Case Studies

Yangjianchen Xu

Department of Biostatistics
University of North Carolina at Chapel Hill

02/11/2022

Overview

- Example: The Proportional Odds Model Under Right Censoring
- Example: Temporal Process Regression

The Proportional Odds Model Under Right Censoring

The proportional odds model for right censored data has been studied extensively in Sections 15.3 and 19.2.2 and several places in Chapter 20. The main steps for estimation and inference that have been discussed for this model are:

- Developing a method of estimation.
- Establishing consistency of the estimator.
- Establishing the rates of convergence.
- Obtaining weak convergence for all regular parameters.
- Establishing efficiency of all regular parameters.
- Obtaining convergence of non-regular parameters.
- Developing a method of inference (**Theorems 19.5 and 19.6 will establish the validity of the profile sampler**).
- Studying the properties of estimation and inference under model misspecification.

- In practice, a researcher may need to iterate between several of these steps before achieving all of the desired conclusions.
- For instance, it may take a few iterations to arrive at a computationally feasible and efficient estimator, or, it may take a few iterations to arrive at the optimal rate of convergence.
- The focus of this example will be on establishing the conditions of Theorems 19.5 and 19.6 as mentioned above, since these are the only missing pieces in establishing the validity of the profile sampler for the proportional odds model under right censoring.

Conditions in Theorem 19.5

Define

$$\ell(t, \theta, \eta) \equiv l\left(t, \eta_t(\theta, \eta)\right),$$

$$\dot{\ell}(t, \theta, \eta) \equiv (\partial/(\partial t))\ell(t, \theta, \eta),$$

$$\ddot{\ell}(t, \theta, \eta) \equiv (\partial/(\partial t))\dot{\ell}(t, \theta, \eta),$$

$$\hat{\eta}_\theta \equiv \operatorname{argmax}_\eta L_n(\theta, \eta),$$

and assume that for any possibly random sequence $\tilde{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$, we have

$$\hat{\eta}_{\tilde{\theta}_n} \xrightarrow{\mathrm{P}} \eta \tag{1}$$

$$P_0 \dot{\ell}(\theta_0, \tilde{\theta}_n, \hat{\eta}_{\tilde{\theta}_n}) = o_{P_0}(\|\tilde{\theta}_n - \theta_0\| + n^{-1/2}) \tag{2}$$

Conditions in Theorem 19.6

Define $\Delta_n(\theta) \equiv n^{-1}(pL_n(\theta) - pL_n(\hat{\theta}_n))$. Assume that for every random sequence $\tilde{\theta}_n \in \Theta$,

$$\Delta_n(\tilde{\theta}_n) = o_{P_0}(1) \quad \text{implies that} \quad \tilde{\theta}_n = \theta_0 + o_{P_0}(1). \tag{3}$$

The Proportional Odds Model Under Right Censoring

Recall from Section 19.2.2 that we use A for the baseline integrated "hazard" instead of Λ and

$$A_t(\beta, A) = \int_0^{(\cdot)} \left(1 + (\beta - t)' h_0(s)\right) dA(s)$$

where $h_0(s) = \left[\sigma_{\theta_0}^{22}\right]^{-1} \sigma_{\theta_0}^{21}(\cdot)(s)$ is the least favorable direction, satisfies Conditions 19.2 and 19.3.

The Proportional Odds Model Under Right Censoring

Hence,

$$
\begin{aligned}
\dot{\ell}(t, \beta, A) &= \left(\frac{\partial}{\partial t}\right) \ell\left(t, A_t(\beta, A)\right) \\
&= \left(\frac{\partial}{\partial t}\right) \left\{\delta \left(\log \Delta A_t(\beta, A)(U) + t' Z\right)\right. \\
&\quad \left. -(1+\delta) \log\left(1 + e^{t'Z} A_t(\beta, A)(U)\right)\right\} \\
&= \int_0^{\tau} \left(Z - \frac{h_0(s)}{1 + (\beta - t)' h_0(s)}\right) \\
&\quad \times \left[dN(s) - (1+\delta) \frac{Y(s) e^{t'Z} dA_t(\beta, A)(s)}{1 + e^{t'Z} A_t(\beta, A)(U \wedge \tau)}\right]
\end{aligned}
\tag{4}
$$

and

$$
\begin{aligned}
\ddot{\ell}(t,\beta,A) &= \left(\frac{\partial}{\partial t}\right)\dot{\ell}(t,\beta,A) \\
&= -\int_0^\tau \frac{h_0^{\otimes 2}(s)}{(1+(\beta-t)'h_0(s))^2}\left[dN(s) - (1+\delta)\frac{Y(s)e^{t'Z}dA_t(\beta,A)(s)}{1+e^{t'Z}A_t(\beta,A)(U\wedge\tau)}\right] \\
&- (1+\delta)\int_0^\tau \left[Z - \frac{h_0(s)}{1+(\beta-t)'h_0(s)}\right]^{\otimes 2}\frac{Y(s)e^{t'Z}dA_t(\beta,A)(s)}{1+e^{t'Z}A_t(\beta,A)(U\wedge\tau)} \\
&+ (1+\delta)\left\{\int_0^\tau \left[Z - \frac{h_0(s)}{1+(\beta-t)'h_0(s)}\right]\frac{Y(s)e^{t'Z}dA_t(\beta,A)(s)}{1+e^{t'Z}A_t(\beta,A)(U\wedge\tau)}\right\}^{\otimes 2}
\end{aligned}
$$

The Proportional Odds Model Under Right Censoring

Hence it is easy to see that both $(t, \beta, A) \mapsto \dot{\ell}(t, \theta, A)(X)$ and $(t, \beta, A) \mapsto \ddot{\ell}(t, \theta, A)(X)$ are continuous for P_0-almost every X. Although it is tedious, it is not hard to verify that for some uniform neighborhood V of (β_0, β_0, A_0),

$$\mathcal{F}_1 \equiv \{\dot{\ell}(t, \beta, A) : (t, \beta, A) \in V\}$$

is P_0-Donsker and

$$\mathcal{F}_2 \equiv \{\ddot{\ell}(t, \beta, A) : (t, \beta, A) \in V\}$$

is P_0-Glivenko-Cantelli (Exercise 22.6.1).

Now consider any sequence $\tilde{\beta}_n \xrightarrow{\mathrm{P}} \beta_0$, and let $\hat{A}_{\tilde{\beta}_n}$ be the profile maximizer at $\beta = \tilde{\beta}_n$, i.e., $\hat{A}_{\tilde{\beta}_n} = \mathrm{argmax}_A\, L_n(A, \tilde{\beta}_n)$, where L_n is the loglikelihood as defined in Section 15.3.1. The arguments in Section 15.3.2 can be modified to verify that $\hat{A}_{\tilde{\beta}_n}(\tau)$ is asymptotically bounded in probability. Since, by definition,

$$L_n(\hat{\beta}_n, \hat{A}_n) \geq L_n(\tilde{\beta}_n, \hat{A}_{\tilde{\beta}_n}) \geq L_n(\tilde{\beta}_n, \tilde{A}_n),$$

where $\tilde{A}_n$ is as defined in Section 15.3.3, we can argue along the lines used in Section 15.3.3 to obtain that $\hat{A}_{\tilde{\beta}_n}$ is uniformly consistent for A_0.

The Proportional Odds Model Under Right Censoring

We now wish to strengthen this last result to

$$\|\hat{A}_{\tilde{\beta}_n} - A_0\|_{[0,\tau]} = O_P(n^{-1/2} + \|\tilde{\beta}_n - \beta_0\|) \tag{5}$$

for any sequence $\tilde{\beta}_n \xrightarrow{\mathrm{P}} \beta_0$. If (5) holds, then, as shown by Murphy and van der Vaart (2000) in the discussion following their Theorem 1,

$$P\dot{\ell}(\beta_0, \tilde{\beta}_n, \hat{A}_{\tilde{\beta}_n}) = o_P(n^{-1/2} + \|\tilde{\beta}_n - \beta_0\|),$$

and thus both Conditions (1) and (2) hold. Hence all of the conditions of Theorem 19.5 hold.

We now show (5). The basic idea of the proof is similar to arguments given in the proof of Theorem 3.4 in Lee (2000). Recall the definition of $V_{n,2}^{\tau}$. The derivative of $L_n(\theta, A_t)$ with respect to t evaluated at $t = 0$ is the score function for A:

$$V_{n,2}^{\tau}(\theta)(h_1) \equiv \mathbb{P}_n \left\{ \int_0^{\tau} h_1(s) dN(s) - (1+\delta) \left[\frac{e^{\beta' Z} \int_0^{U \wedge \tau} h_1(s) dA(s)}{1 + e^{\beta' Z} A(U \wedge \tau)} \right] \right\}$$

Define, for all $h \in \mathcal{H}_{\infty}^2$,

$$\tilde{D}_n(A)(h) \equiv V_{n,2}^{\tau}(\tilde{\beta}_n, A)(h), \quad D_n(A)(h) \equiv V_{n,2}^{\tau}(\beta_0, A)(h),$$

and

$$D_0(A)(h) \equiv P V_{n,2}^{\tau}(\beta_0, A)(h)$$

By definition of a maximizer, $\tilde{D}_n(\hat{A}_{\tilde{\beta}_n})(h) = 0$ and $D_0(A_0)(h) = 0$ for all $h \in \mathcal{H}_{\infty}^2$.

The Proportional Odds Model Under Right Censoring

By Lemma 13.3,

$$\sqrt{n}(\tilde{D}_n - D_0)(\hat{A}_{\tilde{\beta}_n}) - \sqrt{n}(\tilde{D}_n - D_0)\left(A_0\right) = o_P(1)$$

uniformly in $\ell^\infty\left(\mathcal{H}_1^2\right)$, where $\mathcal{H}_1^2$ is the subset of $\mathcal{H}_\infty^2$ consisting of functions of total variation ≤ 1. By the differentiability of the score function, we also have

$$\sqrt{n}(\tilde{D}_n\left(A_0\right) - D_n\left(A_0\right)) = O_P(\sqrt{n}\|\tilde{\beta}_n - \beta_0\|)$$

uniformly in $\ell^\infty\left(\mathcal{H}_1^2\right)$. Combining the previous two displays, we have

$$\begin{aligned}\sqrt{n}(D_0(\hat{A}_{\tilde{\beta}_n}) - D_0\left(A_0\right)) &= -\sqrt{n}(\tilde{D}_n(\hat{A}_{\tilde{\beta}_n}) - D_0(\hat{A}_{\tilde{\beta}_n})) \\ &= -\sqrt{n}(\tilde{D}_n - D_0)(A_0) + o_P(1) \\ &= -\sqrt{n}(D_n - D_0)(A_0) \\ &\quad + O_P(1 + \sqrt{n}\|\tilde{\beta}_n - \beta_0\|) \\ &= O_P(1 + \sqrt{n}\|\tilde{\beta}_n - \beta_0\|).\end{aligned}$$

The Proportional Odds Model Under Right Censoring

Note that $D_0(A)(h) = \int_0^\tau (\sigma^{22}_{\beta_0,A_0} h) dA(s)$, where σ^{22} is as defined in Section 15.3.4. Since $\sigma^{22}_{\beta_0,A_0}$ is continuously invertible as shown in Section 19.2.2, we have that there exists some $c > 0$ such that $\sqrt{n}(D_0(\hat{A}_{\tilde{\beta}_n}) - D_0(A_0)) \geq c \|\hat{A}_{\tilde{\beta}_n} - A_0\|_{\mathcal{H}_1^2}$. Thus (5) is satisfied.

The Proportional Odds Model Under Right Censoring

The only thing remaining to do for this example is to verify (3), after replacing $\tilde{\theta}_n$ with $\tilde{\beta}_n$ and Θ with B. For each $\beta \in B$, let

$$A_\beta \equiv \text{argmax}_A \, P\left[\frac{dP_{\beta,A}}{dP_{\beta_0,A_0}}\right]$$

where $P = P_{\beta_0,A_0}$ by definition, and define

$$\tilde{\Delta}_n(\beta) \equiv P\left[\frac{dP_{\beta,A_\beta}}{dP_{\hat{\beta}_n,\hat{A}_n}}\right] \quad \text{and} \quad \Delta_0(\beta) \equiv P\left[\frac{dP_{\beta,A_\beta}}{dP_{\beta_0,A_0}}\right]$$

Theorem 2 of Kosorok, Lee and Fine (2004) is applicable here since the proportional odds model is a special case of the odds rate model with frailty variance parameter $\gamma = 1$.

The Proportional Odds Model Under Right Censoring

Hence,

$$\sup_{\beta \in B} \left| \Delta_n(\beta) - \tilde{\Delta}_n(\beta) \right| = o_P(1)$$

The smoothness of the model now implies

$$\sup_{\beta \in B} \left| \tilde{\Delta}_n(\beta) - \Delta_0(\beta) \right| = o_P(1)$$

and thus

$$\sup_{\beta \in B} |\Delta_n(\beta) - \Delta_0(\beta)| = o_P(1)$$

As a consequence, we have for any sequence $\tilde{\beta}_n \in B$, that $\Delta_n(\tilde{\beta}_n) = o_P(1)$ implies $\Delta_0(\tilde{\beta}_n) = o_P(1)$. Hence $\tilde{\beta}_n \xrightarrow{\mathrm{P}} \beta_0$ by model identifiability, and (3) follows.

Temporal Process Regression

- In this example, we will study estimation for a varying coefficient regression model for temporal process data (Fine et al., 2004).
- Consider, for example, bone marrow transplantation studies in which the time-varying effect of a certain medication on the prevalence of graft versus host GVH disease may be of scientific interest.
- Let the outcome measure be denoted $Y(t)$, where t is restricted to a finite time interval $[l, u]$.
- In the example, $Y(t)$ is a dichotomous process indicating the presence of GVH at time t.
- More generally, we allow Y to be a stochastic process, but we require Y to have square-integrable total variation over $[l, u]$.

Temporal Process Regression

- We model the mean of the response Y at time t as a function of a k-vector of time-dependent covariates $X(t)$ and a time-dependent stratification indicator $S(t)$ as follows:

$$E[Y(t) \mid X(t), S(t) = 1] = g^{-1}\left(\beta'(t)X(t)\right), \tag{6}$$

  where the link function g is monotone, differentiable and invertible, and $\beta(t) = \{\beta_1(t), \ldots, \beta_k(t)\}'$ is a k-vector of time-dependent regression coefficients.
- For the bone marrow example, $g^{-1}(u) = e^u/(1 + e^u)$ would yield a time-indexed logistic model, with $\beta(t)$ denoting the changes in log odds ratios over time for GVH disease prevalence per unit increase in the covariates.

Temporal Process Regression

- No Markov assumption is involved here since the conditioning in (6) only involves the current time and not previous times.
- In addition to the stratification indicator $S(t)$, it is useful to include a non-missing indicator $R(t)$, for which $R(t) = 1$ if $\{Y(t), X(t), S(t)\}$ is fully observed at t, and $R(t) = 0$ otherwise.
- We assume that $Y(t)$ and $R(t)$ are independent conditionally on $\{X(t), S(t) = 1\}$.

Temporal Process Regression

- The approach we take for inference is to utilize that fact that the model only posits the conditional mean of $Y(t)$ and not the correlation structure.
- Thus we can construct "working independence" estimating equations to obtain simple, nonparametric estimators.
- The pointwise properties of these estimators follows from standard estimating equation results, but uniform properties are quite nontrivial to establish since martingale theory is not applicable here. We will use empirical process methods.

Temporal Process Regression

- The observed data consists of n independent and identically distributed copies of $\{R(t) : t \in [l, u]\}$ and $(\{Y(t), X(t), S(t)\} : R(t) = 1, t \in [l, u])$.
- We can compute an estimator $\hat{\beta}_n(t)$ for each $t \in [l, u]$ as the root of $U_n(\beta(t), t) \equiv \mathbb{P}_n A(\beta(t), t)$, where

$$A(\beta(t), t) \equiv S(t)R(t)D(\beta(t))V(\beta(t), t)\left[Y(t) - g^{-1}\left(\beta'(t)X(t)\right)\right],$$

where $D(u) \equiv \partial\left[g^{-1}\left(u'X(t)\right)\right]/(\partial u)$ is a column k-vector-valued function and $V(\beta(t), t)$ is a time-dependent and possibly data-dependent scalar-valued weight function.

Here are the specific data and estimating equation assumptions we will need:

i $(S_i, R_i, X_i, Y_i), i = 1, \ldots, n$, are i.i.d. and all component processes are cadlag. We require S, R and X to all have total variation over $[l, u]$ bounded by a fixed constant $c < \infty$, and we require Y to have total variation $\tilde{Y}$ over $[l, u]$ which has finite second moment.

ii $t \mapsto \beta(t)$ is cadlag on $[l, u]$.

iii $h \equiv g^{-1}$ and $\dot{h} = \partial h(u)/(\partial u)$ are Lipschitz continuous and bounded above and below on compact sets.

iv We require

$$\inf_{t \in [l,u]} \text{eigmin } P\left[S(t)R(t)X(t)X'(t)\right] > 0,$$

where eigmin denotes the minimum eigenvalue of a matrix.

v For all bounded $B \subset \mathbb{R}^k$, the class of random functions $\{V(b,t) : b \in B, t \in [l, u]\}$ is bounded above and below by positive constants and is BUEI and PM.

Temporal Process Regression

- The form of the estimator will depend on the form of the observed data. The estimator jumps at those M times where

$$t \mapsto (\{Y_i(t), X_i(t), S_i(t)\} : R_i(t) = 1)$$

  and $t \mapsto R_i(t)$ jumps, $i = 1, \ldots, n$.
- If $Y_i(t)$ and $X_i(t)$ are piecewise constant, then so also is $\hat{\beta}_n$. In this situation, finding $\hat{\beta}_n$ (as a process) involves solving $t \mapsto U_n(\beta(t), t)$ at these M time points.
- For most practical applications, Y and X will be either piecewise-constant or continuous, and, therefore, so will $\hat{\beta}_n$.
- In the piecewise-constant case, we can interpolate in a right-continuous manner between the M jump points, otherwise, we can smoothly interpolate between them.
- When M is large, the differences between these two approaches will be small.

Temporal Process Regression

- The bounded total variation assumptions on the data make the transition from pointwise to uniform estimation and inference both theoretically possible and practically feasible.
- In this light, we will assume hereafter that $\hat{\beta}_n$ can be computed at every value of $t \in [l, u]$.
- We will now discuss consistency, asymptotic normality, and inference based on simultaneous confidence bands.
- Several interesting examples of data analyses and simulation studies for this set-up are given in Fine, Yan and Kosorok (2004).

Temporal Process Regression: Consistency

The following theorem gives us existence and consistency of $\hat{\beta}_n$ and the above conditions:

Theorem (22.3)

Assume (6) *holds with true parameter* $\{\beta_0(t) : t \in [l, u]\}$, *where* $\sup_{t\in[l,u]} |\beta_0(t)| < \infty$. *Let* $\hat{\beta}_n = \{\hat{\beta}_n(t) : t \in [l, u]\}$ *be the smallest, in uniform norm, root of* $\{U_n(\beta(t), t) = 0 : t \in [l, u]\}$. *Then such a root exists for all n large enough almost surely, and*

$$\sup_{t\in[l,u]} \left|\hat{\beta}_n(t) - \beta_0(t)\right| \overset{as*}{\to} 0.$$

Proof. Define

$$C(\gamma,\beta,t) \equiv S(t)R(t)D(\gamma(t))V(\gamma,t)\left[Y(t) - h\left(\beta'(t)X(t)\right)\right],$$

where $\gamma, \beta \in \{\ell_c^\infty([l,u])\}^k$ and $\ell_c^\infty(H)$ is the collection of bounded real functions on the set H with absolute value $\leq c$. Let

$$\mathcal{G} \equiv \{C(\gamma,\beta,t) : \gamma,\beta \in \{\ell_c^\infty([l,u])\}^k, t \in [l,u]\}.$$

Temporal Process Regression: Consistency

The first step is to show that $\mathcal{G}$ is BUEI and PM with square-integrable envelope for each $c < \infty$. This implies that $\mathcal{G}$ is P-Donsker and hence also P-Glivenko-Cantelli. We begin by observing that the classes

$$\left\{\beta'(t)X(t) : \beta \in \{\ell_c^\infty([l,u])\}^k, t \in [l,u]\right\}$$

and

$$\left\{b'X(t) : b \in [-c,c]^k, t \in [l,u]\right\}$$

are equivalent.

Temporal Process Regression: Consistency

Next, the following lemma yields that processes with square-integrable total variation are BUEI and PM:

Lemma (22.4)

Let $\{W(t) : t \in [l, u]\}$ be a cadlag stochastic process with square-integrable total variation $\tilde{W}$, then $\{W(t) : t \in [l, u]\}$ is BUEI and PM with envelop $2\tilde{W}$.

Temporal Process Regression: Consistency

Proof. Let $\mathcal{W}$ consist of all cadlag functions $w : [l, u] \mapsto \mathbb{R}$ of bounded total variation and let $\mathcal{W}_0$ be the subset consisting of monotone increasing functions. Then the functions $w_j : \mathcal{W} \mapsto \mathcal{W}_0$, where $w_1(w)$ extracts the monotone increasing part of w and $w_2(w)$ extracts the negative of the monotone decreasing part of w, are both measurable. Moreover, for any $w \in \mathcal{W}, w = w_1(w) - w_2(w)$. Lemma 9.10 tells us that $\{w_j(W)\}$ is VC with index 2 , for both $j = 1, 2$. It is not difficult to verify that cadlag monotone increasing processes are PM (Exercise 22.6.7). Hence we can apply Part (iv) of Lemma 9.17 to obtain the desired result.

- By applying Lemma 9.17, we obtain that $\mathcal{G}$ is BUEI and PM with square-integrable envelope and hence is P-Donsker by Theorem 8.19 and the statements immediately following the theorem.

Temporal Process Regression: Consistency

The second step is to use this Donsker property to obtain existence and consistency. Accordingly, we now have for each $c < \infty$ and all $\tilde{\beta} \in \{\ell_c^\infty([l,u])\}^k$, that

$$
\begin{aligned}
& U_n(\tilde{\beta}(t), t) \\
=& \mathbb{P}_n \Big\{ C(\tilde{\beta}, \beta_0, t) - S(t)R(t)D(\tilde{\beta}(t))V(\tilde{\beta}(t), t) \\
& \times \Big[h(\tilde{\beta}'(t)X(t)) - h\left(\beta_0'(t)X(t)\right)\Big]\Big\} \\
=& -\mathbb{P}_n \Big[S(t)R(t)X(t)X'(t)\dot{h}\left(\check{\beta}'(t)X(t)\right)\dot{h}(\tilde{\beta}'(t)X(t))V(\tilde{\beta}(t), t)\Big] \\
& \times \{\tilde{\beta}(t) - \beta_0(t)\} + \epsilon_n(t)
\end{aligned}
$$

where $\check{\beta}(t)$ is on the line segment between $\tilde{\beta}(t)$ and $\beta_0(t)$ and $\epsilon_n(t) \equiv \mathbb{P}_n C(\tilde{\beta}, \beta_0, t), t \in [l, u]$.

Temporal Process Regression: Consistency

Since $\mathcal{G}$ is P-Glivenko-Cantelli and $PC(\tilde{\beta}, \beta_0, t) = 0$ for all $t \in [l, u]$ and $\tilde{\beta} \in \{\ell_c^\infty([l,u])\}^k$, we have $\sup_{t\in[l,u]} |\epsilon_n(t)| \overset{\text{as*}}{\to} 0$. By Condition (ii) and the uniform positive-definiteness assured by Condition (iv), the above results imply that $U_n(\tilde{\beta}(t), t)$ has a uniformly bounded solution $\hat{\beta}_n$ for all n large enough. Hence $\|U_n(\hat{\beta}_n(t), t)\| \geq c\|\hat{\beta}_n(t) - \beta_0(t)\| - \epsilon_n^*(t)$, where $c > 0$ does not depend on t and $\|\epsilon_n^*\|_\infty \overset{\text{as*}}{\to} 0$. This follows because $\{S(t)R(t)X(t)X(t)' : t \in [l, u]\}$ is P-Glivenko-Cantelli using previous arguments. Thus the desired uniform consistency follows.

Temporal Process Regression: Asymptotic normality

The following theorem establishes both asymptotic normality and an asymptotic linearity structure which we will utilize later for inference:

Theorem (22.5)

Under the conditions of Theorem 22.3, $\hat{\beta}_n$ is asymptotically linear with influence function $\psi(t) \equiv -[H(t)]^{-1} A(\beta_0(t), t)$, where

$$H(t) \equiv P\left[S(t)R(t)D(\beta_0(t))\, V(\beta_0(t), t)\, D'(\beta_0(t))\right]$$

and $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges weakly in $\{\ell^\infty([l, u])\}^k$ to a tight, mean zero Gaussian process $\mathcal{X}(t)$ with covariance

$$\Sigma(s, t) \equiv P\left[\mathcal{X}(s)\mathcal{X}'(t)\right] = P\left[\psi(s)\psi'(t)\right].$$

Proof. By Theorem 22.3, we have for all n large enough,

$$
\begin{aligned}
0 \equiv & \sqrt{n} U_n(\hat{\beta}_n(t), t) \\
= & \sqrt{n} \mathbb{P}_n A\left(\beta_0(t), t\right) + \sqrt{n} \mathbb{P}_n \left[A(\hat{\beta}_n(t), t) - A\left(\beta_0(t), t\right) \right] \\
= & \sqrt{n} \mathbb{P}_n A\left(\beta_0(t), t\right) + \sqrt{n} \mathbb{P}_n \left[C(\hat{\beta}_n, \beta_0, t) - C\left(\beta_0, \beta_0, t\right) \right] \\
& - \sqrt{n} \mathbb{P}_n \left[S(t) R(t) D(\hat{\beta}_n(t)) V(\hat{\beta}_n(t), t) \right. \\
& \left. \times \left\{ h(\hat{\beta}_n'(t) X(t)) - h\left(\beta_0'(t) X(t)\right) \right\} \right] \\
\equiv & \sqrt{n} \mathbb{P}_n A\left(\beta_0(t), t\right) + J_n(t) - K_n(t).
\end{aligned}
$$

Temporal Process Regression: Asymptotic normality

Since $\mathcal{G}$ is P-Donsker and since sums of Donsker classes are Donsker, and also since

$$\sup_{t\in[l,u]} P\left\{C(\hat{\beta}_n, \beta_0, t) - C(\beta_0, \beta_0, t)\right\}^2 \xrightarrow{\mathrm{P}} 0,$$

we have that $\sup_{t\in[l,u]} |J_n(t)| = o_P(1)$. By previous arguments, we also have that

$$K_n(t) = [H(t) + \epsilon_n^{**}(t)]\sqrt{n}(\hat{\beta}_n(t) - \beta_0(t)),$$

where a simple extension of previous arguments yields that $\sup_{t\in[l,u]} |\epsilon_n^{**}(t)| = o_P(1)$. This now yields the desired asymptotic linearity. The weak convergence follows since $\{A(\beta_0(t), t) : t \in [l, u]\}$ is a subset of the Donsker class $\mathcal{G}$.

Temporal Process Regression: Simultaneous confidence bands

We now utilize the asymptotic linear structure derived in the previous theorem to develop simultaneous confidence band inference. Define

$$\hat{H}_n(t) \equiv \mathbb{P}_n \left[S(t)R(t)D(\hat{\beta}_n(t))V(\hat{\beta}_n(t), t)D'(\hat{\beta}_n(t)) \right]$$

and

$$\hat{\Sigma}_n(s,t) \equiv \hat{H}_n^{-1}(s)\mathbb{P}_n \left[A(\hat{\beta}_n(s), s)A'(\hat{\beta}_n(t), t) \right] \hat{H}_n^{-1}(t)$$

and let $\hat{M}(t) = [\hat{M}_1(t), \ldots, \hat{M}_k(t)]'$ be the component-wise square root of the diagonal of $\hat{\Sigma}_n(t,t)$. Define also

$$I_n^{\circ}(t) \equiv n^{-1/2} \sum_{i=1}^{n} G_i [\text{diag}\, \hat{M}(t)]^{-1} \{\hat{H}_n(t)\}^{-1} A_i(\hat{\beta}_n(t), t)$$

where $G_1, \ldots, G_n$ are i.i.d. standard normal deviates independent of the data.

Temporal Process Regression: Simultaneous confidence bands

Now let $m_n^\circ(\alpha)$ be the $1-\alpha$ quantile of the conditional sampling distribution of $\sup_{t\in[l,u]}\|I_n^\circ(t)\|$. The next theorem establishes that $\hat{\Sigma}$ is consistent for Σ and that

$$\hat{\beta}_n(t) \quad \pm \quad n^{-1/2} m_n^\circ(\alpha)\hat{M}(t) \tag{7}$$

is a $1-\alpha$-level simultaneous confidence band for $\beta_0(t)$, simultaneous for all $t \in [u, l]$.

Theorem (22.6)

Under the conditions of Theorem 22.3, $\hat{\Sigma}(s,t)$ is uniformly consistent for $\Sigma(s,t)$, over all $s,t \in [l,u]$, almost surely. If, in addition, $\inf_{t\in[l,u]}$ eigmin $\Sigma(t,t) > 0$, then the $1-\alpha$ confidence band given in (7) is simultaneously valid asymptotically.

Temporal Process Regression: Simultaneous confidence bands

Proof. The proof of uniform consistency of $\hat{\Sigma}$ follows from minor modifications of previous arguments. Provided the minimum eigenvalue condition holds, $\hat{M}(t)$ will be asymptotically bounded both above and below uniformly over $t \in [l, u]$ and uniformly consistent for the component-wise square root of the diagonal of $\Sigma(t, t)$, which we denote $M_0(t)$. The arguments in Section 20.2.3 are applicable, and we can establish, again by recycling earlier arguments, that $I_n^{\circ}(t) \overset{\mathrm{P}}{\underset{G}{\rightsquigarrow}} M_0^{-1}(t)\mathcal{X}(t)$ in $\{\ell^{\infty}([l, u])\}^k$. The desired conclusions now follow.