Theoretical Statistics. Lecture 18.

Peter Bartlett

1. Asymptotic equicontinuity.
2. Donsker property.

Outline of today's lecture

We noticed that, in showing that the asymptotic distribution of Z-estimates is normal, what we really needed was continuity of the sample paths of the stochastic process defined by the estimating equations. This property is called **Asymptotic Equicontinuity**. It is closely related to the **Donsker property**.

Recall: Asymptotics of Z-estimators

Theorem:

$$\text{Suppose} \qquad \Psi(\theta) = P\psi_\theta,$$

$$\Psi_n(\theta) = P_n\psi_\theta,$$

$$\Psi(\theta_0) = 0,$$

$$\Psi_n(\hat\theta_n) = o_P(n^{-1/2}),$$

$$\dot\Psi_{\theta_0}^{-1} \text{ exists},$$

$$\sqrt{n}(\Psi - \Psi_n)(\hat\theta_n) - \sqrt{n}(\Psi - \Psi_n)(\theta_0) = o_P(1 + \sqrt{n}\|\hat\theta_n - \theta_0\|).$$

$$\text{Then } \sqrt{n}(\hat\theta_n - \theta_0) \rightsquigarrow N\left(0, (\dot\Psi_{\theta_0})^{-1} P\psi_{\theta_0}\psi_{\theta_0}^T (\dot\Psi_{\theta_0})^{-1}\right).$$

Recall: Asymptotic equicontinuity

Definition: Define

$$\mathbb{G}_n f = \sqrt{n}(P_n - P)f$$
$$F_{\delta_n} = \{f - g : f, g \in F,\ \rho_P(f-g) < \delta_n\},$$
$$\rho_P(f) = (P(f - Pf)^2)^{1/2}.$$

Then the empirical process $\mathbb{G}_n$ on F is **asymptotically equicontinuous** if, for every sequence $\delta_n \to 0$, $\|\mathbb{G}_n\|_{F_{\delta_n}} \xrightarrow{P} 0$.

Donsker Property

Definition: Suppose F satisfies

$$\text{for all } x,\ \sup_{f \in F} |f(x) - Pf| < \infty.$$

We say F is a **Donsker class** if $\mathbb{G}_n \rightsquigarrow \mathbb{G}$, where $\mathbb{G}$ is a tight random element in $\ell^\infty(F)$.

The limit process is the zero-mean Gaussian process with covariance function

$$\mathbf{E}\mathbb{G}f\mathbb{G}g = P(f - Pf)(g - Pg) = Pfg - PfPg.$$

This process is called the P-Brownian bridge.

Asymptotic equicontinuity

Lemma: F is Donsker iff

1. $\mathbb{G}_n$ is asymptotically equicontinuous on F.
2. F is totally bounded in $L_2(P)$.

Asymptotic equicontinuity

Proof (Donsker implies asymptotically equicontinuous):
Define $g : \ell^\infty(F) \times F \to \mathbb{R}$ by $g(z, f) = z(f)$ (consider the $L_2(P)$ pseudometric on F). Then g is continuous at (z, f) for which $f \mapsto z(f)$ is continuous.

Consider $\|\mathbb{G}_n\|_{F_{\delta_n}}$. For any random difference sequence $f_n - g_n$ in F_{δ_n}, since $\rho_P(f_n - g_n) < \delta_n$, we have $f_n - g_n - \mathbf{E}(f_n - g_n) \rightsquigarrow 0$. Also, since F is Donsker, $\mathbb{G}_n \rightsquigarrow \mathbb{G}_P$ in $\ell^\infty(F)$. Because of the fact (vdV Lemma 18.15) that almost all sample paths of $\mathbb{G}_P$ are continuous on F, g is continuous at almost every point $(\mathbb{G}_P, 0)$. By the definition of $\mathbb{G}_n$ and by the continuous mapping theorem,
$\mathbb{G}_n(f_n - g_n) = \mathbb{G}_n(f_n - g_n - \mathbf{E}(f_n - g_n)) =$
$g(\mathbb{G}_n, f_n - g_n - \mathbf{E}(f_n - g_n)) \rightsquigarrow g(\mathbb{G}_P, 0) = 0$. Thus, $\mathbb{G}_n(f_n - g_n) \xrightarrow{P} 0$.
Since this is true for an arbitrary random difference sequence

$f_n - g_n \in F_{\delta_n}$, it is true for the supremum.

Donsker classes

We say that F has an **envelope function** $x \mapsto B(x)$ if for all x and f, $|f(x)| \leq B(x)$

Donsker classes

Theorem: Suppose F has an envelope function B with $PB^2 < \infty$, and

$$\int_0^\infty \sup_Q \sqrt{\log \mathcal{N}(\epsilon \|B\|_{Q,2}, F, L_2(Q))}\, d\epsilon < \infty,$$

where the supremum is over all finite discrete probability measures on $\mathcal{X}$ satisfying $QB^2 > 0$. Then for all $\delta_n \to 0$, $\|\mathbb{G}_n\|_{F_{\delta_n}} \xrightarrow{P} 0$, where

$$F_\delta = \{f - g : f, g \in F,\ P(f-g)^2 < \delta\}$$

and F is totally bounded in $L_2(P)$. Hence, F is Donsker.

Donsker classes: Proof

$$
\begin{aligned}
\mathbf{E}\|\mathbb{G}_n\|_{F_{\delta_n}} &= \sqrt{n}\mathbf{E}\|P_n - P\|_{F_{\delta_n}} \\
&\le 2\sqrt{n}\mathbf{E}\|R_n\|_{F_{\delta_n}} \\
&\le c\mathbf{E}\int_0^\infty \sqrt{\log \mathcal{N}(\epsilon, F_{\delta_n}, L_2(P_n))}\, d\epsilon.
\end{aligned}
$$

For $\epsilon > \epsilon_n$, $\log \mathcal{N}(\epsilon, F_{\delta_n}, L_2(P_n)) = 0$, where

$$
\epsilon_n^2 = \sup_{f \in F_{\delta_n}} \|f\|_{L_2(P_n)}^2 = \frac{1}{n}\left\|\sum_{i=1}^n f^2(X_i)\right\|_{F_{\delta_n}}.
$$

Donsker classes: Proof

Also, for all Q,

$$\mathcal{N}(\epsilon, F_{\delta_n}, L_2(Q)) \leq N^2(\epsilon/2, F, L_2(Q)).$$

Thus, for $b_n = \|B\|_{L_2(P_n)}$,

$$\begin{aligned}
\mathbf{E}\|\mathbb{G}_n\|_{F_{\delta_n}} &\leq c\mathbf{E}b_n \int_0^{\epsilon_n/b_n} \sqrt{\log \mathcal{N}(\epsilon b_n, F_{\delta_n}, L_2(P_n))}\, d\epsilon, \\
&\leq c\mathbf{E}b_n \int_0^{\epsilon_n/b_n} \sup_Q \sqrt{\log \mathcal{N}(\epsilon b_n/2, F, L_2(Q))}\, d\epsilon.
\end{aligned}$$

Donsker classes: Proof

Since $b_n \stackrel{P}{\to} \mathbf{E}\|B\|_{L_2(P_n)} > 0$, the asymptotic equicontinuity follows from $\epsilon_n \stackrel{P}{\to} 0$. But

$$\begin{aligned}
\epsilon_n^2 &= \sup_{f \in F_{\delta_n}} \|f\|_{L_2(P_n)}^2 \\
&= \sup_{f \in F_{\delta_n}} P_n f^2 \\
&\le \sup_{f \in F_{\delta_n}} (P_n - P) f^2 + \sup_{f \in F_{\delta_n}} P f^2 \\
&\le \sup_{f,g \in F} (P_n - P)(f-g)^2 + \sup_{f \in F_{\delta_n}} P f^2.
\end{aligned}$$

The first term goes to zero because of the entropy condition. The second term goes to zero because of the definition of F_{δ_n}.

Donsker classes: Proof

Hence $\mathbf{E}\|G_n\|_{F_{\delta_n}} \to 0$, and so Markov's inequality implies $\|G_n\|_{F_{\delta_n}} \stackrel{P}{\to} 0$.

The entropy condition also implies F is totally bounded in $L_2(P)$, since the entropy condition means F is totally bounded in $L_2(P_n)$, and $\sup_{f,g\in F} |(P_n - P)(f-g)^2| \to 0$.

We've cheated slightly. (Where?)

The definition of F_δ used $L_2(P)$ instead of ρ_P. That is, we consider a smaller set, since we insist that the mean differences are also included.

But this is a minor difference. In particular, we can make F larger by shifting all means (that is, adding all translates by a constant), and this does not change the log covering numbers by more than a $\log(1/\epsilon)$ term. (Why?)

Recall: Asymptotics of Z-estimators

Theorem:

$$\begin{aligned} \text{Suppose} \qquad \Psi(\theta) &= P\psi_\theta, \\ \Psi_n(\theta) &= P_n\psi_\theta, \\ \Psi(\theta_0) &= 0, \\ \Psi_n(\hat{\theta}_n) &= o_P(n^{-1/2}), \\ \dot{\Psi}_{\theta_0}^{-1} &\text{ exists}, \\ \sqrt{n}(\Psi - \Psi_n)(\hat{\theta}_n) - \sqrt{n}(\Psi - \Psi_n)(\theta_0) &= o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|). \end{aligned}$$

$$\text{Then } \sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N\left(0, (\dot{\Psi}_{\theta_0})^{-1} P\psi_{\theta_0}\psi_{\theta_0}^T (\dot{\Psi}_{\theta_0})^{-1}\right).$$

Asymptotic normality of Z-estimates

We need to check why asymptotic equicontinuity of $F = \{\psi_\theta\}$, that is,

$$\|\mathbb{G}_n\|_{F_{\delta_n}} \xrightarrow{P} 0,$$

implies $\mathbb{G}_n(\psi_{\theta_0} - \psi_{\hat{\theta}_n}) = o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|)$.

For this, we notice that $\psi_{\theta_0} - \psi_{\hat{\theta}_n}$ goes to zero when $P(\psi_{\theta_0} - \psi_{\hat{\theta}_n})^2 \to 0$, which follows from differentiability of $P\psi$ at θ_0 and $\|\theta_0 - \hat{\theta}_n\|^2 \to 0$. Thus, the asymptotic equicontinuity condition implies that $\mathbb{G}_n(\psi_{\theta_0} - \psi_{\hat{\theta}_n}) = o_P(1)$ when $\|\hat{\theta}_n - \theta_0\| \xrightarrow{P} 0$.

Donsker classes: converse result

A class F is star-shaped if, for all $f \in F$, we also have $\lambda f \in F$ for $0 \leq \lambda \leq 1$. [PICTURE]

Theorem: If F is star-shaped, $\|f\|_\infty \leq B$ for all $f \in F$, and for some $\alpha > 0$,

$$\mathbf{E}\|R_n\|_{F-F} = \Omega(n^{-1/2+\alpha}),$$

where $F - F$ is the set of differences of functions in F, then F is not asymptotically equicontinuous.

Donsker classes: converse result

Proof:

We want to show that $\|\mathbb{G}_n\|_{F_{\delta_n}}$ stays large. But we know that

$$\begin{aligned}\|\mathbb{G}_n\|_{F_{\delta_n}} &= \sqrt{n}\|P_n - P\|_{F_{\delta_n}} \\ &\geq c\sqrt{n}\left(\mathbf{E}\|R_n\|_{F_{\delta_n}} - \frac{1}{\sqrt{n}}\right) \qquad \text{(with prob } \geq 1/2\text{).}\end{aligned}$$

Next, we need to relate $\|R_n\|_{F_{\delta_n}}$ to $\|R_n\|_{F-F}$.

Suppose $\|R_n\|_{F-F} = |R_n(f-g)|$. Since F is star-shaped, for any $0 \leq \lambda \leq 1$, λf and λg are in F, and so

$$\rho_P(\lambda(f-g)) = \lambda\rho_P(f-g) = \lambda(P(f-g-(Pf-Pg))^2)^{1/2} \leq 2\lambda B.$$

Donsker classes: converse result

Choosing $\lambda = c_1\delta_n$ (with $c_1 = 1/(2B)$) ensures that $\rho_P(\lambda(f-g)) \leq \delta_n$, so we see that

$$\|R_n\|_{F_{\delta_n}} \geq c_1\delta_n\|R_n\|_{F-F}.$$

Thus, with high probability,

$$\begin{aligned}
\|\mathbb{G}_n\|_{F_{\delta_n}} &\geq c\left(\sqrt{n}\mathbf{E}\|R_n\|_{F_{\delta_n}} - 1\right) \\
&\geq c_2\left(\sqrt{n}\delta_n\|R_n\|_{F-F} - 1\right) \\
&= \Omega\left(\delta_n n^{1/2} n^{-1/2+\alpha}\right) - O(1) \\
&= \Omega\left(\delta_n n^{\alpha}\right) - O(1).
\end{aligned}$$

Clearly, we can have $\delta_n \to 0$ but $\delta_n n^{\alpha} \to \infty$.

Donsker classes

Thus

1. Finiteness of the entropy integral (which suffices for the Donsker property) leads to $\|R_n\|_F$ decreasing at a $n^{-1/2}$ rate, and
2. A slower rate for $\|R_n\|_{F-F}$ implies that the Donsker property does not hold.

(We assumed a uniform bound on $\|f\|_\infty$ for $f \in F$, in order to relate $\|\mathbb{G}_n\|_{F_{\delta_n}}$ to $\mathbf{E}\|\mathbb{G}_n\|_{F_{\delta_n}}$ and to rescale F_{δ_n}, rather than an envelope function—i.e., for all x and all $f \in F$, $|f(x)| \leq B(x)$.)