: If we know $\frac{dQ}{dP}$ and we can compute things like Pf, then we can compute Qf too. Can we do this in an asymptotic sense? We would like to find an asymptotic version of Corollary 1.

2 Contiguity and Asymptotics; LeCam's Lemmas.

The notion that will allow to perform the asymptotic versions of the previous calculations is contiguity.

Definition 2.1. *A sequence $\{Q_n\}$ of distributions is* contiguous *with respect to $\{P_n\}$, written $Q_n \triangleleft P_n$, if $P_n(A_n) \to 0$ implies $Q_n(A_n) \to 0$ for any sequence of sets A_n. Sequences $\{Q_n\}$ and $\{P_n\}$ are* mutually contiguous*, written $Q_n \triangleleft\triangleright P_n$, if $Q_n \triangleleft P_n$ and $P_n \triangleleft Q_n$.*

Below we will characterize contiguity with conditions on the limits of density representations of P_n and Q_n. Because P_n and Q_n need not be absolutely continuous with respect to each other, nor are we a priori provided some mutually dominating measure, the following observations are useful.

Observation 2. *Even if* $Q_n \not\ll P_n$, *we can still consider* $Q_n = Q_n^{\|} + Q_n^{\perp}$ *where* $Q_n^{\|} \ll P$ *and* $Q_n^{\perp} \perp P$. *If we define* $\frac{dQ_n}{dP_n} := \frac{dQ_n^{\|}}{dP_n}$, *then* $\frac{dQ_n}{dP_n} \geq 0$ *and* $\mathbb{E}_{P_n}\left[\frac{dQ_n}{dP_n}\right] = Q_n^{\|}(\Omega) = 1 - Q_n^{\perp}(\Omega) \leq 1$. *Thus, under* P_n, *the sequence* $\frac{dQ_n}{dP_n}$ *is tight.*

Observation 3. *We can always assume without loss of generality that* P_n *and* Q_n *all have densities* p_n *and* q_n *with respect to some finite base measure* μ. *One suitable measure is* $\mu = \sum_{n=1}^{\infty} 2^{-n}(P_n + Q_n)$, *which has total mass 2. We can also without loss of generality take* $\frac{dQ_n}{dP_n} = \frac{q_n}{p_n}$, *so that* $\int \frac{q_n}{p_n} dP_n = \int q_n \mathbf{1}\{p_n > 0\}\, d\mu = Q_n^{\|}(\Omega)$.

We are now ready to state alternative characterizations of contiguity.

Lemma 4 (LeCam's First Lemma, or "Limits determine contiguity")**.** *The following are equivalent:*

1. $Q_n \triangleleft P_n$.
2. *If* $\frac{dQ_n}{dP_n} \xrightarrow[Q_n]{d} U$ *along a subsequence, then* $\mathbb{P}(U > 0) = 1$.
3. *If* $\frac{dQ_n}{dP_n} \xrightarrow[P_n]{d} V$ *along a subsequence, then* $\mathbb{E}[V] = 1$.
4. *If* $T_n \xrightarrow{P_n} 0$, *then* $T_n \xrightarrow{Q_n} 0$.

Proof of (4) $\Rightarrow$ (1): Take A_n such that $P_n(A_n) \to 0$. Define $T_n = \mathbf{1}\{A_n\}$. Then certainly $T_n \overset{P_n}{\to} 0$, so $T_n \overset{Q_n}{\to} 0$. That is, $Q_n(A_n) \to 0$. □

The remaining parts of the proof can be found in Van der Vaart Chapter 6.

The following two examples will be useful in later developments.

Example 1 (Asymptotic log normality)**:** Suppose we have $\log \frac{dP_n}{dQ_n} \xrightarrow[Q_n]{d} \mathsf{N}(\mu, \sigma^2)$. Then, by the continuous mapping theorem, we have $\frac{dP_n}{dQ_n} \xrightarrow[Q_n]{d} \exp(\mathsf{N}(\mu, \sigma^2))$ is greater than 0 with probability 1. Applying the second characterization of LeCam's First Lemma implies $Q_n \triangleleft P_n$. On the other hand, based on our knowledge of the moment generating function for normal random variables, $\mathbb{E}[\exp(\mathsf{N}(\mu, \sigma^2)] = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$. Applying the third characterization of LeCam's First Lemma gives that $P_n \triangleleft Q_n$ if and only if $\exp\left(\mu + \frac{1}{2}\sigma^2\right) = 1$. That is, $\mu = -\frac{1}{2}\sigma^2$. ♣

Example 2 (Smooth likelihoods)**:** Suppose $\{P_\theta\}_{\theta \in \Theta}$ has densities p_θ, and p_θ is smooth enough in θ that $\log p_\theta$ has a Taylor expansion around $\theta_0 \in \operatorname{int} \Theta$. That is,

$$\log p_{\theta_0 + h} = \ell_{\theta_0} + h^T \nabla \ell_{\theta_0} + \frac{1}{2} h^T \nabla^2 \ell_{\theta_0} h + O(\|h\|^3)$$

where $\ell_\theta = \log p_\theta$.

A Taylor expansion gives

$$\begin{aligned}
\log \frac{p_{\theta_0+h/\sqrt{n}}(x_1,\ldots,x_n)}{p_{\theta_0}(x_1,\ldots,x_n)} &= \sum_{i=1}^n \log \frac{p_{\theta_0+h/\sqrt{n}}(x_i)}{p_{\theta_0}(x_i)} = \sum_{i=1}^n \left(\ell_{\theta_0+h/\sqrt{n}}(x_i) - \ell_{\theta_0}(x_i)\right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \nabla \ell_{\theta_0}(x_i) + \frac{1}{2n} h^T \sum_{i=1}^n \nabla^2 \ell_{\theta_0}(x_i) h + o_{P_{\theta_0}}(1) \\
&= \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \nabla \ell_{\theta_0}(x_i)}_{\xrightarrow{d} \mathsf{N}(0, h^T I_\theta h)} - \frac{1}{2} h^T I_{\theta_0} h + o_{P_{\theta_0}}(1) \\
&\xrightarrow{p} \mathsf{N}\left(-\frac{1}{2} h^T I_{\theta_0} h, h^T I_{\theta_0} h\right).
\end{aligned}$$

Surprisingly, this is exactly the condition for mutual contiguity from Example 1! Thus $P_{\theta_0}^n \lhd\rhd P_{\theta_0+h/\sqrt{n}}^n$. Local alternatives and the null are mutually contiguous! ♣

In analogy with Corollary 1 and in light of the previous example, we now wonder whether mutual contiguity tells us something about limit distributions under alternatives. The next theorem is the beginning of an answer.

Theorem 5 (LeCam)**.** *Let P_n, Q_n be distributions on $X_n \in \mathbb{R}^d$. If $Q_n \lhd P_n$ and $\left(X_n, \frac{dQ_n}{dP_n}\right) \xrightarrow[P_n]{d} (X, V)$, then $L(B) := \mathbb{E}[\mathbf{1}\{B\}(X)V]$ is a probability measure (ie. $\mathbb{E}[V] = 1$ and $V \geq 0$) and $X_n \xrightarrow[Q_n]{d} W$ where $W \sim L$.*

Proof (See Van der Vaart Chapter 6). □

The following example is an important application of Theorem 5.

Example 3 (LeCam's Third Lemma)**:** If $\left(X_n, \log \frac{dQ_n}{dP_n}\right) \xrightarrow[P_n]{d} \mathsf{N}\left(\begin{pmatrix}\mu \\ -\frac{1}{2}\sigma^2\end{pmatrix}, \begin{pmatrix}\Sigma & \tau \\ \tau^T & \sigma^2\end{pmatrix}\right)$, then $X_n \xrightarrow[Q_n]{d} \mathsf{N}(\mu+\tau, \Sigma)$. The proof, which we omit, invovles a calculation with characteristic functions. ♣

3 Exact Calculations with Testing

We briefly introduce our next question, which will be answered in future lectures.

Consider the simple hypothesis test P_0 vs. P_1. The total variation distance comes up naturally in this context when considering optimal error rates. Indeed,

$$\begin{aligned}
\inf_{\Phi:\mathcal{X}\to\{0,1\}} \left(P_0(\Phi \neq 0) + P_1(\Phi \neq 1)\right) &= \inf_A \{P_0(A^c) + P_1(A)\} = \inf_A \{1 - (P_0(A) - P_1(A))\} \\
&= 1 - \|P_0 - P_1\|_{\mathrm{TV}}.
\end{aligned}$$

Thus, given a sequence of nulls and alternatives $\{P_{0,n}\}$ and $\{P_{1,n}\}$, if $\|P_{0,n} - P_{1,n}\|_{\mathrm{TV}}$ converges to something in $(0,1)$, then it is impossible to test exactly whether the null or the alternative is true. The result of Problem Set 8, Question 2 implies that this also holds if $d_{\mathrm{hel}}(P_{0,n}, P_{1,n})$ converges to something in $(0,1)$.

We now ask: How/when do we have this type of limiting behavior?