Introduction to Empirical Processes and Semiparametric Inference Lecture 21: Proportional Odds Model, Continued

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Consistency

In this section, we prove uniform consistency of $\hat{\theta}_n$.

Let $\Theta \equiv \mathcal{B}_0 \times \mathcal{A}$ be the parameter space for θ, where

- $\mathcal{B}_0 \subset \mathbb{R}^d$ is the known compact containing β_0 and
- $\mathcal{A}$ is the collection of all monotone increasing functions

$$A : [0, \tau] \mapsto [0, \infty]$$

with $A(0) = 0$.

The following is the main result of this section:

THEOREM 1. *Under the given conditions,* $\hat{\theta}_n \xrightarrow{\text{as}*} \theta_0$.

Proof. Define $\tilde{\theta}_n = (\beta_0, \tilde{A}_n)$, where

$$\tilde{A}_n \equiv \int_0^{(\cdot)} [PW(s;\theta_0)]^{-1} \mathbb{P}_n dN(s).$$

Note that

$$\begin{aligned} L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n) &= \int_0^\tau \log \frac{PW(s;\theta_0)}{\mathbb{P}_n W(s;\hat{\theta}_n)} \mathbb{P}_n dN(s) \\ &+ (\hat{\beta}_n - \beta_0)' \mathbb{P}_n \int_0^\tau Z dN(s) \\ &- \mathbb{P}_n \left[(1+\delta) \log \left(\frac{1 + e^{\hat{\beta}_n' Z} \hat{A}_n(U)}{1 + e^{\beta_0' Z} \tilde{A}_n(U)} \right) \right]. \quad (1) \end{aligned}$$

By Lemma 1 below,

$$(\mathbb{P}_n - P)W(t;\hat{\theta}_n) \stackrel{\text{as}*}{\to} 0.$$

Combining this with Lemma 15.5 yields that

$$\liminf_{n\to\infty} \inf_{t\in[0,\tau]} \mathbb{P}_n W(t;\hat{\theta}_n) > 0$$

and that the $\limsup_{n\to\infty}$ of the total variation of

$$t \mapsto \left[PW(t;\hat{\theta}_n)\right]^{-1}$$

is $< \infty$ with inner probability one.

Since the class

$$\left\{ \int_0^t g(s)dN(s) : t \in [0,\tau], g \in D[0,\tau], \text{the total variation of } g \leq M \right\}$$

is Donsker for every $M < \infty$, we now have

$$\int_0^\tau \log \frac{PW(s;\theta_0)}{\mathbb{P}_n W(s;\hat{\theta}_n)} \mathbb{P}_n dN(s) - \int_0^\tau \log \frac{PW(s;\theta_0)}{PW(s;\hat{\theta}_n)} dQ_0(s) \quad \overset{\text{as}*}{\rightarrow} \quad 0. \tag{2}$$

Combining Lemma 15.5 with the fact that

$$\left\{(1+\delta)\log\left(1+e^{\beta' Z}A(U)\right) :\ \theta \in \Theta, A(\tau) \leq M\right\}$$

is Glivenko-Cantelli for each $M < \infty$ yields

$$(\mathbb{P}_n - P)\left[(1+\delta)\log\left(\frac{1+e^{\hat{\beta}_n' Z}\hat{A}_n(U)}{1+e^{\beta_0' Z}A_0(U)}\right)\right] \overset{\text{as}*}{\to} 0. \qquad (3)$$

Now combining results (2) and (3) with (1), we obtain that

$$
\begin{aligned}
L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n) - \int_0^\tau \log \frac{PW(s;\theta_0)}{PW(s;\hat{\theta}_n)} dQ_0(s) - (\hat{\beta}_n - \beta_0)' P[Z\delta] & \\
+P\left[(1+\delta)\log\left(\frac{1+e^{\hat{\beta}_n' Z}\hat{A}_n(U)}{1+e^{\beta_0' Z}A_0(U)}\right)\right] & \qquad (4) \\
\overset{\text{as}*}{\longrightarrow} \quad 0. &
\end{aligned}
$$

Now select a fixed sequence $X_1, X_2, \ldots$ for which the previous convergence results hold, and note that such sequences occur with inner probability one.

Reapplying (15.2) yields

$$\limsup_{n\to\infty} \sup_{s,t\in[0,\tau]} \frac{|\hat{A}_n(s) - \hat{A}_n(t)|}{|\mathbb{P}_n(N(s) - N(t))|} < \infty.$$

Thus there exists a subsequence $\{n_k\}$ along which both

$$\|\hat{A}_{n_k} - \tilde{A}\|_\infty \to 0$$

and $\hat{\beta}_{n_k} \to \tilde{\beta}$, for some $\tilde{\theta} = (\tilde{\beta}, \tilde{A})$, where $\tilde{A}$ is both continuous and bounded.

Combining this with (4), we obtain

$$0 \leq L_{n_k}(\hat{\theta}_{n_k}) - L_{n_k}(\tilde{\theta}_n) \to P_0 \log \left[\frac{dP_{\tilde{\theta}}}{dP_0} \right] \leq 0,$$

where

- P_θ is the probability measure of a single observation on the specified model at parameter value θ and
- $P_0 \equiv P_{\theta_0}$.

Since the "model is identifiable" (see Exercise 15.6.2), we obtain that $\hat{\theta}_n \to \theta_0$ uniformly.

Since the sequence $X_1, X_2, \ldots$ was an arbitrary representative from a set with inner probability one, we obtain that $\hat{\theta}_n \to \theta_0$ almost surely.

Since $\hat{A}_n$ is a piecewise constant function with jumps $\Delta \hat{A}_n$ only at observed failure times $t_1, \ldots, t_{m_n}$, $\hat{\theta}_n$ is a continuous function of a maximum taken over $m_n + d$ real variables.

This structure implies that

$$\sup_{t\in[0,\tau]} |\hat{A}_n(t) - A_0(t)|$$

is a measurable random variable, and hence the uniform distance between $\hat{\theta}_n$ and θ_0 is also measurable.

Thus the almost sure convergence can be strengthened to the desired outer almost sure convergence.$\Box$

LEMMA 1. *The class of functions* $\{W(t;\theta) : t \in [0,\tau], \theta \in \Theta\}$ *is* P*-Donsker.*

Proof. It is fairly easy to verify that

$$\mathcal{F}_1 \equiv \left\{(1+\delta)e^{\beta' Z}Y(t) : t \in [0,\tau], \beta \in \mathcal{B}_0\right\}$$

is a bounded P-Donsker class.

If we can also verify that

$$\mathcal{F}_2 \equiv \left\{ \left(1 + e^{\beta' Z} A(U)\right)^{-1} : \theta \in \Theta \right\}$$

is P-Donsker, then we are done since the product of two bounded Donsker classes is also Donsker.

To this end, let $\phi : \mathbb{R}^2 \mapsto \mathbb{R}$ be defined by

$$\phi(x, y) = \frac{1 - y}{1 - y + e^x y},$$

and note that ϕ is Lipschitz continuous on sets of the form $[-k, k] \times [0, 1]$, with a finite Lipschitz constant depending only on k, for all $k < \infty$ (see Exercise 15.6.3).

Note also that

$$\mathcal{F}_2 = \left\{ \phi\left(\beta' Z, \frac{A(U)}{1 + A(U)} \right) : \theta \in \Theta \right\}.$$

Clearly,

$$\{\beta' Z : \beta \in \mathcal{B}_0\}$$

is Donsker with range contained in $[-k_0, k_0]$ for some $k_0 < \infty$ by the given conditions.

Moreover,

$$\{A(U)(1 + A(U))^{-1} : A \in \mathcal{A}\}$$

is a subset of all monotone, increasing functions with range $[0, 1]$ and thus, by Theorem 9.24, is Donsker.

Hence, by Theorem 9.31, $\mathcal{F}_2$ is P-Donsker, and the desired conclusions follow.$\square$

Score and Information Operators

Because we have an infinite dimensional parameter A, we need to take care with score and information operator calculations.

The overall idea is that we need these operators in order to utilize the general Z-estimator convergence theorem (Theorem 2.11) in the next section to establish asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and obtain bootstrap validity.

To facilitate the development of these operators, let $\mathcal{H}$ denote the space of elements $h = (h_1, h_2)$ with $h_1 \in \mathbb{R}^d$ and $h_2 \in D[0, \tau]$ of bounded variation.

We supply $\mathcal{H}$ with the norm

$$\|h\|_{\mathcal{H}} \equiv \|h_1\| + \|h_2\|_v,$$

where $\|\cdot\|$ is the Euclidean norm and $\|\cdot\|_v$ is the total variation norm.

Define

$$\mathcal{H}_p \equiv \{h \in \mathcal{H} : \|h\|_{\mathcal{H}} \leq p\},$$

where the inequality is strict when $p = \infty$.

The parameter θ can now be viewed as an element of $\ell^\infty(\mathcal{H}_p)$ if we define

$$\theta(h) \equiv h_1'\beta + \int_0^\tau h_2(s)dA(s), \quad h \in \mathcal{H}_p, \quad \theta \in \Theta.$$

Note that $\mathcal{H}_1$ is sufficiently rich to be able to extract out all components of θ.

For example,

$$\tilde{h} = (h_1, 0) = ((0, 1, 0, \ldots)', 0)$$

extracts out the second component of β, i.e., $\beta_2 = \theta(\tilde{h})$, while

$$\tilde{h}_{*,u} = (0, \mathbf{1}\{(\cdot) \leq u\})$$

extracts out $A(u)$, i.e., $A(u) = \theta(\tilde{h}_{*,u})$.

As a result, the parameter space Θ becomes a subset of $\ell^\infty(\mathcal{H}_p)$ with norm

$$\|\theta\|_{(p)} \equiv \sup_{h \in \mathcal{H}_p} |\theta(h)|.$$

We can study weak convergence in the uniform topology of $\hat{\theta}_n$ via this functional representation of the parameter space since, for all $1 \leq p < \infty$ and every $\theta \in \Theta$,

$$\|\theta\|_\infty \leq \|\theta\|_{(p)} \leq 4p\|\theta\|_\infty$$

(see Exercise 15.6.4).

We now calculate the score operator which will become the Z-estimating equation to which we will apply the Z-estimator convergence theorem.

Consider the one-dimensional submodel defined by the map

$$t \mapsto \theta_t \equiv \theta + t(h_1, \int_0^{(\cdot)} h_2(s)dA(s)), \quad h \in \mathcal{H}_p.$$

The score operator has the form

$$V_n^\tau(\theta)(h) \equiv \left. \frac{\partial}{\partial t} L_n(\theta_t) \right|_{t=0} = V_{n,1}^\tau(\theta)(h_1) + V_{n,2}^\tau(\theta)(h_2),$$

where

$$V_{n,1}^\tau(\theta)(h_1) \equiv \mathbb{P}_n \left\{ h_1' Z N(\tau) - (1+\delta) \left[\frac{h_1' Z e^{\beta' Z} A(U \wedge \tau)}{1 + e^{\beta' Z} A(U \wedge \tau)} \right] \right\},$$

and $V_{n,2}^\tau(\theta)(h_2)$ is defined by replacing h_1 with h_2 in (15.10).

As mentioned earlier, we will need to utilize the dependence on τ at a later point.

Now we have that the NPMLE $\hat{\theta}_n$ can be characterized as a zero of the map

$$\theta \mapsto \Psi_n(\theta) = V_n^\tau(\theta),$$

and thus $\hat{\theta}_n$ is a Z-estimator with the estimating equation residing in $\ell^\infty(\mathcal{H}_\infty)$.

The expectation of Ψ_n is $\Psi \equiv PV^\tau$, where V^τ equals V_1^τ (i.e., V_n^τ with $n = 1$), with X_1 replaced by a generic observation X.

Thus V^τ also satisfies $V_n^\tau = \mathbb{P}_n V^\tau$.

The Gâteaux derivative of Ψ at any $\theta_1 \in \Theta$ exists and is obtained by differentiating over the submodels $t \mapsto \theta_1 + t\theta$.

This derivative is

$$\dot{\Psi}_{\theta_1}(h) \equiv \frac{\partial}{\partial t}\Psi(\theta_1 + t\theta)\bigg|_{t=0} = -\theta(\sigma_{\theta_1}(h)),$$

where the operator $\sigma_\theta : \mathcal{H}_\infty \mapsto \mathcal{H}_\infty$ can be shown to be

$$\begin{aligned} \sigma_\theta(h) &= \begin{pmatrix} \sigma_\theta^{11} & \sigma_\theta^{12} \\ \sigma_\theta^{21} & \sigma_\theta^{22} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\ &\equiv P \begin{pmatrix} \hat{\sigma}_\theta^{11} & \hat{\sigma}_\theta^{12} \\ \hat{\sigma}_\theta^{21} & \hat{\sigma}_\theta^{22} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\ &\equiv P\hat{\sigma}_\theta(h), \end{aligned} \tag{5}$$

where

$$
\begin{aligned}
\hat{\sigma}_\theta^{11}(h_1) &= \hat{\xi}_\theta A(U) Z Z' h_1 \\
\hat{\sigma}_\theta^{12}(h_2) &= \hat{\xi}_\theta Z \int_0^\tau Y(s) h_2(s) dA(s) \\
\hat{\sigma}_\theta^{21}(h_1)(s) &= \hat{\xi}_\theta Y(s) Z' h_1 \\
\hat{\sigma}_\theta^{22}(h_2)(s) &= \frac{(1+\delta) e^{\beta' Z} Y(s) h_2(s)}{1 + e^{\beta' Z} A(U)} \\
&\quad -\hat{\xi}_\theta e^{\beta' Z} \int_0^\tau Y(u) h_2(u) dA(u) Y(s),
\end{aligned}
$$

and

$$
\hat{\xi}_\theta \equiv \frac{(1+\delta) e^{\beta' Z}}{\left[1 + e^{\beta' Z} A(U)\right]^2}.
$$

We need to strengthen this Gâteaux differentiability of Ψ to Fréchet differentiability, at least at $\theta = \theta_0$.

This is accomplished in the following lemma:

LEMMA 2. *Under the given assumptions, the operator* $\theta \mapsto \Psi(\theta)(\cdot)$, *viewed as a map from* $\ell^\infty(\mathcal{H}_p)$ *to* $\ell^\infty(\mathcal{H}_p)$, *is Fréchet differentiable for each* $p < \infty$ *at* $\theta = \theta_0$, *with derivative*

$$\theta \mapsto \dot{\Psi}_{\theta_0}(\theta) \equiv -\theta(\sigma_{\theta_0}(\cdot)).$$

Before giving the proof of this lemma, we note that we also need to verify that $\dot{\Psi}_{\theta_0}$ is continuously invertible.

The following theorem establishes this plus a little more:

THEOREM 2. *Under the given conditions,*

$$\sigma_{\theta_0} : \mathcal{H}_\infty \mapsto \mathcal{H}_\infty$$

is continuously invertible and onto.

Moreover,

$$\dot{\Psi}_{\theta_0} : \overline{\mathit{lin}}\,\Theta \mapsto \overline{\mathit{lin}}\,\Theta$$

is also continuously invertible and onto, with inverse

$$\theta \mapsto \dot{\Psi}_{\theta_0}^{-1}(\theta) \equiv -\theta(\sigma_{\theta_0}^{-1}),$$

where $\sigma_{\theta_0}^{-1}$ *is the inverse of* σ_{θ_0}*.*

The “onto” property is not needed for the Z-estimator convergence theorem, but it will prove useful in Chapter 22 where we will revisit this example and show that $\hat{\theta}_n$ is asymptotically efficient.

We conclude this section with the proofs of Lemma 2 and Theorem 2.

Proof of Lemma 2. By the smoothness of $\theta \mapsto \sigma_\theta(\cdot)$, we have

$$\lim_{t \downarrow 0} \sup_{\theta : \|\theta\|_{(p)} \leq 1} \sup_{h \in \mathcal{H}_p} \left| \int_0^\tau \theta \left(\sigma_{\theta_0 + ut\theta}(h) - \sigma_{\theta_0}(h) \right) du \right| = 0.$$

Thus

$$\sup_{h \in \mathcal{H}_p} \left| \Psi(\theta_0 + \theta)(h) - \Psi(\theta_0)(h) + \theta \left(\sigma_{\theta_0}(h) \right) \right| = o\left(\|\theta\|_{(p)} \right)$$

as $\|\theta\|_{(p)} \to 0.\Box$

Proof of Theorem 2. From the explicit form of the operator σ_{θ_0} defined above, we have that

$$\sigma_{\theta_0} = \sigma_1 + \sigma_2,$$

where

$$\sigma_1 \equiv \left(\begin{array}{cc} I & 0 \\ 0 & g_0(\cdot) \end{array} \right),$$

where I is the $d \times d$ identity matrix,

$$g_0(s) = P\left[\frac{(1+\delta)e^{\beta_0' Z} Y(s)}{1 + e^{\beta_0' Z} A_0(U)} \right],$$

and where

$$\sigma_2 \equiv \sigma_{\theta_0} - \sigma_1.$$

It is not hard to verify that σ_2 is a compact operator, i.e., that the range of $h \mapsto \sigma_2(h)$ over the unit ball in $\mathcal{H}_\infty$ lies within a compact set (see Exercise 15.6.6).

Note that since $1/g_0$ has bounded total variation, we have for any $g = (g_1, g_2) \in \mathcal{H}$, that $g = \sigma_1(h)$, where

$$h = \left(g_1, \frac{g_2(\cdot)}{g_0(\cdot)} \right) \in \mathcal{H}.$$

Thus σ_1 is onto.

It is also true that

$$\|g_0(\cdot)h_2(\cdot)\|_{\mathcal{H}} \geq \left(\inf_{s\in[0,\tau]} |g_0(s)| \right) \|h_2\|_{\mathcal{H}} \geq c_0 \|h_2\|_{\mathcal{H}},$$

and thus σ_1 is both continuously invertible and onto.

If we can also verify that σ_{θ_0} is one-to-one, we then have by Lemma 6.17 that

$$\sigma_{\theta_0} = \sigma_1 + \sigma_2$$

is both continuously invertible and onto.

We will now verify that σ_{θ_0} is one-to-one by showing that for any $h \in \mathcal{H}_\infty$, $\sigma_{\theta_0}(h) = 0$ implies that $h = 0$.

Fix an $h \in \mathcal{H}_\infty$ for which $\sigma_{\theta_0}(h) = 0$, and define the one-dimensional submodel

$$t \mapsto \theta_{0t} = (\beta_{0t}, A_{0t}) \equiv (\beta_0, A_0) + t\left(h_1, \int_0^{(\cdot)} h_2(s)dA(s)\right).$$

Note that $\sigma_{\theta_0}(h) = 0$ implies

$$P\left\{ -\frac{\partial^2}{(\partial t)^2} \ell_n(\theta_{0t}) \bigg|_{t=0} \right\} \;=\; P[V^\tau(\theta_0)(h)]^2 \;=\; 0, \qquad (6)$$

where we are using the original form of the likelihood ℓ_n instead of the modified form L_n because the submodels A_{0t} are differentiable for all t small enough.

It can be verified that $V^u(\theta_0)(h)$ is a continuous time martingale over $u \in [0, \tau]$ (here is where we need the dependency of V^τ on τ).

The basic idea is that $V^u(\theta_0)(h)$ can be reexpressed as

$$\int_0^u \left(\frac{\dot\lambda_{\theta_0}(s)}{\lambda_{\theta_0}(s)} \right) \big(h_1' Z + h_2(s)\big)\, dM(s),$$

where

$$\dot\lambda_{\theta_0} \equiv \frac{\partial}{\partial t} \lambda_{\theta_0 t} \Big|_{t=0}, \qquad \lambda_\theta(u) \equiv \frac{e^{\beta' Z} a_0(u)}{1 + e^{\beta' Z} A(u)},$$

and where

$$M(u) = N(u) - \int_0^u Y(s) \lambda_{\theta_0}(s) ds$$

is a martingale since λ_{θ_0} is the correct hazard function for the failure time T given Z.

Thus

$$P[V^{\tau}(\theta_0)(h)]^2 = P[V^{u}(\tau)(\theta_0)(h)]^2 + P[V^{\tau}(\theta_0)(h) - V^{u}(\theta_0)(h)]^2$$

for all $u \in [0, \tau]$, and hence $P[V^{u}(\theta_0)(h)]^2 = 0$ for all $u \in [0, \tau]$.

Thus $V^{u}(\theta_0)(h) = 0$ almost surely, for all $u \in [0, \tau]$.

Hence, if we assume that the failure time T is censored at some $U \in (0, \tau]$, we have almost surely that

$$\frac{e^{\beta_0' Z} \int_0^u (h_1' Z + h_2(s)) Y(s) dA_0(s)}{1 + \int_0^u e^{\beta_0' Z} Y(s) dA_0(s)} = 0,$$

for all $u \in [0, \tau]$.

Hence

$$\int_0^u (h_1' Z + h_2(s)) Y(s) dA_0(s) = 0$$

almost surely for all $u \in [0, \tau]$.

Taking the derivative with respect to u yields that $h_1' Z + h_2(u) = 0$ almost surely for all $u \in [0, \tau]$.

This of course forces $h = 0$ since var$[Z]$ is positive definite.

Thus σ_{θ_0} is one-to-one since h was an arbitrary choice satisfying $\sigma_{\theta_0}(h) = 0$.

Hence σ_{θ_0} is continuously invertible and onto, and the first result of the theorem is proved.

We now prove the second result of the theorem.

Note that since

$$\sigma_{\theta_0} : \mathcal{H}_\infty \mapsto \mathcal{H}_\infty$$

is continuously invertible and onto, for each $p > 0$, there is a $q > 0$ such that

$$\sigma_{\theta_0}^{-1}(\mathcal{H}_q) \subset \mathcal{H}_p.$$

Fix $p > 0$, and note that

$$\begin{aligned}
\inf_{\theta \in \text{lin}\,\Theta} \frac{\|\theta(\sigma_{\sigma_0}(\cdot))\|_{(p)}}{\|\theta(\cdot)\|_{(p)}} &\geq \inf_{\theta \in \text{lin}\,\Theta} \left[\frac{\sup_{h \in \sigma_{\theta_0}^{-1}(\mathcal{H}_{(q)})} |\theta(\sigma_{\theta_0}^{-1}(h))|}{\|\theta\|_{(p)}} \right] \\
&= \inf_{\theta \in \Theta} \frac{\|\theta\|_{(q)}}{\|\theta\|_{(p)}} \geq \frac{q}{2p}.
\end{aligned}$$

(See Exercise 15.6.4 to verify the last inequality.)

Thus

$$\|\theta(\sigma_{\theta_0})\|_{(p)} \geq c_p\|\theta\|_{(p)},$$

for all $\theta \in \text{lin}\,\Theta$, where $c_p > 0$ depends only on p.

Lemma 6.16, Part (i), now implies that $\theta \mapsto \theta(\sigma_{\theta_0})$ is continuously invertible.

For any $\theta_1 \in \overline{\text{lin}}\,\Theta$, we have $\theta_2(\sigma_{\theta_0}) = \theta_1$, where

$$\theta_2 = \theta_1(\sigma_{\theta_0}^{-1}) \in \overline{\text{lin}}\,\Theta.$$

Thus $\theta \mapsto \theta(\sigma_{\theta_0})$ is also onto.

Hence

$$\theta \mapsto \dot{\Psi}_{\theta_0}(\theta) = -\theta(\sigma_{\theta_0})$$

is both continuously invertible and onto, and the theorem is proved.$\square$

Weak Convergence and Bootstrap Validity

Our approach to establishing weak convergence will be through verifying the conditions of Theorem 2.11 via the Donsker class result of Lemma 13.3.

After establishing weak convergence, we will use a similar technical approach, but with some important differences, to obtain validity of a simple weighted bootstrap procedure.

Recall that

$$\Psi_n(\theta)(h) = \mathbb{P}_n V^\tau(\theta)(h),$$

and note that $V^\tau(\theta)(h)$ can be expressed as

$$V^\tau(\theta)(h) = \int_0^\tau (h_1' Z + h_2(s)) dN(s) - \int_0^\tau (h_1' Z + h_2(s)) W(s;\theta) dA(s).$$

We now show that for any $0 < \epsilon < \infty$,

$$\mathcal{G}_\epsilon \equiv \{V^\tau(\theta)(h) : \theta \in \Theta_\epsilon, h \in \mathcal{H}_1\},$$

where

$$\Theta_\epsilon \equiv \{\theta \in \Theta : \; \|\theta - \theta_0\|_{(1)} \leq \epsilon\}$$

is P-Donsker.

First, Lemma 1 above tells us that $\{W(t;\theta) : t \in [0,\tau], \theta \in \Theta\}$ is Donsker.

Second, it is easily seen that the class

$$\{h_1'Z + h_2(t) : t \in [0,\tau], h \in \mathcal{H}_1\}$$

is also Donsker.

Since the product of bounded Donsker classes is also Donsker, we have that

$$\{f_{t,\theta}(h) \equiv (h_1'Z + h_2(t))W(t;\theta) : t \in [0,\tau], \theta \in \Theta_\epsilon, h \in \mathcal{H}_1\}$$

is Donsker.

Third, consider the map

$$\phi : \ell^\infty([0,\tau] \times \Theta_\epsilon \times \mathcal{H}_1) \mapsto \ell^\infty(\Theta_\epsilon \times \mathcal{H}_1 \times \mathcal{A}_\epsilon)$$

defined by

$$\phi(f_{\cdot,\theta}(h)) \equiv \int_0^\tau f_{s,\theta}(h) d\tilde{A}(s),$$

for $\tilde{A}$ ranging over

$$\mathcal{A}_\epsilon \equiv \{A \in \mathcal{A} : \sup_{t \in [0,\tau]} |A(t) - A_0(t)| \leq \epsilon\}.$$

Note that for any $\theta_1, \theta_2 \in \Theta_\epsilon$ and $h, \tilde{h} \in \mathcal{H}_1$,

$$\left|\phi(f_{\cdot,\theta_1}(h)) - \phi(f_{\cdot,\theta_2}(\tilde{h}))\right| \leq \sup_{t\in[0,\tau]} \left|f_{t,\theta_1}(h) - f_{t,\theta_2}(\tilde{h})\right| \times (A_0(\tau)+\epsilon).$$

Thus ϕ is continuous and linear, and hence the class

$$\{\phi(f_{\cdot,\theta}(h)) : \theta \in \Theta_\epsilon, h \in \mathcal{H}_1\}$$

is Donsker by Lemma 3 below.

Thus also

$$\left\{ \int_0^\tau (h_1' Z + h_2(s)) W(s; \theta) dA(s) : \ \theta \in \Theta_\epsilon, h \in \mathcal{H}_1 \right\}$$

is Donsker.

Since it not hard to verify that

$$\left\{ \int_0^\tau (h_1' Z + h_2(s)) dN(s) : h \in \mathcal{H}_1 \right\}$$

is also Donsker, we now have that $\mathcal{G}_\epsilon$ is indeed Donsker as desired.

We now present the needed lemma and its proof before continuing:

LEMMA 3. *Suppose $\mathcal{F}$ is Donsker and*

$$\phi : \ell^\infty(\mathcal{F}) \mapsto \mathbb{D}$$

is continuous and linear.

Then $\phi(\mathcal{F})$ is Donsker.

Proof. Observe that

$$\mathbb{G}_n\phi(\mathcal{F}) = \phi(\mathbb{G}_n\mathcal{F}) \rightsquigarrow \phi(\mathbb{G}\mathcal{F}) = \mathbb{G}(\phi(\mathcal{F})),$$

where

- the first equality follows from linearity,
- the weak convergence follows from the continuous mapping theorem,
- the second equality follows from a reapplication of linearity, and
- the meaning of the "abuse in notation" is obvious.$\square$

We now have that both

$$\{V^\tau(\theta)(h) - V^\tau(\theta_0)(h) : \ \theta \in \Theta_\epsilon, h \in \mathcal{H}_1\}$$

and

$$\{V^\tau(\theta_0)(h) : h \in \mathcal{H}_1\}$$

are also Donsker.

Thus

$$\sqrt{n}(\Psi_n(\theta_0) - \Psi(\theta_0)) \rightsquigarrow \mathbb{G}V^\tau(\theta_0)$$

in $\ell^\infty(\mathcal{H}_1)$.

Moreover, since it is not hard to show (see Exercise 15.6.7) that

$$\sup_{h\in\mathcal{H}_1} P\left(V^\tau(\theta)(h) - V^\tau(\theta_0)(h)\right)^2 \to 0, \qquad \text{as } \theta \to \theta_0, \tag{7}$$

Lemma 13.3 yields that

$$\left\|\sqrt{n}(\Psi_n(\theta) - \Psi(\theta)) - \sqrt{n}(\Psi_n(\theta_0) - \Psi(\theta_0))\right\|_{(1)} = o_P(1), \text{ as } \theta \to \theta_0.$$

Combining these results with Theorem 2, we have that all of the conditions of Theorem 2.11 are satisfied, and thus

$$\left\| \sqrt{n}\dot{\Psi}_{\theta_0}(\hat{\theta}_n - \theta_0) + \sqrt{n}(\Psi_n(\theta_0) - \Psi(\theta_0)) \right\|_{(1)} = o_P(1) \tag{8}$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{Z}_0 \equiv -\dot{\Psi}_{\theta_0}^{-1}\left(\mathbb{G}V^{\tau}(\theta_0)\right)$$

in $\ell^{\infty}(\mathcal{H}_1)$.

We can observe from this result that $\mathcal{Z}_0$ is a tight, mean zero Gaussian process with covariance

$$P[\mathcal{Z}_0(h)\mathcal{Z}_0(\tilde{h})] = P\left[V^\tau(\theta_0)(\sigma_{\theta_0}^{-1}(h))V^\tau(\theta_0)(\sigma_{\theta_0}^{-1}(\tilde{h})\right],$$

for any $h, \tilde{h} \in \mathcal{H}_1$.

As pointed out earlier, this is in fact uniform convergence since any component of θ can be extracted via $\theta(h)$ for some $h \in \mathcal{H}_1$.

Now we will establish validity of a weighted bootstrap procedure for inference.

Let $w_1, \ldots, w_n$ be positive, i.i.d., and independent of the data $X_1, \ldots, X_n$, with

- $0 < \mu \equiv Pw_1 < \infty$,
- $0 < \sigma^2 \equiv$ var$(w_1) < \infty$, and
- $\|w_1\|_{2,1} < \infty$.

Define the weighted bootstrapped empirical process

$$\tilde{\mathbb{P}}_n \equiv n^{-1} \sum_{i=1}^{n} (w_i/\bar{w}) \Delta_{X_i},$$

where $\bar{w} \equiv n^{-1} \sum_{i=1}^{n} w_i$ and Δ_{X_i} is the empirical measure for the observation X_i.

This particular bootstrap was introduced in Section 2.2.3.

Let $\tilde{L}_n(\theta)$ be $L_n(\theta)$ but with $\mathbb{P}_n$ replaced by $\tilde{\mathbb{P}}_n$, and let $\tilde{\Psi}_n$ be Ψ_n but with $\mathbb{P}_n$ replaced by $\tilde{\mathbb{P}}_n$.

Define $\tilde{\theta}_n$ to be the maximizer of $\theta \mapsto \tilde{L}_n(\theta)$.

The idea is, after conditioning on the data sample $X_1, \ldots, X_n$, to compute $\tilde{\theta}_n$ for many replications of the weights $w_1, \ldots, w_n$ to form confidence intervals for θ_0.

We want to show that

$$\sqrt{n}(\mu/\sigma)(\tilde{\theta}_n - \hat{\theta}_n) \quad \underset{w}{\overset{\mathsf{P}}{\rightsquigarrow}} \quad \mathcal{Z}_0. \tag{9}$$

We first study the unconditional properties of $\tilde{\theta}_n$.

Note that for maximizing $\theta \mapsto \tilde{L}_n(\theta)$ and for zeroing $\theta \mapsto \tilde{\Psi}_n$, we can temporarily drop the $\bar{w}$ factor since neither the maximizer nor zero of a function is modified when multiplied by a positive constant.

Let w be a generic version of w_1, and note that if a class of functions $\mathcal{F}$ is Glivenko-Cantelli, then so also is the class of functions $w \cdot \mathcal{F}$ via Theorem 10.13.

Likewise, if the class $\mathcal{F}$ is Donsker, then so is $w \cdot \mathcal{F}$ via the multiplier central limit theorem, Theorem 10.1.

Also, $Pwf = \mu Pf$, trivially.

What this means, is that the arguments in Sections 15.3.2 and 15.3.3 can all be replicated for $\tilde{\theta}_n$ with only trivial modifications.

This means that $\tilde{\theta}_n \overset{\text{as}*}{\rightarrow} \theta_0$.

Now, reinstate the $\bar{w}$ everywhere, and note by Corollary 10.3, we can verify that both

$$\sqrt{n}(\tilde{\Psi} - \Psi)(\theta_0) \rightsquigarrow (\sigma/\mu)\mathbb{G}_1 V^\tau(\theta_0) + \mathbb{G}_2 V^\tau(\theta_0),$$

where $\mathbb{G}_1$ and $\mathbb{G}_2$ are independent Brownian bridge random measures, and

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

Thus reapplication of Theorem 2.11 yields that

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

Combining this with (8), we obtain

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

Now, using

- the linearity of $\dot{\Psi}_{\theta_0}$,
- the continuity of $\dot{\Psi}_{\theta_0}^{-1}$, and
- the bootstrap central limit theorem, Theorem 2.6,

we have the desired result that

$$\sqrt{n}(\mu/\sigma)(\tilde{\theta}_n - \hat{\theta}_n) \underset{w}{\overset{\mathsf{P}}{\rightsquigarrow}} \mathcal{Z}_0.$$

Thus the proposed weighted bootstrap is valid.

We also note that it is not clear how to verify the validity of the usual nonparametric bootstrap, although its validity probably does hold.

The key to the relative simplicity of the theory for the proposed weighted bootstrap is that Glivenko-Cantelli and Donsker properties of function classes are not altered after multiplying by independent random weights satisfying the given moment conditions.

We also note that the weighted bootstrap is computationally simple, and thus it is quite practical to generate a reasonably large number of replications of $\tilde{\theta}_n$ to form confidence intervals.

This is demonstrated numerically in Kosorok, Lee and Fine (2004).