Chapter 3.3: Z-Estimators

Rens Kamphuis

University of Amsterdam

July 6, 2020

Overview

M-Estimators Recap

Consider the criterion function $\theta \mapsto M_n(\theta)$.

M-Estimators Recap

Consider the criterion function $\theta \mapsto M_n(\theta)$.

Estimators $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ that maximise a certain criterion function are called M-estimators.

M-Estimators Recap

Consider the criterion function $\theta \mapsto M_n(\theta)$.

Estimators $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ that maximise a certain criterion function are called M-estimators.

We have seen results that ensure:

- Consistency: $\hat{\theta}_n \xrightarrow{P^*} \theta_0$
- Rate of convergence: $r_n d(\hat{\theta}_n, \theta_0) = O_P^*(1)$
- Limiting distribution

Z-Estimators

We again consider a criterion function

$$\Psi_n : \Theta \to \mathbb{L},$$

where Θ is a subset of a Banach space and $\mathbb{L}$ is another Banach space.

Z-Estimators

We again consider a criterion function

$$\Psi_n : \Theta \to \mathbb{L},$$

where Θ is a subset of a Banach space and $\mathbb{L}$ is another Banach space.

The estimators $\hat{\theta}_n$ solve

$$\Psi_n(\hat{\theta}_n) = 0.$$

Z-Estimators

We again consider a criterion function

$$\Psi_n : \Theta \to \mathbb{L},$$

where Θ is a subset of a Banach space and $\mathbb{L}$ is another Banach space.

The estimators $\hat{\theta}_n$ solve

$$\Psi_n(\hat{\theta}_n) = 0.$$

Let $\Psi : \Theta \to \mathbb{L}$ be the deterministic limiting process of Ψ_n. It is again reasonable to assume

$$\Psi_n(\theta) \stackrel{P^*}{\to} \Psi(\theta) \quad \text{for every } \theta.$$

Z-Estimators

We again consider a criterion function

$$\Psi_n : \Theta \to \mathbb{L},$$

where Θ is a subset of a Banach space and $\mathbb{L}$ is another Banach space.

The estimators $\hat{\theta}_n$ solve

$$\Psi_n(\hat{\theta}_n) = 0.$$

Let $\Psi : \Theta \to \mathbb{L}$ be the deterministic limiting process of Ψ_n. It is again reasonable to assume

$$\Psi_n(\theta) \xrightarrow{P^*} \Psi(\theta) \quad \text{for every } \theta.$$

We hope that $\hat{\theta}_n$ tends to a value θ_0 satisfying

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

Z-Estimators

We are interested in the same questions as last week. Namely,

- Do we have consistency?
- What is the rate of convergence?
- What is the limiting distribution?

Z-Estimators

We are interested in the same questions as last week. Namely,

- Do we have consistency?
- What is the rate of convergence?
- What is the limiting distribution?

Fortunately, there is a very direct relationship between M-estimators and Z-estimators.

A zero of Ψ_n maximises the function $\theta \mapsto -\|\Psi_n(\theta)\|$.

Z-Estimators

We are interested in the same questions as last week. Namely,

- Do we have consistency?
- What is the rate of convergence?
- What is the limiting distribution?

Fortunately, there is a very direct relationship between M-estimators and Z-estimators.

A zero of Ψ_n maximises the function $\theta \mapsto -\|\Psi_n(\theta)\|$.

Results from last week allow us to derive similar results for Z-estimators very easily! Therefore, we will only be looking at the limiting distribution.

Fréchet Derivative

Recall that $\Psi : \Theta \to \mathbb{L}$, where Θ is a subset of a Banach space and $\mathbb{L}$ is another Banach space.

Fréchet Derivative

Recall that $\Psi : \Theta \to \mathbb{L}$, where Θ is a subset of a Banach space and $\mathbb{L}$ is another Banach space. We assume that Ψ is Fréchet-differentiable at θ_0. That is,

$$\|\Psi(\theta) - \Psi(\theta_0) - \dot{\Psi}_{\theta_0}(\theta - \theta_0)\| = o(\|\theta - \theta_0\|) \quad \text{as } \theta \to \theta_0,$$

where $\dot{\Psi}_{\theta_0} : \operatorname{lin} \Theta \to \mathbb{L}$ is a continuous, linear, and one-to-one map.

Fréchet Derivative

Recall that $\Psi : \Theta \to \mathbb{L}$, where Θ is a subset of a Banach space and $\mathbb{L}$ is another Banach space. We assume that Ψ is Fréchet-differentiable at θ_0. That is,

$$\|\Psi(\theta) - \Psi(\theta_0) - \dot{\Psi}_{\theta_0}(\theta - \theta_0)\| = o(\|\theta - \theta_0\|) \quad \text{as } \theta \to \theta_0,$$

where $\dot{\Psi}_{\theta_0} : \operatorname{lin} \Theta \to \mathbb{L}$ is a continuous, linear, and one-to-one map.

In the following theorem we will make the assumption that $\dot{\Psi}_{\theta_0}^{-1}$ exists and is continuous on the range of $\dot{\Psi}_{\theta_0}$.

Fréchet Derivative

Recall that $\Psi : \Theta \to \mathbb{L}$, where Θ is a subset of a Banach space and $\mathbb{L}$ is another Banach space. We assume that Ψ is Fréchet-differentiable at θ_0. That is,

$$\|\Psi(\theta) - \Psi(\theta_0) - \dot{\Psi}_{\theta_0}(\theta - \theta_0)\| = o(\|\theta - \theta_0\|) \quad \text{as } \theta \to \theta_0,$$

where $\dot{\Psi}_{\theta_0} : \text{lin}\ \Theta \to \mathbb{L}$ is a continuous, linear, and one-to-one map.

In the following theorem we will make the assumption that $\dot{\Psi}_{\theta_0}^{-1}$ exists and is continuous on the range of $\dot{\Psi}_{\theta_0}$.

Remark: this assumption is easy to verify if Θ were finite-dimensional.

Limiting Distribution

Theorem (3.3.1)

Assume that

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

Limiting Distribution

Theorem (3.3.1)

Assume that

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

and that the sequence $\sqrt{n}(\Psi_n - \Psi)(\theta_0)$ *converges in distribution to a tight random element* Z.

Limiting Distribution

Theorem (3.3.1)

Assume that

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

and that the sequence $\sqrt{n}(\Psi_n - \Psi)(\theta_0)$ *converges in distribution to a tight random element* Z*. If* $\Psi(\theta_0) = 0$ *and* $\Psi_n(\hat{\theta}_n) = o_P^*(n^{-1/2})$*, and* $\hat{\theta}_n \xrightarrow{P^*} \theta_0$*, then*

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

Limiting Distribution

Theorem (3.3.1)

Assume that

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

and that the sequence $\sqrt{n}(\Psi_n - \Psi)(\theta_0)$ *converges in distribution to a tight random element* Z*.If* $\Psi(\theta_0) = 0$ *and* $\Psi_n(\hat{\theta}_n) = o_P^*(n^{-1/2})$*, and* $\hat{\theta}_n \xrightarrow{P^*} \theta_0$*, then*

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

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

Theorem 3.3.1

Since $\Psi(\theta_0) = 0$ and $\Psi_n(\hat{\theta}_n) = o_P^*(n^{-1/2})$, we have

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

Theorem 3.3.1

Since $\Psi(\theta_0) = 0$ and $\Psi_n(\hat{\theta}_n) = o_P^*(n^{-1/2})$, we have

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

Reshuffling terms in the first assumption gives

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

Theorem 3.3.1

Since $\Psi(\theta_0) = 0$ and $\Psi_n(\hat{\theta}_n) = o_P^*(n^{-1/2})$, we have

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

Reshuffling terms in the first assumption gives

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

Since $\dot{\Psi}_{\theta_0}$ is continuously invertible, there exists $c > 0$ such that

$$\begin{aligned} \|\theta - \theta_0\| = \|\dot{\Psi}_{\theta_0}^{-1}\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| &\leq \frac{1}{c}\|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| \\ \implies \|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| &\geq c\|\theta - \theta_0\|. \end{aligned}$$

Theorem 3.3.1

Since $\Psi(\theta_0) = 0$ and $\Psi_n(\hat{\theta}_n) = o_P^*(n^{-1/2})$, we have

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

Reshuffling terms in the first assumption gives

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

Since $\dot{\Psi}_{\theta_0}$ is continuously invertible, there exists $c > 0$ such that

$$\begin{aligned}\|\theta - \theta_0\| = \|\dot{\Psi}_{\theta_0}^{-1}\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| &\leq \frac{1}{c}\|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| \\ &\implies \|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| \geq c\|\theta - \theta_0\|.\end{aligned}$$

Using the differentiability of Ψ then gives

$$\|\Psi(\theta) - \Psi(\theta_0)\| \geq c\|\theta - \theta_0\| + o(\|\theta - \theta_0\|).$$

Theorem 3.3.1

Since $\Psi(\theta_0) = 0$ and $\Psi_n(\hat{\theta}_n) = o_P^*(n^{-1/2})$, we have

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

Reshuffling terms in the first assumption gives

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

Since $\dot{\Psi}_{\theta_0}$ is continuously invertible, there exists $c > 0$ such that

$$\|\theta - \theta_0\| = \|\dot{\Psi}_{\theta_0}^{-1}\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| \leq \frac{1}{c}\|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\|$$
$$\implies \|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| \geq c\|\theta - \theta_0\|.$$

Using the differentiability of Ψ then gives

$$\|\Psi(\theta) - \Psi(\theta_0)\| \geq c\|\theta - \theta_0\| + o(\|\theta - \theta_0\|).$$

Combining now gives

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

Theorem 3.3.1

Since $\Psi(\theta_0) = 0$ and $\Psi_n(\hat{\theta}_n) = o_P^*(n^{-1/2})$, we have

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

Reshuffling terms in the first assumption gives

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

Since $\dot{\Psi}_{\theta_0}$ is continuously invertible, there exists $c > 0$ such that

$$\|\theta - \theta_0\| = \|\dot{\Psi}_{\theta_0}^{-1}\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| \leq \frac{1}{c}\|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\|$$
$$\implies \|\dot{\Psi}_{\theta_0}(\theta - \theta_0)\| \geq c\|\theta - \theta_0\|.$$

Using the differentiability of Ψ then gives

$$\|\Psi(\theta) - \Psi(\theta_0)\| \geq c\|\theta - \theta_0\| + o(\|\theta - \theta_0\|).$$

Combining now gives

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

and we conclude that $\hat{\theta}_n$ is $\sqrt{n}$-consistent for θ_0 in norm.

Theorem 3.3.1

The differentiability of Ψ allows us to write

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

Note that the last term is $o_P^*(1)$.

Theorem 3.3.1

The differentiability of Ψ allows us to write

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

Note that the last term is $o_P^*(1)$. Recall that

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

Theorem 3.3.1

The differentiability of Ψ allows us to write

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

Note that the last term is $o_P^*(1)$. Recall that

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

We conclude that indeed

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

Theorem 3.3.1

The differentiability of Ψ allows us to write

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

Note that the last term is $o_P^*(1)$. Recall that

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

We conclude that indeed

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

Using continuity of $\dot{\Psi}_{\theta_0}^{-1}$, the continuous mapping theorem gives

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

Remark: if $\sqrt{n}\|\hat{\theta}_n - \theta_0\|$ were asymptotically tight, the first conclusion is valid without without requiring continuous invertibility of $\dot{\Psi}_{\theta_0}$.

The i.i.d. case

Recall that we are looking for the zero of

$$\Psi_n : \Theta \to \mathbb{L}.$$

The i.i.d. case

Recall that we are looking for the zero of

$$\Psi_n : \Theta \to \mathbb{L}.$$

If $\mathbb{L}$ is an $\ell^\infty(\mathcal{H})$ space, then $\hat{\theta}_n$ is the zero of Ψ_n if it solves the system of estimating equations

$$\Psi_n(\hat{\theta}_n)h = 0 \text{ for all } h \in \mathcal{H}.$$

The i.i.d. case

Recall that we are looking for the zero of

$$\Psi_n : \Theta \to \mathbb{L}.$$

If $\mathbb{L}$ is an $\ell^\infty(\mathcal{H})$ space, then $\hat{\theta}_n$ is the zero of Ψ_n if it solves the system of estimating equations

$$\Psi_n(\hat{\theta}_n)h = 0 \text{ for all } h \in \mathcal{H}.$$

In case of i.i.d. observations, we may use $\Psi_n(\theta)h = \mathbb{P}_n\psi_{\theta,h}$ and $\Psi(\theta)h = \mathbb{P}\psi_{\theta,h}$ for given measurable functions $\psi_{\theta,h}$.

The i.i.d. case

Recall that we are looking for the zero of

$$\Psi_n : \Theta \to \mathbb{L}.$$

If $\mathbb{L}$ is an $\ell^\infty(\mathcal{H})$ space, then $\hat{\theta}_n$ is the zero of Ψ_n if it solves the system of estimating equations

$$\Psi_n(\hat{\theta}_n)h = 0 \text{ for all } h \in \mathcal{H}.$$

In case of i.i.d. observations, we may use $\Psi_n(\theta)h = \mathbb{P}_n\psi_{\theta,h}$ and $\Psi(\theta)h = \mathbb{P}\psi_{\theta,h}$ for given measurable functions $\psi_{\theta,h}$.

The stochastic condition reduces to

$$\|\mathbb{G}_n(\psi_{\hat{\theta}_n,h} - \psi_{\theta_0,h})\|_{\mathcal{H}} = o_P^*(1 + \sqrt{n}\|\hat{\theta}_n - \theta\|).$$

The stochastic condition revisited

Lemma (3.3.5)

Suppose the class of functions

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

is P-Donsker for some $\delta > 0$ and that

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

If $\hat{\theta}_n \xrightarrow{P^*} \theta_0$, then the stochastic condition in Theorem 3.3.1 is satisfied.

Lemma 3.3.5

Wlog, $\hat{\theta}_n$ takes its values in $\Theta_\delta = \{\theta : \|\theta - \theta_0\| < \delta\}$.

Lemma 3.3.5

Wlog, $\hat{\theta}_n$ takes its values in $\Theta_\delta = \{\theta : \|\theta - \theta_0\| < \delta\}$.
Define $f : \ell^\infty(\Theta_\delta \times \mathcal{H}) \times \Theta_\delta \to \ell^\infty(\mathcal{H})$ by $f(z, \theta)h = z(\theta, h)$. This function is continuous at (z, θ_0) such that

$$\|z(\theta, h) - z(\theta_0, h)\|_{\mathcal{H}} \to 0 \quad \text{as } \theta \to \theta_0.$$

Lemma 3.3.5

Wlog, $\hat{\theta}_n$ takes its values in $\Theta_\delta = \{\theta : \|\theta - \theta_0\| < \delta\}$.
Define $f : \ell^\infty(\Theta_\delta \times \mathcal{H}) \times \Theta_\delta \to \ell^\infty(\mathcal{H})$ by $f(z, \theta)h = z(\theta, h)$. This function is continuous at (z, θ_0) such that

$$\|z(\theta, h) - z(\theta_0, h)\|_{\mathcal{H}} \to 0 \quad \text{as } \theta \to \theta_0.$$

Define a stochastic process Z_n indexed by $\Theta_\delta \times \mathcal{H}$ by

$$Z_n(\theta, h) = \mathbb{G}_n(\psi_{\theta,h} - \psi_{\theta_0,h}).$$

This sequence converges in $\ell^\infty(\Theta_\delta \times \mathcal{H})$ to a tight Gaussian process Z.

Lemma 3.3.5

Wlog, $\hat{\theta}_n$ takes its values in $\Theta_\delta = \{\theta : \|\theta - \theta_0\| < \delta\}$.
Define $f : \ell^\infty(\Theta_\delta \times \mathcal{H}) \times \Theta_\delta \to \ell^\infty(\mathcal{H})$ by $f(z, \theta)h = z(\theta, h)$. This function is continuous at (z, θ_0) such that

$$\|z(\theta, h) - z(\theta_0, h)\|_{\mathcal{H}} \to 0 \quad \text{as } \theta \to \theta_0.$$

Define a stochastic process Z_n indexed by $\Theta_\delta \times \mathcal{H}$ by

$$Z_n(\theta, h) = \mathbb{G}_n(\psi_{\theta,h} - \psi_{\theta_0,h}).$$

This sequence converges in $\ell^\infty(\Theta_\delta \times \mathcal{H})$ to a tight Gaussian process Z. Moreover, it has continuous samples paths with respect to the semimetric ρ given by

$$\rho^2((\theta_1, h_1), (\theta_2, h_2)) = P(\psi_{\theta_1,h_1} - \psi_{\theta_0,h_1} - \psi_{\theta_2,h_2} + \psi_{\theta_0,h_2})^2.$$

Lemma 3.3.5

By assumption

$$\sup_h \rho^2((\theta, h), (\theta_0, h)) = \sup_h P(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0 \quad \text{as } \theta \to \theta_0.$$

It follows that f is continuous at almost all sample paths of Z.

Lemma 3.3.5

By assumption

$$\sup_h \rho^2((\theta, h), (\theta_0, h)) = \sup_h P(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0 \quad \text{as } \theta \to \theta_0.$$

It follows that f is continuous at almost all sample paths of Z.

By Slutsky's lemma (Example 1.4.7), $(Z_n, \hat{\theta}_n) \rightsquigarrow (Z, \theta_0)$.

Lemma 3.3.5

By assumption

$$\sup_h \rho^2((\theta, h), (\theta_0, h)) = \sup_h P(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0 \quad \text{as } \theta \to \theta_0.$$

It follows that f is continuous at almost all sample paths of Z.

By Slutsky's lemma (Example 1.4.7), $(Z_n, \hat{\theta}_n) \rightsquigarrow (Z, \theta_0)$.

The continuous mapping theorem then gives

$$Z_n(\hat{\theta}_n) = f(Z_n, \hat{\theta}_n) \rightsquigarrow f(Z, \theta_0) = 0 \text{ in } \ell^\infty(\mathcal{H}).$$

Examples

In case of Euclidean Θ and i.i.d. random variables, the following examples satisfy the stochastic condition:

Example 3.3.7 (Lipschitz): For every θ_1, θ_2 in a neighbourhood of θ_0,

$$\|\psi_{\theta_1}(x) - \psi_{\theta_2}(x)\| \leq \dot{\psi}(x)\|\theta_1 - \theta_2\|,$$

where $P\dot{\psi}^2 < \infty$.

Examples

In case of Euclidean Θ and i.i.d. random variables, the following examples satisfy the stochastic condition:

Example 3.3.7 (Lipschitz): For every θ_1, θ_2 in a neighbourhood of θ_0,

$$\|\psi_{\theta_1}(x) - \psi_{\theta_2}(x)\| \leq \dot{\psi}(x)\|\theta_1 - \theta_2\|,$$

where $P\dot{\psi}^2 < \infty$.

Example 3.3.8 (Classical smoothness): Assume $\theta \mapsto \psi_\theta(x)$ is twice continuously differentiable for each x, with derivatives satisfying

$$P\|\dot{\psi}_{\theta_0}\|^2 < \infty; \qquad P^* \sup_{\|\theta-\theta_0\|<\delta} \|\ddot{\psi}_\theta\| < \infty.$$

Idea: Taylor $\mathbb{G}_n(\psi_\theta - \psi_{\theta_0})$ to show that

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