Introduction to Empirical Processes and Semiparametric Inference Lecture 17: Z-Estimators

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Z-Estimators

Recall from Section 2.2.5 that Z-estimators are approximate zeros of data-dependent functions.

These data-dependent functions, denoted Ψ_n, are maps between a possibly infinite dimensional normed parameter space Θ and a normed space $\mathbb{L}$, where the respective norms are $\|\cdot\|$ and $\|\cdot\|_{\mathbb{L}}$.

The Ψ_n are frequently called estimating equations.

A quantity $\hat{\theta}_n \in \Theta$ is a Z-estimator if

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

In this chapter, we extend and prove the results of Section 2.2.5.

As part of this, we extend the Z-estimator master theorem, Theorem 10.16, to the infinite dimensional case, divided into two parts: consistency and weak convergence.

We first discuss consistency and present a Z-estimator master theorem for consistency.

We then discuss weak convergence and examine closely the special case of Z-estimators which are empirical measures of Donsker classes.

We then use this structure to develop a Z-estimator master theorem for weak convergence.

Both master theorems, the one for consistency and the one for weak convergence, will include results for the bootstrap.

Finally, we demonstrate how Z-estimators can be viewed as Hadamard differentiable functionals of the involved estimating equations and how this structure enables use of a modified delta method to obtain very general results for Z-estimators.

Recall from Section 2.2.5 that the Kaplan-Meier estimator is an important and instructive example of a Z-estimator.

A more sophisticated example, which will be presented later in the case studies of Chapter 15, is the nonparametric maximum likelihood estimator for the proportional odds survival model.

Consistency

The main consistency result we have already presented in Theorem 2.10 of Section 2.2.5, and the proof of this theorem was given as an exercise (Exercise 2.4.2).

We will now extend this result to the bootstrapped Z-estimator.

First, we restate the identifiability condition of Theorem 2.10:

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

$$\|\Psi(\theta_n)\|_{\mathbb{L}} \to 0 \text{ implies } \|\theta_n - \theta_0\| \to 0 \text{ for any } \{\theta_n\} \in \Theta. \qquad (1)$$

Note that there are alternative identifiability conditions that will also work, including the stronger condition that both $\Psi(\theta_0) = 0$ and $\Psi : \Theta \mapsto \mathbb{L}$ be one-to-one.

Nevertheless, Condition (1) seems to be the most efficient for our purposes.

In what follows, we will use the bootstrap-weighted empirical process $\mathbb{P}_n^\circ$ to denote either the nonparametric bootstrapped empirical process (with multinomial weights) or the multiplier bootstrapped empirical process defined by

$$f \mapsto \mathbb{P}_n^\circ f = n^{-1} \sum_{i=1}^n (\xi_i/\bar{\xi}) f(X_i),$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. positive weights with $0 < \mu = \mathrm{E}\xi_1 < \infty$ and

$$\bar{\xi} = n^{-1} \sum_{i=1}^n \xi_i.$$

Note that this is a special case of the weighted bootstrap introduced in Theorem 10.13 but with the addition of $\bar{\xi}$ in the denominator.

We leave it as an exercise (Exercise 13.4.1) to verify that the conclusions of Theorem 10.13 are not affected by this addition.

Let

$$\mathcal{X}_n \equiv \{X_1, \ldots, X_n\}$$

as given in Theorem 10.13.

The following is the main result of this section:

THEOREM 1. *(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 *satisfies (1) 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

$$\text{pr}\left(\|\Psi_n^\circ(\hat{\theta}_n^\circ)\|_{\mathbb{L}} > \eta \,\middle|\, \mathcal{X}_n\right) \quad = \quad o_P(1) \;\; \textit{for every } \eta > 0, \qquad (2)$$

we have both

$$\|\hat{\theta}_n - \theta_0\| = o_P(1)$$

and

$$\text{pr}\left(\|\hat{\theta}_n^\circ - \theta_0\| > \eta \,\middle|\, \mathcal{X}_n\right) = o_P(1)$$

for every $\eta > 0$*.*

Note in (2) the absence of an outer probability on the left side.

This is because, as argued in the paragraph following Theorem 10.4, a continuous, real valued map of either of these bootstrapped empirical processes is measurable with respect to the random weights conditional on the data.

Note that we might have worked toward obtaining outer almost sure results since we are making a strong Glivenko-Cantelli assumption for the class of functions involved.

However, we only need convergence in probability for statistical applications.

Notice also that we only assumed

$$\|\Psi_n^\circ(\hat{\theta}_n^\circ)\|$$

goes to zero conditionally rather than unconditionally as done in Theorem 10.16.

However, it seems to be easier to check the conditional version in practice.

Moreover, the unconditional version is actually stronger than the conditional version, since

$$\mathsf{E}^*\mathsf{pr}\left(\|\Psi_n^\circ(\hat{\theta}_n^\circ)\|_{\mathbb{L}} > \eta \middle| \mathcal{X}_n \right) \leq \mathrm{P}^*\left(\|\Psi_n^\circ(\hat{\theta}_n^\circ)\|_{\mathbb{L}} > \eta \right)$$

by the version of Fubini's theorem given as Theorem 6.14.

It is unclear how to extend this argument to the outer almost sure setting.

This is another reason for restricting our attention to the convergence in probability results.

Nevertheless, we still need the strong Glivenko-Cantelli assumption since this enables the use of Theorems 10.13 and 10.15.

While the above approach will be helpful for some Z-estimators, many Z-estimators are complex enough to require individually Tailored approaches to establishing consistency.

Later on, we will revisit the Kaplan-Meier estimator example of Section 2.2.5 to which we can apply a generalization of Theorem 1, Theorem 3, which includes weak convergence.

In contrast, the proportional odds model for right-censored survival data, which will be presented in Chapter 15, requires a more individualized approach to establishing consistency.

Weak Convergence

Recall Theorem 2.11:

THEOREM 2. *Assume that*

- $\Psi(\theta_0) = 0$ *for some* θ_0 *in the interior of* Θ*,*
- $\sqrt{n}\Psi_n(\hat{\theta}_n) \xrightarrow{\mathsf{P}} 0$*, and*
- $\|\hat{\theta}_n - \theta_0\| \xrightarrow{\mathsf{P}} 0$ *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*

$$\frac{\left\|\sqrt{n}(\Psi_n(\hat{\theta}_n) - \Psi(\hat{\theta}_n)) - \sqrt{n}(\Psi_n(\theta_0) - \Psi(\theta_0))\right\|_{\mathbb{L}}}{1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|} \xrightarrow{\mathsf{P}} 0. \quad (3)$$

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

$$\|\sqrt{n}\dot{\Psi}_{\theta_0}(\hat{\theta}_n - \theta_0) + \sqrt{n}(\Psi_n - \Psi)(\theta_0)\|_{\mathbb{L}} \xrightarrow{\mathsf{P}} 0 \quad (4)$$

and thus

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

An important thing to note is that no assumptions about the data being i.i.d. are required.

Proof of Theorem 2. By the definitions of $\hat{\theta}_n$ and θ_0,

$$\begin{aligned} \sqrt{n}\left(\Psi(\hat{\theta}_n) - \Psi(\theta_0)\right) &= -\sqrt{n}\left(\Psi_n(\hat{\theta}_n) - \Psi(\hat{\theta}_n)\right) + o_P(1) \\ &= -\sqrt{n}(\Psi_n - \Psi)(\theta_0) + o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|), \quad (5) \end{aligned}$$

by Assumption (3).

Note the error terms throughout this theorem are with respect to the norms of the spaces, e.g. Θ or $\mathbb{L}$, involved.

Since $\dot{\Psi}_{\theta_0}$ is continuously invertible, we have by Part (i) of Lemma 6.16 that 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 $\overline{\text{lin}}\,\Theta$.

Combining this with the differentiability of Ψ yields

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

Combining this with (5), we obtain

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

We now have that $\hat{\theta}_n$ is $\sqrt{n}$-consistent for θ_0 with respect to $\| \cdot \|$.

By the differentiability of Ψ, the left side of (5) can be replaced by

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

This last error term is now $o_P(1)$ as also is the error term on the right side of (5).

Now the result (4) follows.

Next the continuity of $\dot{\Psi}_{\theta_0}^{-1}$ and the continuous mapping theorem yield

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

as desired.$\square$

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

LEMMA 1. *Assume the conditions of Theorem 2 except that consistency of* $\hat{\theta}_n$ *is strengthened to asymptotic tightness of* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *and the Fréchet differentiability of* Ψ *is weakened to Hadamard differentiability at* θ_0*.*

Then the results of Theorem 2 still hold.

Proof. The asymptotic tightness of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ enables (5) to imply

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

The Hadamard differentiability of Ψ yields

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

Combining, we now have

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

and all of the results of the theorem follow.$\square$

Using Donsker Classes

We now consider the special case where the data involved are i.i.d., i.e.,

$$\Psi_n(\theta)(h) = \mathbb{P}_n \psi_{\theta,h}$$

and

$$\Psi(\theta)(h) = P\psi_{\theta,h},$$

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 (3) to hold:

LEMMA 2. *Suppose the class of functions*

$$\{\psi_{\theta,h} - \psi_{\theta_0,h} : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\} \tag{6}$$

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

$$\sup_{h \in \mathcal{H}} P\left(\psi_{\theta,h} - \psi_{\theta_0,h}\right)^2 \quad \to \quad 0, \quad \text{as } \theta \to \theta_0. \tag{7}$$

Then if $\hat{\theta}_n \xrightarrow{\mathsf{P}} \theta_0$,

$$\sup_{h \in \mathcal{H}} \left| \mathbb{G}_n \psi_{\hat{\theta}_n,h} - \mathbb{G}_n \psi_{\theta_0,h} \right| = o_P(1).$$

Before giving the proof of this lemma, we make the somewhat trivial observation that the conclusion of this lemma implies (3).

Proof of Lemma 2. Let

$$\Theta_\delta \equiv \{\theta : \|\theta - \theta_0\| < \delta\}$$

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 almost all sample paths of Z_0.

By Slutsky's theorem (Theorem 7.15),

$$(Z_n, \hat{\theta}_n) \rightsquigarrow (Z_0, \theta_0).$$

The continuous mapping theorem (Theorem 7.7) now implies that

$$Z_n(\hat{\theta}_n) = f(Z_n, \hat{\theta}_n) \rightsquigarrow f(Z_0, \theta_0) = 0.\square$$

If, in addition to the assumptions of Lemma 2, we are willing to assume

$$\{\psi_{\theta_0,h} : h \in \mathcal{H}\} \tag{8}$$

is P-Donsker, then

$$\sqrt{n}(\Psi_n - \Psi)(\theta_0) \rightsquigarrow Z,$$

and all of the weak convergence assumptions of Theorem 2 are satisfied.

Alternatively, we could just assume that

$$\mathcal{F}_\delta \equiv \{\psi_{\theta,h} : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\} \tag{9}$$

is P-Donsker for some $\delta > 0$, then both (6) and (8) are P-Donsker for some $\delta > 0$.

We are now well poised for a Z-estimator master theorem for weak convergence.

A Master Theorem and the Bootstrap

In this section, we augment the results of the previous section to achieve a general Z-estimator master theorem that includes both weak convergence and validity of the bootstrap.

Here we consider the two bootstrapped Z-estimators described in Section 13.1, except that for the multiplier bootstrap we make the additional requirements that $0 < \tau^2 = \text{var}(\xi_1) < \infty$ and $\|\xi_1\|_{2,1} < \infty$.

We use $\underset{\circ}{\overset{\mathsf{P}}{\rightsquigarrow}}$ to denote either $\underset{\xi}{\overset{\mathsf{P}}{\rightsquigarrow}}$ or $\underset{W}{\overset{\mathsf{P}}{\rightsquigarrow}}$ depending on which bootstrap is being used, and we let the constant $k_0 = \tau/\mu$ for the multiplier bootstrap and $k_0 = 1$ for the multinomial bootstrap.

Here is the main result:

THEOREM 3. *Assume $\Psi(\theta_0) = 0$ and the following hold:*

(A) $\theta \mapsto \Psi(\theta)$ satisfies (1);

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

(C) The class $\mathcal{F}_\delta$ in (9) is P-Donsker for some $\delta > 0$;

(D) Condition (7) holds;

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

(F) $\theta \mapsto \Psi(\theta)$ is Fréchet-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*

$$\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) \underset{\circ}{\overset{\mathsf{P}}{\rightsquigarrow}} k_0 Z.$$

About the conditions:

- Condition (A) is identifiability.
- Conditions (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.
- Condition (F) specifies the smoothness and invertibility requirements of the derivative of Ψ.

Except for the last half of Condition (E), all of the conditions are requirements for asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

What is perhaps surprising is how little additional assumptions are needed to obtain bootstrap validity.

Only an assurance that the bootstrapped estimator is an approximate zero of the bootstrapped estimating equation is required.

Thus bootstrap validity is almost an automatic consequence of asymptotic normality.

Before giving the proof of the theorem, we will present an example.

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

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

where

- the observed data U, δ is the right-censored survival time and censoring indicator, respectively, and where
- we have replaced the survival function S with θ in the notation of Section 2.2.5 to obtain greater consistency with the notation of the current chapter.

The limiting estimating function $\Psi(\theta) = P\psi_\theta$, given in (2.11), is

$$\begin{aligned}\Psi(\theta)(t) &= P\psi_{\theta,t} \\ &= \theta_0(t)L(t) + \int_0^t \frac{\theta_0(u)}{\theta(u)} dG(u)\theta(t) - \theta(t), \quad (10)\end{aligned}$$

where

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

Thus, if we make the substitution

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

$\Psi(\theta_n)(t) \to 0$ 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 the same interval.

By solving this integral equation, we obtain

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

which implies $\epsilon_n(t) \to 0$ uniformly, since $L(t-) \geq L(\tau-) > 0$.

Thus $\|\theta_n - \theta_0\|_\infty \to 0$, implying the desired identifiability, and hence Condition (A) of the theorem are satisfied.

Exercise 2.4.3 verifies that $\Psi(\theta)$ is Fréchet differentiable with derivative $\dot{\Psi}_{\theta_0}$ defined in (2.16).

Exercise 2.4.4 verifies that $\dot{\Psi}_{\theta_0}$ is continuously invertible with inverse $\dot{\Psi}_{\theta_0}^{-1}$ given explicitly in (2.17).

Thus Condition (F) of the theorem is also established.

In the paragraphs in Section 2.2.5 after the presentation of Theorem 2.1, the class

$$\{\psi_\theta(t) : \theta \in \Theta, t \in [0, \tau]\},$$

where Θ is the class of all survival functions $t \mapsto \theta(t)$ with $\theta(0) = 0$ and with t restricted to $[0, \tau]$, was shown to be Donsker.

Note that in this setting, $\mathcal{H} = [0, \tau]$.

Thus Conditions (B) and (C) of the theorem hold.

It is also quite easy to verify directly that

$$\sup_{t\in[0,\tau]} P\left[\psi_\theta(t) - \psi_{\theta_0}(t)\right]^2 \to 0,$$

as $\|\theta - \theta_0\|_\infty \to 0$, and thus Condition (D) of the theorem is satisfied.

If $\hat{\theta}_n$ is the Kaplan-Meier estimator, then

$$\|\Psi_n(\hat{\theta}_n)(t)\|_\infty = 0$$

almost surely.

If the bootstrapped version is

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

where $\tilde{T}_1, \ldots, \tilde{T}_{m_n}$ are the observed failure times in the sample, then also

$$\|\Psi_n^\circ(\hat{\theta}_n^\circ)\|_\infty = 0$$

almost surely.

Thus Condition (E) of the theorem is satisfied, and hence all of the conditions of theorem are satisfied.

Thus we obtain consistency, weak convergence, and bootstrap consistency for the Kaplan-Meier estimator all at once.

As mentioned at the end of Section 13.1, a master result such as this will not apply to all Z-estimator settings.

Many interesting and important Z-estimators require an individualized approach to obtaining consistency, such as the Z-estimator for the proportional odds model for right-censored data which we examine in Chapter 15.

Proof of Theorem 3. The consistency of $\hat{\theta}_n$ and weak convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ follow from Theorems 1 and 2 and Lemma 2.

Theorem 1 also yields that there exists a decreasing sequence $0 < \eta_n \downarrow 0$ such that

$$\text{pr}\left(\|\hat{\theta}_n^\circ - \theta_0\| > \eta_n \middle| \mathcal{X}_n \right) = o_P(1).$$

Now we can use the same arguments used in the proof of Lemma 2, in combination with Theorem 2.6, to obtain that

$$\sqrt{n}(\Psi_n^\circ - \Psi)(\hat{\theta}_n^\circ) - \sqrt{n}(\Psi_n^\circ - \Psi)(\theta_0) = E_n,$$

where pr$(E_n > \eta | \mathcal{X}_n) = o_P(1)$ for all $\eta > 0$.

Combining this with arguments used in the proof of Theorem 2, we can deduce that

$$\sqrt{n}(\hat{\theta}_n^\circ - \theta_0) = -\dot{\Psi}_{\theta_0}^{-1}\sqrt{n}(\Psi_n^\circ - \Psi)(\theta_0) + E_n',$$

where pr$(E_n' > \eta | \mathcal{X}_n) = o_P(1)$ for all $\eta > 0$.

Combining this with the conclusion of Theorem 2, we obtain

$$\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) = -\dot{\Psi}_{\theta_0}^{-1}\sqrt{n}(\Psi_n^\circ - \Psi_n)(\theta_0) + E_n'',$$

where pr$(E_n'' > \eta | \mathcal{X}_n) = o_P(1)$ for all $\eta > 0$.

The final conclusion now follows from reapplication of Theorem 2.6.$\square$