Introduction to Empirical Processes and Semiparametric Inference Lecture 20: Examples

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Example: A Change-Point Model

For this model, we observe i.i.d. realizations of $X = (Y, Z)$, where

$$Y = \alpha \mathbf{1}\{Z \leq \zeta\} + \beta \mathbf{1}\{Z > \zeta\} + \epsilon,$$

- Z and ϵ are independent
- with ϵ continuous, $\mathrm{E}\epsilon = 0$ and $\sigma^2 \equiv \mathrm{E}\epsilon^2 < \infty$,
- $\gamma \equiv (\alpha, \beta) \in \mathbb{R}^2$
- and ζ is known to lie in a bounded interval $[a, b]$.

The unknown parameters can be collectedas $\theta = (\gamma, \zeta)$, and the subscript zero will be used to denote the true parameter values.

We make the very important assumptions

- $\alpha_0 \neq \beta_0$
- Z has a strictly bounded and positive density f over $[a, b]$ with pr$(Z < a) > 0$ and pr$(Z > b) > 0$.

Our goal is to estimate θ through least squares.

This is the same as maximizing $M_n(\theta) = \mathbb{P}_n m_\theta$, where

$$m_\theta(x) \equiv -\left(y - \alpha \mathbf{1}\{z \leq \zeta\} - \beta \mathbf{1}\{z > \zeta\}\right)^2 .$$

Let $\hat{\theta}_n$ be maximizers of $M_n(\theta)$, where

$$\hat{\theta}_n \equiv (\hat{\gamma}_n, \hat{\zeta}_n)$$

and

$$\hat{\gamma}_n \equiv (\hat{\alpha}_n, \hat{\beta}_n).$$

Since we are not assuming that γ is bounded, we first need to prove the existence of $\hat{\gamma}_n$, i.e., we need to prove that

$$\|\hat{\gamma}_n\| = O_P(1).$$

We then need to provide consistency of all parameters and then establish the rates of convergence for the parameters.

Finally, we need to obtain the joint limiting distribution of the parameter estimates.

Existence

Note that the covariate Z and parameter ζ can be partitioned into four mutually exclusive sets:

- $\{Z \leq \zeta \wedge \zeta_0\}$,
- $\{\zeta < Z \leq \zeta_0\}$,
- $\{\zeta_0 < Z \leq \zeta\}$ and
- $\{Z > \zeta \vee \zeta_0\}$.

Since also

$$\mathbf{1}\{Z < a\} \leq \mathbf{1}\{Z \leq \zeta \wedge \zeta_0\}$$

and

$$\mathbf{1}\{Z > b\} \leq \mathbf{1}\{Z > \zeta \vee \zeta_0\}$$

by assumption, we obtain

$$\begin{aligned} -\mathbb{P}_n\epsilon^2 &= M_n(\theta_0) \\ &\leq M_n(\hat{\theta}_n) \\ &\leq -\mathbb{P}_n\left[(\epsilon - \hat{\alpha}_n + \alpha_0)^2\mathbf{1}\{Z < a\} + (\epsilon - \hat{\beta}_n + \beta_0)^2\mathbf{1}\{Z > b\}\right]. \end{aligned}$$

By decomposing the squares, we now have

$$
\begin{aligned}
(\hat{\alpha}_n - \alpha_0)^2 \mathbb{P}_n[\mathbf{1}\{Z < a\}] + (\hat{\beta}_n - \beta_0)^2 \mathbb{P}_n[\mathbf{1}\{Z > b\}] & \\
\leq \quad & \mathbb{P}_n[\epsilon^2 \mathbf{1}\{a \leq z \leq b\}] \\
& +2|\hat{\alpha}_n - \alpha_0| \mathbb{P}_n[\epsilon \mathbf{1}\{Z < a\}] \\
& +2|\hat{\beta}_n - \beta_0| \mathbb{P}_n[\epsilon \mathbf{1}\{Z > b\}] \\
\leq \quad & O_p(1) + o_P(1) \|\hat{\gamma}_n - \gamma_0\|.
\end{aligned}
$$

Since

$$\text{pr}(Z < a) \wedge \text{pr}(Z > b) > 0,$$

the above now implies that

$$\|\hat{\gamma}_n - \gamma_0\|^2 = O_P(1 + \|\hat{\gamma}_n - \gamma_0\|)$$

and hence that

$$\|\hat{\gamma}_n - \gamma_0\| = O_P(1).$$

Thus all the parameters are bounded in probability and therefore exist.

Consistency

Our approach to establishing consistency will be to utilize the argmax theorem (Theorem 14.1).

We first need to establish that $M_n \rightsquigarrow M$ in $\ell^\infty(K)$ for all compact $K \subset H \equiv \mathbb{R}^2 \times [a, b]$, where $M(\theta) \equiv Pm_\theta$.

We then need to show that $\theta \mapsto M(\theta)$ is upper semicontinuous with a unique maximum at θ_0.

We already know from the previous paragraph that $\hat{\theta}_n$ is asymptotically tight (i.e., $\|\hat{\theta}_n\| = O_P(1)$).

The argmax theorem will then yield that $\hat{\theta}_n \rightsquigarrow \theta_0$ as desired.

Fix a compact $K \subset H$.

We now verify that $\mathcal{F}_K \equiv \{m_\theta : \theta \in K\}$ is Glivenko-Cantelli.

Note that

$$\begin{aligned} m_\theta(X) &= -(\epsilon - \alpha + \alpha_0)^2 \mathbf{1}\{Z \leq \zeta \wedge \zeta_0\} \\ &\quad -(\epsilon - \beta + \alpha_0)^2 \mathbf{1}\{\zeta < Z \leq \zeta_0\} \\ &\quad -(\epsilon - \alpha + \beta_0)^2 \mathbf{1}\{\zeta_0 < Z \leq \zeta\} \\ &\quad -(\epsilon - \beta + \beta_0)^2 \mathbf{1}\{Z > \zeta \vee \zeta_0\}. \end{aligned}$$

It is not difficult to verify that

$$\{(\epsilon - \alpha + \alpha_0)^2 : \theta \in K\}$$

and

$$\mathbf{1}\{Z \leq \zeta \wedge \zeta_0 : \theta \in K\}$$

are separately Glivenko-Cantelli classes.

Thus the product of the two class is also Glivenko-Cantelli by Corollary 9.27 since the product of the two envelopes is integrable.

Similar arguments reveal that the remaining components of the sum are also Glivenko-Cantelli, and reapplication of Corollary 9.27 yields that $\mathcal{F}_K$ itself is Glivenko-Cantelli.

Thus $M_n \rightsquigarrow M$ in $\ell^\infty(K)$ for all compact K.

We now establish upper semicontinuity of $\theta \mapsto M(\theta)$ and uniqueness of the maximum.

Using the decomposition of the sets for (Z, ζ) used in the *Existence* paragraph above, we have

$$\begin{aligned} M(\theta) &= -\text{pr}\epsilon^2 - (\alpha - \alpha_0)^2 \text{pr}(Z \leq \zeta \wedge \zeta_0) \\ &\quad -(\beta - \alpha_0)^2 \text{pr}(\zeta < Z \leq \zeta_0) \\ &\quad -(\alpha - \beta_0)^2 \text{pr}(\zeta_0 < Z \leq \zeta) \\ &\quad -(\beta - \beta_0)^2 \text{pr}(Z > \zeta \vee \zeta_0) \\ &\leq -\text{pr}\epsilon^2 \\ &= M(\theta_0). \end{aligned}$$

Because Z has a bounded density on $[a, b]$, we obtain that M is continuous.

It is also clear that M has a unique maximum at θ_0 because the density of Z is bounded below and $\alpha_0 \neq \beta_0$ (see Exercise 14.6.5).

Now the conditions of the argmax theorem are met, and the desired consistency follows.

Rate of Convergence

We will utilize Corollary 14.5 to obtain the convergence rates via the discrepancy function

$$\tilde{d}(\theta, \theta_0) \equiv \|\gamma - \gamma_0\| + \sqrt{|\zeta - \zeta_0|}.$$

Note that this is not a norm since it does not satisfy the triangle inequality.

Nevertheless, $\tilde{d}(\theta, \theta_0) \to 0$ if and only if $\|\theta - \theta_0\| \to 0$.

Moreover, from the *Consistency* paragraph above, we have that

$$
\begin{aligned}
M(\theta) - M(\theta_0) &= -P\{Z \leq \zeta \wedge \zeta_0\}(\alpha - \alpha_0)^2 \\
&\quad -P\{Z > \zeta \vee \zeta_0\}(\beta - \beta_0)^2 \\
&\quad -P\{\zeta < Z \leq \zeta_0\}\}(\beta - \alpha_0)^2 \\
&\quad -P\{\zeta_0 < Z \leq \zeta\}(\alpha - \beta_0)^2 \\
&\leq -P\{Z < a\}(\alpha - \alpha_0)^2 \\
&\quad -P\{Z > b\}(\beta - \beta_0)^2 \\
&\quad -k_1(1 - o(1))|\zeta - \zeta_0| \\
&\leq -(k_1 \wedge \delta_1 - o(1))\tilde{d}^2(\theta, \theta_0),
\end{aligned}
$$

where

- the first inequality follows from the fact that the product of the density of Z and $(\alpha_0 - \beta_0)^2$ is bounded below by some $k_1 > 0$, and
- the second inequality follows from both pr$(Z < a)$ and pr$(Z > b)$ being bounded below by some $\delta_1 > 0$.

Thus

$$M(\theta) - M(\theta_0) \lesssim -\tilde{d}^2(\theta, \theta_0)$$

for all $\|\theta - \theta_0\|$ small enough, as desired.

Consider now the class of functions

$$\mathcal{M}_\delta \equiv \{m_\theta - m_{\theta_0} : \tilde{d}(\theta, \theta_0) < \delta\}.$$

Using previous calculations, we have

$$
\begin{aligned}
m_\theta - m_{\theta_0} &= 2(\alpha-\alpha_0)\epsilon\mathbf{1}\{Z \le \zeta \wedge \zeta_0\} + 2(\beta-\beta_0)\epsilon\mathbf{1}\{Z > \zeta \vee \zeta_0\} \\
&\quad +2(\beta-\alpha_0)\epsilon\mathbf{1}\{\zeta < Z \le \zeta_0\} + 2(\alpha-\beta_0)\epsilon\mathbf{1}\{\zeta_0 < Z \le \zeta\} \\
&\quad -(\alpha-\alpha_0)^2\mathbf{1}\{Z \le \zeta \wedge \zeta_0\} - (\beta-\beta_0)^2\mathbf{1}\{Z > \zeta \vee \zeta_0\} \\
&\quad -(\beta-\alpha_0)^2\mathbf{1}\{\zeta < Z \le \zeta_0\} - (\alpha-\beta_0)^2\mathbf{1}\{\zeta_0 < Z \le \zeta\} \\
&\equiv A_1(\theta) + A_2(\theta) + B_1(\theta) + B_2(\theta) \\
&\quad -C_1(\theta) - C_2(\theta) - D_1(\theta) - D_2(\theta). \qquad (1)
\end{aligned}
$$

Consider first A_1.

Since $\{\mathbf{1}\{Z \leq t\} : t \in [a, b]\}$ is a VC class, it is easy to compute that

$$\mathsf{E}^* \sup_{\tilde{d}(\theta,\theta_0)<\delta} |\mathbb{G}_n A_1(\theta)| \lesssim \delta,$$

as a consequence of Lemma 8.17.

Note: statements for fixed n are considered true if they hold for all n large enough.

Similar calculations apply to A_2.

Similar calculations also apply to C_1 and C_2, except that the upper bounds will be $\lesssim \delta^2$ instead of $\lesssim \delta$.

Now we consider B_1.

An envelope for the class

$$\mathcal{F} = \{B_1(\theta) : \tilde{d}(\theta, \theta_0) < \delta\}$$

is

$$F = 2(|\beta_0 - \alpha_0| + \delta)|\epsilon|\mathbf{1}\{\zeta_0 - \delta^2 < Z \leq \zeta_0\}.$$

It is not hard to verify that

$$\log N_{[]}(\eta\|F\|_{P,2}, \mathcal{F}, L_2(P)) \quad \lesssim \quad \log(1/\eta) \qquad (2)$$

(see Exercise 14.6.6).

Now Theorem 11.2 yields that

$$\mathsf{E}^* \sup_{\tilde{d}(\theta,\theta_0)<\delta} |\mathbb{G}_n B_1(\theta)| = \mathsf{E}^* \|\mathbb{G}_n\|_{\mathcal{F}} \times \|F\|_{P,2} \lesssim \delta^2.$$

Similar calculations apply also to B_2, D_1 and D_2.

Combining all of these results with the fact that $O(\delta^2) = O(\delta)$, we obtain

$$\mathsf{E}^*\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim \delta.$$

Now when $\delta \mapsto \phi(\delta) = \delta$, $\phi(\delta)/\delta^\alpha$ is decreasing for any $\alpha \in (1, 2)$.

Thus the conditions of Corollary 14.5 are satisfied with $\phi(\delta) = \delta$.

Since $r_n^2\phi(1/r_n) = r_n$, we obtain that

$$\sqrt{n}\tilde{d}(\hat{\theta}_n, \theta_0) = O_P(1).$$

By the form of $\tilde{d}$, this now implies that

$$\sqrt{n}\|\hat{\gamma}_n - \gamma_0\| = O_P(1)$$

and

$$n|\hat{\zeta}_n - \zeta_0| = O_P(1).$$

Weak Convergence

We utilize a minor modification of the argmax theorem and the rate result above to obtain that

$$\begin{aligned}
\hat{h}_n &\equiv (\sqrt{n}(\hat{\gamma}_n - \gamma_0), n(\hat{\zeta}_n - \zeta_0)) \\
&\equiv \left((\hat{h}_1, \hat{h}_2), \hat{h}_3\right) \\
&\rightsquigarrow \left((\tilde{h}_1, \tilde{h}_2), \tilde{h}_3\right) \\
&\equiv \tilde{h},
\end{aligned}$$

where

- the three components of $\tilde{h}$ are mutually independent,
- $\tilde{h}_1$ and $\tilde{h}_2$ are mean zero Gaussians
- with respective variances

$$\frac{\sigma^2}{P(Z \leq \zeta_0)} \text{ and } \frac{\sigma^2}{P(Z > \zeta_0)},$$

- and $\tilde{h}_3$ is the smallest argmax of $Q(h_3)$ for a certain, two-sided Poisson process.

We establish both finite-dimensional convergence and also tightness, but we omit the details.

Case Study: The Proportional Odds Model Under Right Censoring

In the right-censored regression set-up, we observe $X = (U, \delta, Z)$, where

- $U = T \wedge C$,
- $\delta = \mathbf{1}\{U = T\}$,
- $Z \in \mathbb{R}^d$ is a covariate vector,
- T is a failure time of interest,
- and C is a right censoring time.

We assume that C and T are independent given Z.

The proportional odds regression model stipulates that the survival function of T given Z has the form

$$S_Z(t) = \left(1 + e^{\beta' Z} A(t)\right)^{-1}, \tag{3}$$

where $t \mapsto A(t)$ is nondecreasing on $[0, \tau]$, with $A(0) = 0$ and $\tau < \infty$ being the upper limit of the censoring distribution, i.e., we assume that pr$(C > \tau) = 0$.

We also assume

- that pr$(C = \tau) > 0$,
- that var$[Z]$ is positive definite, and
- that the distribution of Z and C are uninformative of S_Z.

Let the true parameter values be denoted β_0 and A_0.

We make the additional assumptions that the support of Z is compact and that β_0 lies in a known compact $\subset \mathbb{R}^d$.

Murphy, Rossini and van der Vaart (1997) develop asymptotic theory for maximum likelihood estimation of this model under general conditions for A_0 which permit ties in the failure time distribution.

To simplify the exposition, we will make stronger assumptions on A_0 in order to facilitate arguments similar to those used in Lee (2000) and Kosorok, Lee and Fine (2004).

Specifically, we assume that A_0 has a derivative a_0 that satisfies

$$0 < \inf_{t \in [0,\tau]} a_0(t) \leq \sup_{t \in [0,\tau]} a_0(t) < \infty.$$

Let $F_Z \equiv 1 - S_Z$.

For distinct covariate values Z_1 and Z_2, we can deduce from (3) that

$$\frac{F_{Z_1}(t) S_{Z_2}(t)}{F_{Z_2}(t) S_{Z_1}(t)} = \frac{e^{\beta' Z_1}}{e^{\beta' Z_2}},$$

which justifies the "proportional odds" designation.

A motivation for this model is that in some settings it can be easier to justify on scientific grounds than other common alternatives such as the proportional hazards or accelerated failure time models (Murphy, Rossini and van der Vaart, 1997).

Define the composite model parameter $\theta \equiv (\beta, A)$.

In the following sections,

- we derive the nonparametric maximum likelihood estimator (NPMLE) $\hat{\theta}_n$,
- prove its existence,
- establish consistency,
- obtain weak convergence, and
- verify bootstrap validity for all model parameters.

Certain score and information operators will be needed for the weak convergence component, and these will be introduced just before we establish weak convergence but after we have proven consistency.

While β is assumed to lie in a known compact set, we make no such restrictions on A for estimation.

Hence the NPMLE $\hat{A}_n$ might be unbounded: for this reason, we need to verify that $\hat{A}_n(\tau) = O_P(1)$ and thus $\hat{\theta}_n$ "exists."

Nonparametric Maximum Likelihood Estimation

The likelihood for a sample of n i.i.d. observations $(U_1, \delta_1, Z_1), \ldots, (U_n, \delta_n, Z_n)$ is

$$\ell_n(\theta) = \mathbb{P}_n \left\{ \delta(\log a(U) + \beta' Z) - (1 + \delta) \log \left(1 + e^{\beta' Z} A(U) \right) \right\},$$

where a is the derivative of A.

As discussed in Murphy (1994), maximizing ℓ_n over A for fixed β results in an maximizer that is piecewise constant with jumps at the observed failure times and thus does not have a continuous density.

To address this issue, Murphy (1994) and Parner (1998) suggest replacing $a(u)$ in ℓ_n with $n\Delta A(u)$, the jump size of A at u, which modified "empirical log-likelihood" we will denote $L_n(\theta)$.

We will show later that when A_0 is continuous, the step sizes of the maximizer over A of L_n will go to zero as $n \to \infty$.

The procedure we will use to estimate θ is to maximize the profile log-likelihood

$$pL_n(\beta) \equiv \sup_A L_n(\beta, A)$$

to obtain $\hat{\theta}_n$.

The associated maximizer over A we will denote $\hat{A}_n$.

In other words, $\hat{A}_n = \hat{A}_{\hat{\beta}_n}$, where

$$\hat{A}_\beta \equiv \arg\max_A L_n(\beta, A).$$

We also define

$$\hat{\theta}_\beta \equiv (\beta, \hat{A}_\beta).$$

Obviously $\hat{\theta}_n \equiv \hat{\theta}_{\hat{\beta}_n}$ is just the joint maximizer of $L_n(\theta)$.

To characterize $\hat{A}_\beta$, consider one-dimensional submodels for A defined by the map

$$t \mapsto A_t \equiv \int_0^{(\cdot)} (1 + th_1(s))dA(s),$$

where h_1 is an arbitrary total variation bounded cadlag function on $[0, \tau]$.

The derivative of $L_n(\theta, A_t)$ with respect to t evaluated at $t = 0$ is the score function for A, $V^{\tau}_{n,2}(\theta)(h_1)$

$$\equiv \quad \mathbb{P}_n \left\{ \int_0^{\tau} h_1(s)dN(s) - (1+\delta)\left[\frac{e^{\beta' Z}\int_0^{U\wedge\tau} h_1(s)dA(s)}{1+e^{\beta' Z}A(U\wedge\tau)}\right]\right\}, \tag{4}$$

where the subscript "2" denotes that this is the score of the second parameter A.

The dependence on τ will prove useful in later sections, but, for now, it can be ignored since $P(U \leq \tau) = 1$ by assumption.

Choose $h_1(u) = \mathbf{1}\{u \leq t\}$, insert into (4), and equate the result to zero to obtain

$$\mathbb{P}_n N(t) \quad = \quad \mathbb{P}_n \left\{ \frac{(1+\delta)e^{\beta' Z} \int_0^t Y(s) d\hat{A}_\beta(s)}{1 + e^{\beta' Z} \hat{A}_\beta(U)} \right\}, \qquad (5)$$

where $N(t) \equiv \delta \mathbf{1}\{U \leq t\}$ and $Y(t) \equiv \mathbf{1}\{U \geq t\}$ are the usual counting and at-risk processes for right-censored survival data.

Next define

$$W(t;\theta) \equiv \frac{(1+\delta)e^{\beta' Z}Y(t)}{1+e^{\beta' Z}A(U)}$$

and solve (5) to obtain

$$\hat{A}_\beta(t) \quad = \quad \int_0^t \left\{\mathbb{P}_n W(s;\hat{\theta}_\beta)\right\}^{-1} \mathbb{P}_n dN(s). \qquad (6)$$

Thus $\hat{A}_\beta$ can be characterized as a stationary point of (6).

This structure will prove useful in later developments and can also be used to calculate $\hat{\theta}_n$ from data.

One approach to accomplishing this is to

- first use (6) to facilitate calculating $pL_n(\beta)$ so that $\hat{\beta}_n$ can be determined via a simple search algorithm
- and then take $\hat{A}_n$ to be the solution of (6) corresponding to $\beta = \hat{\beta}_n$.

In the case of multiple solutions, we take the one corresponding to the maximizer of $A \mapsto L_n(\hat{\beta}_n, A)$.

Existence

While we are assuming that β lies in a known, compact $\subset \mathbb{R}^d$, we are not setting boundedness restrictions on A.

Fortunately, such restrictions are not necessary, as can be seen in the following lemma, which is the contribution of this section:

LEMMA 1. *Under the given conditions,*

$$\limsup_{n\to\infty} \hat{A}_n(\tau) < \infty$$

with inner probability one.

Proof. Note that

$$\|\mathbb{P}_n N - Q_0\|_\infty \stackrel{\text{as}*}{\to} 0 \quad \text{and} \quad \sup_{u\geq 0} |(\mathbb{P}_n - P)\mathbf{1}\{U \leq u\}| \stackrel{\text{as}*}{\to} 0, \text{(7)}$$

where $Q_0 \equiv PN$.

Let $X_1, X_2, \ldots$ be a fixed data sequence satisfying (7).

Note that, without loss of generality, all data sequences will satisfy this since the probability of such a sequence is 1 by definition of outer-almost sure convergence.

Under this set-up, we can treat the maximum likelihood estimators as a sequence of fixed quantities.

By the assumed compactness of the parameter space for β, there exists a subsequence of $\{n\}$ for which $\hat{\beta}_n$ converges to a bounded vector β_* along that subsequence.

Now choose a further subsequence $\{n_k\}$ for which both

$$\hat{\beta}_{n_k} \to \beta_* \in \mathbb{R}^d$$

and

$$\hat{A}_{n_k}(\tau) \to \infty.$$

We will now work towards a contradiction.

Let $\theta_n \equiv (\beta_0, \mathbb{P}_n N)$, and note that, by definition of the NPMLE,

$$\begin{aligned} 0 &\leq L_{n_k}(\hat{\theta}_n) - L_{n_k}(\theta_n) \\ &\leq O(1) + \int_0^\tau \log(n\Delta A_{n_k}(s))\mathbb{P}_{n_k} dN(s) \\ &\quad -\mathbb{P}_{n_k}\left[(1+\delta)\log(1+\hat{A}_{n_k}(U))\right]. \end{aligned} \tag{8}$$

Let $\{u_0, u_1, \ldots, u_M\}$ be a partition of $[0, \tau]$ for some finite M, with

$$0 = u_0 < u_1 < \cdots < u_M = \tau,$$

which we will specify in more detail shortly, and define

$$N^j(s) \equiv N(s)\mathbf{1}\{U \in [u_{j-1}, u_j]\}$$

for $1 \leq j \leq M$.

Now by Jensen's inequality,

$$\begin{aligned}
&\int_0^\tau \log(n_k \Delta \hat{A}_{n_k}) \mathbb{P}_{n_k} dN^j(s) \\
&\leq \quad \mathbb{P}_{n_k} N^j(\tau) \log\left(\frac{\int_0^{u_j} n\Delta \hat{A}_{n_k}(s) d\mathbb{P}_n N^j(s)}{\mathbb{P}_{n_k} N^j(\tau)} \right) \\
&\leq \quad O(1) + \log(\hat{A}_{n_k}(u_j)) \mathbb{P}_{n_k} \delta \mathbf{1}\{U \in [u_{j-1}, u_j]\}.
\end{aligned}$$

Thus the right-side of (8) is dominated by

$$
\begin{aligned}
O(1) + \sum_{j=1}^{M-1} \log \hat{A}_{n_k}(u_j) \\
\times \mathbb{P}_{n_k}\left(\delta \mathbf{1}\{U \in [u_{j-1}, u_j]\} - (1+\delta)\mathbf{1}\{U \in [u_j, u_{j+1}]\}\right) \\
+ \log \hat{A}_{n_k}(\tau)\mathbb{P}_{n_k}\left(\delta \mathbf{1}\{U \in [u_{M-1}, \infty]\} - (1+\delta)\mathbf{1}\{U \in [\tau, \infty]\}\right).
\end{aligned}
\tag{9}
$$

Without loss of generality, assume $Q_0(\tau) > 0$.

Because of the assumptions, we can choose $u_0, u_1, \ldots, u_M$ with $M < \infty$ such that for some $\eta > 0$,

$$P\left((1+\delta)\mathbf{1}\{U \in [\tau, \infty]\}\right) \geq \eta + P\left(\delta \mathbf{1}\{U \in [u_{M-1}, \infty]\}\right)$$

and

$$P\left((1+\delta)\mathbf{1}\{U \in [u_j, u_{j+1}]\}\right) \geq \eta + P\left(\delta \mathbf{1}\{U \mathbf{1}[u_{j-1}, u_j]\}\right),$$

for all $1 \leq j \leq M-1$.

Hence

$$\text{(9)} \leq -(\eta + o(1)) \log \hat{A}_{n_k}(\tau) \rightarrow -\infty,$$

as $k \rightarrow \infty$, which yields the desired contradiction.

Thus the lemma follows since the data sequence was arbitrary.$\square$

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,* $\hat{\theta}_n \stackrel{\text{as}*}{\longrightarrow} \theta_0$.

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

$$
\begin{aligned}
L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n) &= \int_0^\tau \log \frac{PW(s;\theta_0)}{\mathbb{P}_n W(s;\hat{\theta}_n)} \mathbb{P}_n dN(s) \\
&+ (\hat{\beta}_n - \beta_0)' \mathbb{P}_n \int_0^\tau Z dN(s) \\
&- \mathbb{P}_n \left[(1+\delta) \log \left(\frac{1 + e^{\hat{\beta}_n' Z} \hat{A}_n(U)}{1 + e^{\beta_0' Z} \tilde{A}_n(U)} \right) \right] \qquad (10)
\end{aligned}
$$

By Lemma 2 below,

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

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

$$\int_0^\tau \log \frac{PW(s;\theta_0)}{\mathbb{P}_n W(s;\hat{\theta}_n)} \mathbb{P}_n dN(s) - \int_0^\tau \log \frac{PW(s;\theta_0)}{PW(s;\hat{\theta}_n)} dQ_0(s) \quad \overset{\text{as}*}{\to} \quad 0. \tag{11}$$

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

$$(\mathbb{P}_n - P)\left[(1+\delta)\log\left(\frac{1+e^{\hat{\beta}_n' Z}\hat{A}_n(U)}{1+e^{\beta_0' Z}A_0(U)}\right)\right]\overset{\text{as}*}{\to} 0. \qquad (12)$$

Now combining results (11) and (12) with (10), we obtain that

$$
\begin{aligned}
L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n) - \int_0^\tau \log \frac{PW(s;\theta_0)}{PW(s;\hat{\theta}_n)} dQ_0(s) - (\hat{\beta}_n - \beta_0)' P[Z\delta] \\
+P\left[(1+\delta)\log\left(\frac{1+e^{\hat{\beta}_n' Z}\hat{A}_n(U)}{1+e^{\beta_0' Z}A_0(U)}\right)\right] \qquad (13) \\
\overset{\text{as}*}{\longrightarrow} \quad 0.
\end{aligned}
$$

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 (13), 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.$\square$

LEMMA 2. *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.$\square$