Theoretical Statistics. Lecture 17.

Peter Bartlett

1. Asymptotic normality of Z-estimators: classical conditions.
2. Asymptotic equicontinuity.

Recall: Delta method

Theorem: Suppose $\phi : \mathbb{R}^k \to \mathbb{R}^m$ is differentiable at θ, and $\sqrt{n}(T_n - \theta) \rightsquigarrow T$, then

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(T)$$

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) - \phi'_\theta(\sqrt{n}(T_n - \theta)) \xrightarrow{P} 0.$$

Here, ϕ'_θ is the derivative (linear map) satisfying

$$\phi(\theta + h) - \phi(\theta) = \phi'_\theta(h) + o(\|h\|)$$

for $h \to 0$.

Asymptotic normality of Z-estimators

Theorem: Consider

$$\Psi_n(\theta) = P_n\psi_\theta, \qquad \Psi(\theta) = P\psi_\theta.$$

Suppose $\hat{\theta}_n \in \mathbb{R}$ is a zero of Ψ_n, $\theta_0 \in \mathbb{R}$ is a zero of Ψ, $\hat{\theta}_n \stackrel{P}{\to} \theta_0$. Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{-\sqrt{n}\Psi_n(\theta_0)}{\dot{\Psi}_n(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)\ddot{\Psi}_n(\tilde{\theta}_n)}$$

where $\tilde{\theta}_n = \lambda\hat{\theta}_n + (1-\lambda)\theta_0$ for some $0 \leq \lambda \leq 1$.
If $P\psi_{\theta_0}^2$ exists, $P\dot{\psi}_{\theta_0}$ exists and is non-zero, and $\ddot{\Psi}_n(\tilde{\theta}_n) = O_P(1)$, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N\left(0, P\psi_{\theta_0}^2/(P\dot{\psi}_{\theta_0})^2\right).$$

Asymptotic normality of Z-estimators: Proof

We take a Taylor series expansion of $\Psi_n(\hat{\theta}_n)$ around θ_0:

$$
\begin{aligned}
0 &= \Psi_n(\theta_0) + (\hat{\theta}_n - \theta_0)\dot{\Psi}_n(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 \ddot{\Psi}_n(\tilde{\theta}_n), \\
&= \Psi_n(\theta_0) + (\hat{\theta}_n - \theta_0)\left(\dot{\Psi}_n(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)\ddot{\Psi}_n(\tilde{\theta}_n)\right),
\end{aligned}
$$

where $\tilde{\theta}_n$ is between $\hat{\theta}_n$ and θ_0. Rearranging gives the first equality of the theorem:

$$
\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{-\sqrt{n}\Psi_n(\theta_0)}{\dot{\Psi}_n(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)\ddot{\Psi}_n(\tilde{\theta}_n)}
$$

Since $P\psi_{\theta_0}^2$ exists,

$$
-\sqrt{n}\Psi_n(\theta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta_0}(X_i) \rightsquigarrow N(P\psi_{\theta_0}, \mathrm{var}(\psi_{\theta_0})) = N(0, P\psi_{\theta_0}^2).
$$

Asymptotic normality of Z-estimators: Proof

Since $P\dot{\psi}_{\theta_0}$ exists,

$$\dot{\Psi}_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} \dot{\psi}_{\theta_0}(X_i) \xrightarrow{P} P\dot{\psi}_{\theta_0}.$$

Finally,

$$\frac{1}{2}(\hat{\theta}_n - \theta_0)\ddot{\Psi}_n(\tilde{\theta}_n) = \frac{1}{2}o_P(1)O_P(1) = o_P(1).$$

Slutsky's lemma gives the result.

Asymptotic normality of Z-estimators

Analogous result for $\theta \in \mathbb{R}^p$:

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

Asymptotic normality of Z-estimators

Consider the (classical) conditions we used:

- $P\psi_{\theta_0}^2$ exists:
  The form of the estimating equations (and the distribution) keep the variance under control.
- $P\dot{\psi}_{\theta_0}$ exists and is non-singular:
  This requires the function ψ to be regular at its zero.
- $\ddot{\Psi}_n(\tilde{\theta}_n) = O_P(1)$: We used this to control the remainder term in the Taylor series. But it is not necessary to have these derivatives existing. We can replace this with a stochastic equicontinuity condition: showing that $\{\psi_\theta : \|\theta - \theta_0\| \leq \epsilon\}$ is a **Donsker class** for some $\epsilon > 0$.

Asymptotics of Z-estimators

[Construct conditions of theorem below as we proceed through the proof.]

Suppose $\Psi(\theta) = P\psi_\theta$, $\Psi_n(\theta) = P_n\psi_\theta$ and $\Psi(\theta_0) = 0$. Then

$$\sqrt{n}(\Psi - \Psi_n)(\theta_0) = \sqrt{n}(P - P_n)\psi_{\theta_0} \rightsquigarrow N(0, P\psi_{\theta_0}\psi_{\theta_0}^T).$$

Suppose also that $\hat{\theta}_n$ is an approximate zero of Ψ_n (we'll assume that $\Psi_n(\hat{\theta}_n) = o_P(n^{-1/2})$). We'd like to show that $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow Z$ for some normal Z.

If Ψ is differentiable at θ_0, we can write

$$\Psi(\hat{\theta}_n) = \Psi(\theta_0) + \dot{\Psi}_{\theta_0}(\hat{\theta}_n - \theta_0) + o_P(\|\hat{\theta}_n - \theta_0\|).$$

Asymptotics of Z-estimators

Assuming that the inverse of $\dot\Psi_{\theta_0}$ exists, we can rearrange this to:

$$\begin{aligned}\sqrt{n}(\hat\theta_n - \theta_0) &= \sqrt{n}(\dot\Psi_{\theta_0})^{-1}\left(\Psi(\hat\theta_n) - \Psi(\theta_0)\right) + o_P(\sqrt{n}\|\hat\theta_n - \theta_0\|)\\ &= \sqrt{n}(\dot\Psi_{\theta_0})^{-1}\left(\Psi(\hat\theta_n) - \Psi_n(\hat\theta_n)\right) + o_P(1 + \sqrt{n}\|\hat\theta_n - \theta_0\|),\end{aligned}$$

from the definition of θ_0 and the condition that $\hat\theta_n$ is an approximate solution to the estimating equations Ψ_n.

We would like to relate the term $\sqrt{n}(\Psi - \Psi_n)(\hat\theta_n)$ to the asymptotically normal $\sqrt{n}(\Psi - \Psi_n)(\theta_0)$.

Asymptotics of Z-estimators

If we knew that

$$\sqrt{n}(\Psi - \Psi_n)(\hat{\theta}_n) - \sqrt{n}(\Psi - \Psi_n)(\theta_0) = o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|),$$

then we would have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\dot{\Psi}_{\theta_0})^{-1}\left(\Psi(\theta_0) - \Psi_n(\theta_0)\right) + o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|).$$

Dividing both sides by $\sqrt{n}$ shows that the norm of the parameter error decreases as $\|\hat{\theta}_n - \theta_0\| = o_P(1/\sqrt{n})$, so

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

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 equicontinuity

We need to know that, as $\hat{\theta}_n$ approaches θ_0, we have that $\mathbb{G}_n(\psi_{\hat{\theta}_n} - \psi_{\theta_0})$ becomes small, where $\mathbb{G}_n$ is the scaled empirical process

$$\mathbb{G}_n = \sqrt{n}(P - P_n).$$

This is a continuity condition: the random variable $\mathbb{G}_n \psi_\theta$ is continuous in its indexing variable θ. That is, the sample paths are continuous. In the case of vector ψ_θ, we want the changes in $\mathbb{G}_n$ to be small uniformly across the dimensions. More generally, when we consider infinite-dimensional θ, we can think of $\psi_\theta(x) = h \mapsto \psi_{\theta,h}(x)$ where $h \in H$ (a set of size the dimensionality of θ). In that case, we need the changes in $\mathbb{G}_n$ to be uniformly small over $h \in H$. This is called asymptotic continuity of the stochastic process. We'll see later that the pseudometric involves the variance.

Stochastic Convergence in Metric Spaces

vdV18.

Definition: For a set T, define $\ell^\infty(T)$ as the set of functions $z : T \to \mathbb{R}$ with $\|z\|_T < \infty$, where $\|z\|_T = \sup_{t \in T} |z(t)|$.

We can define convergence in a metric space through the characterization given by the portmanteau lemma:

Definition: For random elements X_n, X of a metric space (M, d), we say $X_n \rightsquigarrow X$ if $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$ for all bounded, continuous $f : M \to \mathbb{R}$.

Central Limit Theorem: Empirical Distribution Functions

The law of large numbers: $|F_n(t) - F(t)| \stackrel{P}{\to} 0$.

The uniform law of large numbers (GC): $\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \stackrel{P}{\to} 0$.

The central limit theorem:
$\sqrt{n}(F_n(t_1) - F(t_1), F_n(t_2) - F(t_2), F_n(t_k) - F(t_k)) \rightsquigarrow$
$(\mathbb{G}_F(t_1), \mathbb{G}_F(t_2), \ldots, \mathbb{G}_F(t_k))$, where the limit is a multivariate normal distribution with mean zero and covariance

$$\begin{aligned} \mathbf{E}\mathbb{G}_F(t_i)\mathbb{G}_F(t_j) &= \mathbf{E}1[X \leq t_i]1[X \leq t_j] - \mathbf{E}1[X \leq t_i]\mathbf{E}1[X \leq t_j] \\ &= F[t_i \wedge t_j] - F(t_i)F(t_j). \end{aligned}$$

Central Limit Theorem: Empirical Distribution Functions

The Donsker theorem shows that the sequence of empirical processes (random functions) $\sqrt{n}(F_n - F)$ converges weakly to a Gaussian process $\mathbb{G}_F$ with zero mean and this covariance. This is an F-Brownian bridge process. If F is uniform, it is a uniform-Brownian bridge. (Bridge because it is constrained to be 0 at 0 and 1.) For a uniform bridge $\mathbb{G}$, the F-Brownian bridge is $t \mapsto \mathbb{G}(F(t))$.

Weak convergence in a metric space

Definition: For (T, ρ) a totally bounded pseudometric space, define $UC(T, \rho)$ as the set of uniformly continuous functions $z : T \to \mathbb{R}$.

Notice that $UC(T, \rho) \subseteq \ell^\infty(T)$.

Weak convergence in a metric space

vdV Thm 18.14, Lemma 18.15:

Theorem: A sequence $X_n : \Omega_n \to \ell^\infty(T)$ converges weakly to a tight random element X (that is, $\forall \epsilon$, $\exists$ compact K, $\Pr(X \notin K) < \epsilon$) iff

1. $\forall k$, $\forall t_1, \ldots, t_k \in T$, $\exists Z$, $(X_n(t_1), \ldots, X_n(t_k)) \rightsquigarrow Z$, and
2. $\forall \epsilon, \eta$, $\exists$ partition $T_1, \ldots, T_k \subseteq T$ such that

$$\lim_{n\to\infty} \sup P^*(\sup_i \sup_{s,t \in T_i} |X_n(s) - X_n(t)| \geq \epsilon) \leq \eta.$$

Furthermore, under (1), (2), there is a pseudometric ρ on T such that (T, ρ) is totally bounded and X has almost all sample paths in $UC(T, \rho)$.
If X is zero-mean Gaussian, $\rho(s,t) = s.d.(X_s - X_t)$.

Weak convergence in a metric space

ρ is defined in terms of the sequence of partitions—as something like a tree distance in terms of the successive refinements of the $T_1, \ldots, T_k$.

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 $\delta_n \to 0$, $\|\mathbb{G}_n\|_{F_{\delta_n}} \stackrel{P}{\to} 0$.