Z-estimators

Jianqiao Wang

September 9, 2021

Overview

1. Consistency
2. Weak Convergence
3. Using the Delta Method

Z-Estimators

A Z-estimator $\hat{\theta}_n$ is the approximate zero of a data-dependent function. More precisely, let the parameter space be Θ and $\Psi_n : \Theta \to \mathbb{L}$ be a data-dependent function between two norm spaces, with norms $\|\cdot\|$ and $\|\cdot\|_{\mathbb{L}}$. A quantity $\hat{\theta}_n \in \Theta$ is a Z-estimator if

$$\|\Psi_n(\hat{\theta}_n)\|_{\mathbb{L}} \xrightarrow{P} 0$$

Consistency

The main consistency result is stated in Chapter 2 and we will now extend it to the bootstrapped Z-estimator.

The map $\Psi : \Theta \to \mathbb{L}$ is identifiable at $\theta_0 \in \Theta$ if

$$\|\Psi(\theta_n)\|_{\mathbb{L}} \to 0 \text{ implies } \|\theta_n - \theta_0\| \to 0 \text{ for any } \{\theta_n\} \in \Theta$$

We will use the bootstrap-weighted empirical process $\mathbb{P}_n^\circ$ to denote either the nonparametric bootstrapped empirical process or the multiplier bootstrapped empirical process defined by
$f \to \mathbb{P}_n^\circ f = n^{-1} \sum_{i=1}^n (\xi_i/\bar{\xi}) f(X_i)$, where $\xi_1, \ldots, \xi_n$ are i.i.d. positive weights with $0 < \mu = E\xi_1$ and $\bar{\xi} = n^{-1} \sum_{i=1}^n \xi_i$.
Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$.

Consistency

Theorem (Master Z-estimator theorem for consistency)

Let $\theta \mapsto \Psi(\theta) = P\psi_\theta, \theta \mapsto \Psi_n(\theta) = \mathbb{P}_n\psi_\theta$ *and* $\theta \mapsto \Psi_n^\circ(\theta) = \mathbb{P}_n^\circ\psi_\theta$ *where* Ψ *is identifiable and the class* $\{\psi_\theta : \theta \in \Theta\}$ *is P-Glivenko-Cantelli. Then, provided* $\|\Psi_n(\hat{\theta}_n)\|_{\mathbb{L}} = o_P(1)$ *and*

$$P(\|\Psi_n^\circ(\hat{\theta}_n^\circ)\|_{\mathbb{L}} > \eta | \mathcal{X}_n) = o_P(1) \text{ for every } \eta > 0 \tag{1}$$

we have both $\|\hat{\theta}_n - \theta_0\| = o_P(1)$ *and* $P(\|\hat{\theta}_n^\circ - \theta_0\| > \eta | \mathcal{X}_n) = o_P(1)$ *for every* $\eta > 0$.

Proof

$\|\hat{\theta}_n - \theta_0\| = o_P(1)$ is a conclusion from theorem 2.10. For conditional bootstrap result, (1) implies that for some sequence $\eta_n \downarrow 0$, $P(\|\Psi(\hat{\theta}_n^\circ)\|_{\mathbb{L}} > \eta_n | \mathcal{X}_n) = o_P(1)$, since

$$P(\sup_{\theta \in \Theta} \|\Psi_n^\circ(\theta) - \Psi(\theta)\| > \eta | \mathcal{X}_n) = o_P(1)$$

for all $\eta > 0$ by theorem 10.13 and 10.15. Thus, for any $\epsilon > 0$,

$$\begin{aligned} P(\|\hat{\theta}_n^\circ - \theta_0\| > \epsilon | \mathcal{X}_n) \le & P(\|\hat{\theta}_n^\circ - \theta_0\| > \epsilon, \|\Psi(\hat{\theta}_n^\circ)\|_{\mathbb{L}} \le \eta_n | \mathcal{X}_n) \\ & + P(\|\Psi(\hat{\theta}_n^\circ)\|_{\mathbb{L}} > \eta_n | \mathcal{X}_n) \\ \xrightarrow{P} & 0. \end{aligned}$$

Identifiability condition of Ψ implies that for all $\delta > 0$ there exists an $\eta > 0$ such that $\|\Psi(\theta)\|_{\mathbb{L}} < \eta$ implies $\|\theta - \theta_0\| < \delta$.

Weak Convergence

Theorem (Theorem 2.11 in Chapter 2)

Assume that $\Psi(\theta_0) = 0$ *for some* θ_0 *in the interior of* Θ, $\sqrt{n}\Psi_n(\hat{\theta}_n) \xrightarrow{P} 0$, *and* $\|\hat{\theta}_n - \theta_0\| \xrightarrow{P} 0$ *for the random sequence* $\{\hat{\theta}_n\} \in \Theta$. *Assume also that* $\sqrt{n}(\Psi_n - \Psi)(\theta_0) \rightsquigarrow Z$, *for some tight random* Z, *and that*

$$\frac{\|\sqrt{n}(\Psi_n(\hat{\theta}_n) - \Psi(\hat{\theta}_n)) - \sqrt{n}(\Psi_n(\theta_0) - \Psi(\theta_0))\|_{\mathbb{L}}}{1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|} \xrightarrow{P} 0 \tag{2}$$

If $\theta \mapsto \Psi(\theta)$ *is Frechet-differentiable at* θ_0 *with continuously-invertible derivative* $\dot{\Psi}_{\theta_0}$, *then*

$$\|\sqrt{n}\dot{\Psi}_{\theta_0}(\hat{\theta}_n - \theta_0) + \sqrt{n}(\Psi_n - \Psi)(\theta_0)\|_{\mathbb{L}} \xrightarrow{P} 0 \tag{3}$$

and thus $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Psi}_{\theta_0}^{-1}(Z)$.

Previous definition

A map $\phi : \Theta \subset \mathbb{D} \mapsto \mathbb{L}$ is *Frechet-differentiable* at $\theta \in \Theta$ if there exists a continuous linear map $\phi'_\theta : \mathbb{D} \mapsto \mathbb{L}$ with

$$\frac{\|\phi(\theta + h_n) - \phi(\theta) - \phi'_\theta(h_n)\|_{\mathbb{L}}}{1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|} \to 0$$

for all sequences $\{h_n\} \subset \mathbb{D}$ with $\|h_n\| \to 0$ and $\theta + h_n \in \Theta$ for all $n \geq 1$.

An operator A is *continuous invertible* if A is invertible with the property that for a constant $c > 0$ and all $\theta_1, \theta_2 \in \Theta$,

$$\|A(\theta_1) - A(\theta_2)\|_{\mathbb{L}} \geq c\|\theta_1 - \theta_2\|.$$

Proof

By the definitions of $\hat{\theta}_n$ and θ_0 and assumption (2),

$$\begin{aligned}\sqrt{n}(\Psi(\hat{\theta}_n) - \Psi(\theta_0)) &= -\sqrt{n}(\Psi_n(\hat{\theta}_n) - \Psi(\hat{\theta}_n)) + o_P(1) \\ &= -\sqrt{n}(\Psi_n - \Psi)(\theta_0) + o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|).\end{aligned} \tag{4}$$

Since $\dot{\Psi}_{\theta_0}$ is continuously invertible, there exists a constant $c > 0$ such that $\|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| \geq c\|\theta - \theta_0\|$ for all θ and θ_0 in $\overline{lin}\Theta$. Combining this with the Frechet differentiability of Ψ yields $\|\Psi(\theta) - \Psi(\theta_0)\| \geq c\|\theta - \theta_0\| + o(\|\theta - \theta_0\|)$. Combining this with above equation, we obtain

$$\sqrt{n}\|\hat{\theta}_n - \theta_0\|(c + o_P(1)) \leq O_P(1) + o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|).$$

Proof

Now we have that $\hat{\theta}_n$ is $\sqrt{n}$-consistent for θ_0 wit respect to $\|\cdot\|$. By the differentiability of Ψ, the left side of (4) can be replaced by $\sqrt{n}\dot{\Psi}_{\theta_0}(\hat{\theta}_n - \theta_0) + o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|)$. And the error terms on both sides is $o_P(1)$. We obtain (3).

Next the continuity of $\dot{\Psi}_{\theta_0}^{-1}$ and the continuous maping theorem yield $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Psi}_{\theta_0}^{-1}(Z)$.

Weak Convergence

The following lemma allows us to weaken the Frechet differentiability requirement to Hadamard differentiability when it is also know that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically tight:

Lemma

Assume the conditions of theorem 2.11 except that consistency of $\hat{\theta}_n$ is strengthened to asymptotic tightness of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and the Frechet differentiability of Ψ is weakened to Hadamard differentiability at θ_0. Then the results of theorem 2.11 still hold.

Proof

The asymptotic tightness of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ enables expression (4) to imply $\sqrt{n}(\Psi(\hat{\theta}_n) - \Psi(\theta_0)) = -\sqrt{n}(\Psi_n - \Psi)(\theta_0) + o_P(1)$. The Hadamard differentiability of Ψ yields $\sqrt{n}(\Psi(\hat{\theta}_n) - \Psi(\theta_0)) = \sqrt{n}\dot{\Psi}_{\theta_0}(\hat{\theta}_n - \theta_0) + o_P(1)$. Combining, we have $\sqrt{n}\dot{\Psi}_{\theta_0}(\hat{\theta}_n - \theta_0) = -\sqrt{n}(\Psi_n - \Psi)(\theta_0) + o_P(1)$.

Weak Convergence

We now consider the special case where the data involved are i.i.d., i.e. $\Psi_n(\theta)(h) = \mathbb{P}_n\psi_{\theta,h}$ amd $\Psi(\theta)(h) = P\psi_{\theta,h}$, for measurable functions $\psi_{\theta,h}$, where h ranges over an index set $\mathcal{H}$. The following lemma gives us reasonably verifiable sufficient conditions for (2) to hold:

Lemma

Suppose the class of functions $\{\psi_{\theta,h} - \psi_{\theta_0,h} : \|\theta - \theta_0\| \leq \delta, h \in \mathcal{H}\}$ *is P-Donsker for some* $\delta > 0$ *and*

$$\sup_{h \in \mathcal{H}} P(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0, \text{ as } \theta \to \theta_0$$

Then if $\Psi_n(\hat{\theta}_n) = o_P(n^{-1/2})$ *and* $\hat{\theta}_n \xrightarrow{P} \theta_0$,

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

Proof

Let $\Theta_\delta \equiv \{\theta : \|\theta - \theta_0\| \leq \delta\}$ and define the extraction function $f : \ell^\infty(\Theta_\delta \times \mathcal{H}) \times \Theta_\delta \mapsto \ell^\infty(\mathcal{H})$ as $f(z,\theta)(h) \equiv z(\theta, h)$, where $z \in \ell^\infty(\Theta_\delta \times \mathcal{H})$. Note that f is continuous at every point (z, θ_1) such that $\sup_{h\in\mathcal{H}} |z(\theta, h) - z(\theta_1, h)| \to 0$ as $\theta \to \theta_1$. Define the stochastic process $Z_n(\theta, h) \equiv \mathbb{G}_n(\psi_{\theta,h} - \psi_{\theta_0,h})$ indexed by $\Theta_\delta \times \mathcal{H}$. As assumed, the process Z_n converges weakly in $\ell^\infty(\Theta_\delta \times \mathcal{H})$ to a tight Gaussian process Z_0 with continuous sample paths with respect to the metric ρ defined by $\rho^2((\theta_1, h_1), (\theta_2, h_2)) = P(\psi_{\theta_1,h_1} - \psi_{\theta_0,h_1} - \psi_{\theta_2,h_2} + \psi_{\theta_0,h_2})^2$. Since, $\sup_{h\in\mathcal{H}} \rho((\theta, h), (\theta_0, h)) \to 0$ by assumption, we have that f is continuous at most all sample paths of Z_0. By Slutsky's theorem, $(Z_n, \hat{\theta}_n) \rightsquigarrow (Z_0, \theta_0)$. The continuous mapping theorem now implies that $Z_n(\hat{\theta}_n) = f(Z_n, \hat{\theta}_n) \rightsquigarrow f(Z_0, \theta_0) = 0$.

Weak Convergence

In addition to the assumptions in above lemma, we are willing to assume

$$\{\psi_{\theta_0,h} : h \in \mathcal{H}\}$$

is P-Donsker, then $\sqrt{n}(\Psi_n - \Psi)(\theta_0) \rightsquigarrow Z$ and all of the weak convergence assumptions of theorem 2.11 are satisfied.

Alternatively, we could just assume that

$$\mathcal{F}_\delta \equiv \{\psi_{\theta,h} : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\}$$

is P-Donsker for some $\delta > 0$.

Weak Convergence

Now, consider the two bootstrapped Z-estimators previously discussed, multinomial bootstrap and multiplier bootstrap. For the multiplier bootstrap we make the additional requirements that $0 < \tau^2 = var(\xi_1) < \infty$ and $\|\xi_1\|_{2,1} \leq \infty$. We use $\underset{\circ}{\overset{P}{\rightsquigarrow}}$ to denote either $\underset{\xi}{\overset{P}{\rightsquigarrow}}$ or $\underset{W}{\overset{P}{\rightsquigarrow}}$ depending on which bootstrap is being used. Let the constant $k_0 = \tau/\mu$ for the multiplier bootstrap and $k_0 = 1$ for the multinomial bootstrap.

Weak Convergence

Theorem

Assume $\Psi(\theta_0) = 0$ *and the following hold:*

(A) $\theta \mapsto \Psi(\theta)$ *is identifiable.*

(B) *The class* $\{\psi_{\theta,h}; \theta \in \Theta, h \in \mathcal{H}\}$ *is P-Glivenko-Cantelli.*

(C) *The class* $\mathcal{F}_\delta \equiv \{\psi_{\theta,h} : h \in \mathcal{H}\}$ *is P-Donsker for some* $\delta > 0$.

(D) $\sup_{h \in \mathcal{H}} P(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0$, *as* $\theta \to \theta_0$, *for some* $\delta > 0$.

(E) $\|\Psi_n(\hat{\theta}_n)\|_{\mathbb{L}} = o_P(n^{-1/2})$ *and* $P(\sqrt{n}\|\Psi_n^\circ(\hat{\theta}_n^\circ)\|_{\mathbb{L}} > \eta | \mathcal{X}_n) = o_P(1)$ *for every* $\eta > 0$.

(F) $\theta \mapsto \Psi(\theta)$ *is Frechet-differentiable at* θ_0 *with continuously invertible derivative* $\dot{\Psi}_{\theta_0}$.

Then $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Psi}_{\theta_0}^{-1} Z$, *where* $Z \in \ell^\infty(\mathcal{H})$ *is the tight, mean zero Gaussian limiting distribution of* $\sqrt{n}(\Psi_n - \Psi)(\theta_0)$, *and* $\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) \underset{\circ}{\overset{P}{\rightsquigarrow}} k_0 Z$

Weak Convergence

Condition (A) is identifiability. Condition (B) and (C) are consistency and asymptotic normality conditions for the estimating equation. Condition (D) is an asymptotic equicontinuity condition for the estimating equation at θ_0. Condition (E) simply states that the estimators are approximate zeros of the estimating equation, while condition (F) specifies the smoothness and invertibility requirements of the derivative of Ψ.

Example: Right-censored Kaplan-Meier Estimator

In Section 2.2.5, right-censored Kaplan-Meier estimator is shown to be a Z-estimator with a certain estimating equation $\Psi_n(\theta) = \mathbb{P}_n \psi_\theta(t)$, where

$$\psi_\theta(t) = 1\{U > t\} + (1 - \delta)1\{U \leq t\}1\{\theta(U) > 0\}\frac{\theta(t)}{\theta(U)} - \theta(t)$$

where U, δ is the observed right-censored survival time and censoring indicator. The limiting estimating function $\Psi(\theta) = P\psi_\theta$ is

$$\Psi(\theta)(t) = \theta_0(t)L(t) + \int_0^t \frac{\theta_0(u)}{\theta(u)} dG(u)\theta(t) - \theta(t)$$

where G is the censoring distribution and $L = 1 - G$. t and $[0, \tau]$ play the roles of h and $\mathcal{H}$.

Example: Right-censored Kaplan-Meier Estimator

Verify identifiablity: let

$$\epsilon_n(t) = \frac{\theta_0(t)}{\theta_n(t)} - 1$$

Then $\Psi(\theta_n)(t) \to 0$ uniformly over $t \in [0, \tau]$ implies that

$$u_n(t) = \epsilon_n(t)L(t) + \int_0^t \epsilon_n(u)dG(u) \to 0$$

uniformly over $[0, \tau]$.

By solving this integral equation, we obatin

$$\epsilon_n(t) = u_n(0) + \int_0^t \frac{du_n(s)}{L(s-)}$$

which implies $\epsilon_n(t) \to 0$ uniformly. Thus, $\|\theta_n - \theta_0\|_\infty \to 0$.

Example: Right-censored Kaplan-Meier Estimator

Condition (B), (C) and (F) were established in Chapter 2.

As $\|\theta - \theta_0\|_\infty \to 0$, we can directly verify condition (D) by

$$\sup_{h \in \mathcal{H}} P(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0$$

If $\hat{\theta}_n$ is Kaplen-Meier estimator, then $\|\Psi_n(\hat{\theta}_n)(t)\|_\infty = 0$ almost surely.

If the bootstrapped version is

$$\hat{\theta}_n^\circ(t) = \prod_{j:\tilde{T}_j \le t} (1 - \frac{n\mathbb{P}_n^\circ[\delta 1\{U = \tilde{T}_j\}]}{n\mathbb{P}_n^\circ[1\{U \ge \tilde{T}_j\}]})$$

where $\tilde{T}_j$ are observed failure times, then $\|\Psi_n^\circ(\hat{\theta}_n^\circ)\|_\infty = 0$ almost surely.

Then we can obtain consistency, weak convergence, and bootstrap consistency for the Kaplan-Meier estimator.

Using the Delta Method

An alternative approach to Z-estimators is to view the extraction of the zero from the estimating equation as a continuous mapping.
Let Θ be the subset of a Banach space and $\mathbb{L}$ to be a Banach space. Let $\ell^\infty(\Theta, \mathbb{L})$ to be the Banach space of all uniformly norm-bounded functions $z : \Theta \mapsto \mathbb{L}$. Let $Z(\Theta, \mathbb{L})$ be the subset consisting of all maps with at least one zero, and let $\Phi(\Theta, \mathbb{L})$ be the collection of all maps $\phi : Z(\Theta, \mathbb{L}) \mapsto \Theta$ that for each element $z \in Z(\Theta, \mathbb{L})$ extract one of its zeros $\phi(z)$. This structure allows for multiple zeros. Define $\ell_0^\infty(\Theta, \mathbb{L})$ to be the elements $z \in \ell^\infty(\Theta, \mathbb{L})$ for which $\|z(\theta) - z(\theta_0)\|_{\mathbb{L}} \to 0$ as $\theta \to \theta_0$.

Using the Delta Method

Theorem (13.5)

Assume $\Psi : \Theta \to \mathbb{L}$ is uniformly norm-bounded over Θ, $\Psi(\theta_0) = 0$, and identifiable. Let Ψ also be Frechet differentiable at θ_0 with continuously invertible derivative $\dot{\Psi}_{\theta_0}$. Then the continuous linear operator $\phi'_\Psi : \ell_0^\infty(\Theta, \mathbb{L}) \mapsto lin\Theta$ defined by $z \mapsto \phi'_\Psi(z) \equiv -\dot{\Psi}_{\theta_0}^{-1}(z(\theta_0))$ satisfies:

$$\sup_{\phi \in \Phi(\Theta, \mathbb{L})} \| \frac{\phi(\Psi + t_n z_n) - \phi(\Psi)}{t_n} - \phi'_\Psi(z(\theta_0)) \| \to 0$$

as $n \to \infty$, for any sequences $(t_n, z_n) \in (0, \infty) \times \ell^\infty(\Theta, \mathbb{L})$ such that $t_n \downarrow 0$, $\Psi + t_n z_n \in Z(\Theta, \mathbb{L})$, and $z_n \to z \in \ell_0^\infty(\Theta, \mathbb{L})$.

Proof

Let $0 < t_n \downarrow 0$ and $z_n \to z \in \ell_0^\infty(\Theta, \mathbb{L})$ such that $\Psi + t_n z_n \in Z(\Theta, \mathbb{L})$. Choose any sequence $\{\phi_n\} \in \Phi(\Theta, \mathbb{L})$, and note that $\theta_n \equiv \phi_n(\Psi + t_n Z_n)$ satisfies $\Psi(\theta_n) + t_n z_n = 0$ by construction. Hence $\Psi(\theta_n) = O(t_n)$. By indentifiability, $\theta_n \to \theta_0$. By the Frechet differentiability of Ψ,

$$\liminf_{n\to\infty} \frac{\|\Psi(\theta_n) - \Psi(\theta_0)\|_{\mathbb{L}}}{\|\theta_n - \theta_0\|} \geq \liminf_{n\to\infty} \frac{\|\dot{\Psi}_{\theta_0}(\theta_n - \theta_0)\|_{\mathbb{L}}}{\|\theta_n - \theta_0\|} \geq \inf_{\|g\|=1} \|\dot{\Psi}_{\theta_0}(g)\|_{\mathbb{L}}$$

where g ranges over $lin\Theta$. Since the inverse of $\dot{\Psi}_{\theta_0}$ is continuous, the right side of the above is positive. Thus there exists a universal constant $c < \infty$ for which $\|\theta_n - \theta_0\| < c\|\Psi(\theta_n) - \Psi(\theta_0)\|_{\mathbb{L}} = c\|t_n z_n(\theta_n)\|_{\mathbb{L}}$ for all n large enough.

Proof

Hence $\|\theta_n - \theta_0\| = O(t_n)$. By Frechet differentiability, $\Psi(\theta_n) - \Psi(\theta_0) = \dot{\Psi}_{\theta_0}(\theta_n - \theta_0) + o(\|\theta_n - \theta_0\|)$, where $\dot{\Psi}_{\theta_0}$ is linear and continuous on $lin\Theta$. The remainder term is $o(t_n)$ by previous arguments. Combining this with the fact that $t_n^{-1}(\Psi(\theta_n) - \Psi(\theta_0)) = -z_n(\theta_n) \to z(\theta_0)$, we obtain

$$\frac{\theta_n - \theta}{t_n} = \dot{\Psi}_{\theta_0}^{-1}(\frac{\Psi(\theta_n) - \Psi(\theta_0)}{t_n} + o(1)) \to -\dot{\Psi}_{\theta_0}^{-1}(z(\theta_0))$$

The conclusion now follows since the sequence ϕ_n was arbitrary.

Using the Delta Method

The following simple corollary allows the delta method to be applied to Z-estimators. Let $\tilde{\phi} : \ell^\infty(\Theta, \mathbb{L}) \mapsto \Theta$ be a map such that for each $x \in \ell^\infty(\Theta, \mathbb{L}), \tilde{\phi}(x) = \theta_1 \neq \theta_0$ when $x \notin Z(\Theta, \mathbb{L})$ and $\tilde{\phi}(x) = \phi(x)$ for some $\phi \in \Phi(\Theta, \mathbb{L})$ otherwise.

Corollary

Suppose Ψ satisfies the conditions of theorem 13.5, $\hat{\theta}_n = \tilde{\phi}(\Psi_n)$, and Ψ_n has at least one zero for all n large enough, outer almost surely. Suppose also that $r_n(\Psi_n - \Psi) \rightsquigarrow X$ in $\ell^\infty(\Theta, \mathbb{L})$, with X tight and $\|X(\theta)\|_{\mathbb{L}} \to 0$ as $\theta \to \theta_0$ almost surely. Then $r_n(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Psi}_{\theta_0}^{-1} X(\theta_0)$.

Using the Delta Method

A drawback with this approach is that root finding algorithms in practice are seldom exact, and room needs to be allowed for computational error. The following corollary yields a very general Z-estimator result based on a modified delta method. We make the fairly realistic assumption that the Z-estimator $\hat{\theta}_n$ is computed from Ψ_n using a deterministic algorithm (e.g., a computer program) that is allowed to depend on n and which is not required to yield an exact root of Ψ_n.

Using the Delta Method

Corollary

Suppose Ψ *satisfies the conditions of theorem 13.5, and* $\hat{\theta}_n = A_n(\Psi_n)$ *for some sequence of deterministic algorithms* $A_n : \ell^\infty(\Theta, \mathbb{L}) \mapsto \Theta$ *and random sequence* $\Psi_n : \Theta \mapsto \mathbb{L}$ *of estimating equations such that* $\Psi_n \xrightarrow{P} \Psi$ *in* $\ell^\infty(\Theta, \mathbb{L})$ *and* $\Psi_n(\hat{\theta}_n) = o_P(r_n^{-1})$ *, where* $0 < r_n \to \infty$ *is a sequence of constants for which* $r_n(\Psi_n - \Psi) \rightsquigarrow X$ *in* $\ell^\infty(\Theta_0, \mathbb{L})$ *for some closed* $\Theta_0 \subset \Theta$ *containing an open neighborhood of* θ_0*, with* X *tight and* $\|X(\theta)\|_{\mathbb{L}} \to 0$ *as* $\theta \to \theta_0$ *almost surely. Then* $r_n(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Psi}_{\theta_0}^{-1} X(\theta_0)$.

Proof

Let $X_n \equiv r_n(\Psi_n - \Psi)$. By theorem 7.26, there exists a new probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ on which: $E^* f(\tilde{\Psi}_n) = E^* f(\Psi_n)$ for all bounded $f : \ell^\infty(\Theta, \mathbb{L}) \mapsto \mathbb{R}$ and all $n \geq 1$; $\tilde{\Psi}_n \overset{as*}{\to} \Psi$ in $\ell^\infty(\Theta, \mathbb{L})$; $r_n(\tilde{\Psi}_n - \Psi) \overset{as*}{\to} \tilde{X}$ in $\ell^\infty(\Theta_0, \mathbb{L})$; $\tilde{X}$ and X have the same distributions; and $r_n(A_n(\tilde{\Psi}_n) - \theta_0)$ and $r_n(A_n(\Psi_n) - \theta_0)$ have the same distributions. Note that for any bounded f, $\tilde{\Psi}_n \mapsto f(\tilde{\Psi}_n(A_n(\tilde{\Psi}_n))) = g(\tilde{\Psi}_n)$ for some bounded g. Thus $\tilde{\Psi}_n(\tilde{\theta}_n) = o_{\tilde{P}}(r_n^{-1})$ for $\tilde{\theta}_n \equiv A_n(\tilde{\Psi}_n)$.

Proof

Hence for any subsequence n' there exists a further subsequence n'' such that $\tilde{\Psi}_{n''} \stackrel{as*}{\to} \Psi$ in $\ell^\infty(\Theta, \mathbb{L})$, $r_{n''}(\tilde{\Psi}_{n''} - \Psi) \stackrel{as*}{\to} \tilde{X}$ in $\ell^\infty(\Theta_0, \mathbb{L})$, and $\tilde{\Psi}_{n''}(\tilde{\theta}_{n''}) \stackrel{as*}{\to} 0$ in $\mathbb{L}$. Thus also $\Psi(\tilde{\theta}_{n''}) \stackrel{as*}{\to} 0$, which implies $\tilde{\theta}_{n''} \stackrel{as*}{\to} \theta_0$. Note that $\tilde{\theta}_{n''}$ is a zero of $\tilde{\Psi}_{n''}(\theta) - \tilde{\Psi}_{n''}(\tilde{\theta}_{n''})$ by definition and is contained in Θ_0 for all n large enough. Hence, for all n large enough, $r_{n''}(\tilde{\theta}_{n''} - \theta_0) = r_{n''}(\phi_{n''}(\tilde{\Psi}_{n''} - \tilde{\Psi}_{n''}(\tilde{\theta}_{n''})) - \phi_{n''}(\Psi))$ for some sequence $\phi_n \in \Phi(\Theta_0, \mathbb{L})$ possibly dependent on sample realization $\tilde{\omega} \in \tilde{\Omega}$. This implies that for all n large enough, by theorem 13.5,

$$
\begin{aligned}
&\|r_{n''}(\tilde{\theta}_{n''} - \theta_0) - \phi'_{\Psi}(\tilde{X})\| \\
\leq &\sup_{\phi \in \Phi(\Theta_0, \mathbb{L})} |r_{n''}(\phi(\tilde{\Psi}_{n''} - \tilde{\Psi}_{n''}(\tilde{\theta}_{n''})) - \phi(\Psi)) - \phi'_{\Psi}(\tilde{X})| \\
\stackrel{as*}{\to} &0.
\end{aligned}
$$

Proof

This implies $\|r_{n''}(A_{n''}(\tilde{\Psi}_{n''}) - \theta_0) - \phi'_{\Psi}(\tilde{X})\| \overset{as*}{\to} 0$. Since this hold for every subsequence, we have $r_n(A_n(\tilde{\Psi}_n) - \theta_0) \rightsquigarrow \phi'_{\Psi}(\tilde{X})$. This implies $r_n(A_n(\Psi_n) - \theta_0) \rightsquigarrow \phi'_{\Psi}(X)$.

Using the Delta Method

The following corollary extends the previous result to generalized bootstrapped processes. Let Ψ_n° be a bootstrapped version of Ψ_n based on both the data sequence X_n and a sequence of weights $W = \{W_n, n \geq 1\}$.

Corollary

Assume the conditions of previous corollary and, in addition that $\hat{\theta}_n^\circ = A_n(\Psi_n^\circ)$ for a sequence of bootstrapped estimating equations $\Psi_n^\circ(X_n, W_n)$, with $\Psi_n^\circ - \Psi \underset{W}{\overset{P}{\rightsquigarrow}} 0$ and $r_n \Psi_n^\circ(\hat{\theta}_n^\circ) \underset{W}{\overset{P}{\rightsquigarrow}} 0$ in $\ell^\infty(\Theta, \mathbb{L})$, and with $r_n c(\Psi_n^\circ - \Psi) \underset{W}{\overset{P}{\rightsquigarrow}} X$ in $\ell^\infty(\Theta_0, \mathbb{L})$ for some $0 < c < \infty$, where the maps $W_n \mapsto h(\Psi_n^\circ)$ are measurable for every $h \in C_b(\ell^\infty(\Theta, \mathbb{L}))$ outer almost surely. Then $r_n c(\hat{\theta}_n^\circ - \hat{\theta}_n) \underset{W}{\overset{P}{\rightsquigarrow}} \phi'_\Psi(X)$.