TECHNICAL REPORT

Proofs of Asymptotic Results for "Maximum Likelihood Estimation in Semiparametric Transformation Models for Counting Processes"

BY DONGLIN ZENG and D. Y. LIN

This report contains the proofs for the asymptotic properties of the maximum likelihood estimators $(\widehat{\beta}_n, \widehat{\Lambda}_n)$. We conjecture that the results hold generally, but we only provide the proofs under the following set of conditions:

Condition 1. The function $\Lambda_0(t)$ is strictly increasing and continuously differentiable, and β_0 lies in the interior of a compact set $\mathcal{C}$.

Condition 2. With probability one, $Z(.)$ has bounded total variation in $[0,\tau]$. In addition, if there exists a vector γ and a deterministic function $\gamma_0(t)$ such that $\gamma_0(t) + \gamma^T Z(t) = 0$ with probability one, then $\gamma = 0$ and $\gamma_0(t) = 0$.

Condition 3. With probability one, there exists a positive constant δ such that $\text{pr}(C \geq \tau|Z) > \delta$ and $\text{pr}(\overline{Y}^*(\tau) = 1|Z) > \delta$, where $\overline{Y}^*(\tau) = 1$ means that $Y^*(t) = 1$ for all $t \in [0,\tau]$.

Condition 4. For any positive c_0, $\limsup_{x\to\infty}\{G(c_0 x)\}^{-1}\log\{x\sup_{y\leq x} G'(y)\} = 0$. This condition is satisfied by $G(x) = \{(1+x)^\rho - 1\}/\rho$ with $\rho > 0$.

Consistency. The proof consists of three steps: first, we show that the maximum likelihood estimators exist or equivalently that the jump sizes of $\widehat{\Lambda}_n$ are finite; secondly, we show that, for almost every sample, $\widehat{\Lambda}_n$ is bounded, so that by the Helly selection, along a subsequence, $\widehat{\Lambda}_n \to \Lambda^*$ weakly and $\widehat{\beta}_n \to \beta^*$; finally, we show that $\Lambda^* = \Lambda_0$ and $\beta^* = \beta_0$.

Step 1. Let $(X_{i1}, \ldots, , X_{i,n_i})$ be the ordered observed event times for the ith subject and define $X_{i0} = 0$. Let M be a constant such that $\sup_{\beta\in\mathcal{C}, t\in[0,\tau]}|\beta^T Z(t)| \leq M$ with probability one. Condition 2 implies that such a constant exists. Thus, the ith term in (4) satisfies

$$\int_0^\tau \log\Lambda\{t\}e^{\beta^T Z_i(t)}dN_i(t) + \int_0^\tau \log G'(\int_0^t Y_i(s)e^{\beta^T Z_i(s)}d\Lambda)dN_i(t) - G(\int_0^\tau Y_i(s)e^{\beta^T Z_i(s)}d\Lambda)$$

$$\leq \quad n_i G(\Lambda(\tau\wedge C_i)e^M)\left[\frac{\log\left\{\int_0^\tau Y_i(t)d\Lambda e^M \sup_{y\leq\int_0^\tau Y_i(t)d\Lambda e^M} G'(y)\right\}}{G(\int_0^\tau Y_i(t)d\Lambda e^{-M})} - \frac{1}{n_i}\right].$$

Under Condition 4, this quantity diverges to $-\infty$ if $\Lambda\{X_{ij}\}$ tends to ∞ for some X_{ij}. Thus, the jump sizes of Λ must be finite.

Step 2. We show that $\sup_n \widehat{\Lambda}_n(\tau) < \infty$ with probability one. Since $l_n(\Lambda, \beta)$ achieves its maximum at $(\widehat{\Lambda}_n, \widehat{\beta}_n)$, the following inequality holds

$$\frac{1}{n}\left\{l_n(\xi_n\overline{\Lambda}_n, \widehat{\beta}_n) - l_n(\overline{\Lambda}_n, \widehat{\beta}_n)\right\} \geq 0, \qquad (A.1)$$

where $\xi_n = \widehat{\Lambda}_n(\tau)$ and $\overline{\Lambda}_n = \widehat{\Lambda}_n/\xi_n$. To show that $\sup_n \widehat{\Lambda}_n(\tau) < \infty$ with probability one, it suffices to show that ξ_n is bounded almost surely. We prove this result by contradiction. Suppose that, for every sample in a probability set with positive probability, $\xi_n \to \infty$ for some subsequence, which we still denote by ξ_n. From (A.1), we obtain

$$\frac{1}{n}\sum_{i=1}^n \int_0^\tau \log\left\{\xi_n G'(\xi_n \int_0^t Y_i(t)e^{\widehat{\beta}_n^T Z_i(s)} d\overline{\Lambda}_n)\right\} dN_i(t) - \frac{1}{n}\sum_{i=1}^n G(\xi_n \int_0^\tau Y_i(s)e^{\widehat{\beta}_n^T Z_i(s)} d\overline{\Lambda}_n)$$
$$\geq \frac{1}{n}\sum_{i=1}^n \int_0^\tau \log G'(\int_0^t Y_i(s)e^{\widehat{\beta}_n^T Z_i(s)} d\overline{\Lambda}_n) dN_i(t) - \frac{1}{n}\sum_{i=1}^n G(\int_0^\tau Y_i(s)e^{\widehat{\beta}_n^T Z_i(s)} d\overline{\Lambda}_n).$$

Note that the right-hand side is bounded from below by

$$\log \min_{y \leq e^M} G'(y) \left\{\frac{1}{n}\sum_{i=1}^n N_i(\tau)\right\} - G(e^M) > -\infty.$$

However, the left-hand side is bounded from above by

$$\frac{1}{n}\sum_{i=1}^n \int_0^{\tau\wedge C_i} dN_i(t) \log \xi_n \sup_{y \leq \xi_n e^M} G'(y) - \frac{1}{n}\sum_{i=1}^n I(\overline{Y}_i^*(\tau) = 1, C_i \geq \tau) G(e^{-M}\xi_n).$$

Under Condition 4, $\log \xi_n \sup_{y \leq \xi_n e^M} G'(y) \leq \epsilon G(\xi_n e^{-M})$ for any ϵ when n is large enough. Thus,

$$\left\{\frac{\epsilon}{n}\sum_{i=1}^n N_i(\tau) - \frac{1}{n}\sum_{i=1}^n I(\overline{Y}_i^*(\tau) = 1, C_i \geq \tau)\right\} G(\xi_n e^{-M}) > -\infty.$$

If we choose ϵ such that $\epsilon E[N(\tau)] \leq \mathrm{pr}(\overline{Y}^*(\tau) = 1, C \geq \tau)/2$, the left-hand side diverges to $-\infty$ when $\xi_n \to \infty$. This is a contradiction. Therefore, $\widehat{\Lambda}_n$ is bounded with probability one. By the Helly selection, along a subsequence, we assume that $\widehat{\Lambda}_n \to \Lambda^*$ weakly and $\widehat{\beta}_n \to \beta^*$.

Step 3. We show that $\Lambda^* = \Lambda_0$ and $\beta^* = \beta_0$. By differentiating $l_n(\Lambda, \beta)$ with respect to $\Lambda\{X_{ij}\}$ and setting the derivative be zero, we obtain

$$\frac{1}{n\widehat{\Lambda}_n\{X_{ij}\}} = \phi_n(X_{ij}; \widehat{\Lambda}_n, \widehat{\beta}_n),$$

where

$$\phi_n(s; \widehat{\Lambda}_n, \widehat{\beta}_n) = \frac{1}{n}\sum_{k=1}^n G'(\int_0^\tau Y_k(t)e^{\widehat{\beta}_n^T Z_k(t)} d\widehat{\Lambda}_n) e^{\widehat{\beta}_n^T Z_k(s)} Y_k(s)$$

$$-\frac{1}{n}\sum_{k=1}^{n}\int_0^\tau \frac{I(t\geq s)Y_k(s)e^{\widehat{\beta}^T Z_k(s)}G''(\int_0^t Y_k(\widetilde{s})e^{\widehat{\beta}^T Z_k(\widetilde{s})}d\widehat{\Lambda}_n)}{G'(\int_0^t Y_k(\widetilde{s})e^{\widehat{\beta}^T Z_k(\widetilde{s})}d\widehat{\Lambda}_n)}dN_k(t).$$

It follows immediately that

$$\widehat{\Lambda}_n(t)=\int_0^t \frac{\sum_{i=1}^n dN_i(s)/n}{|\phi_n(s;\widehat{\Lambda}_n,\widehat{\beta}_n)|}. \qquad (A.2)$$

By the Glivenko-Cantelli theorem, $\phi_n(t;\widehat{\Lambda}_n,\widehat{\beta}_n)$ uniformly converges to a continuously differentiable function $\phi^*(s;\Lambda^*,\beta^*)$. We show that $\min_{s\in[0,\tau]}|\phi^*(s;\Lambda^*,\beta^*)|\geq 2\epsilon_0$ for some positive constant ϵ_0 by contradiction. If this inequality does not hold, then $\phi^*(s_0;\Lambda^*;\beta^*)=0$ for some $s_0\in[0,\tau]$. It follows from (A.2) that, for any $\epsilon>0$,

$$\widehat{\Lambda}_n(\tau)\geq\int_0^\tau \frac{\sum_{i=1}^n dN_i(s)/n}{|\phi_n(s;\widehat{\Lambda}_n,\widehat{\beta}_n)|+\epsilon}\to E\left[\int_0^\tau \frac{dN(s)}{|\phi^*(s;\Lambda^*,\beta^*)|+\epsilon}\right].$$

Letting ϵ decrease to zero, we obtain

$$E\left[\int_0^\tau \frac{dN(s)}{|\phi^*(s;\Lambda^*,\beta^*)|}\right]<\infty.$$

However, $|\phi^*(s;\Lambda^*,\beta^*)| = |\phi^*(s;\Lambda^*,\beta^*)-\phi^*(s_0;\Lambda^*,\beta^*)| \leq c_1|s-s_0|$ for some constant c_1 and $\int_0^\tau |s-s_0|^{-1}E[dN(s)]=\infty$. This is a contradiction. Thus, when n is large enough, $|\phi_n(t;\widehat{\Lambda}_n,\widehat{\beta}_n)|>\epsilon_0>0$ for some constant ϵ_0.

By replacing $\widehat{\Lambda}_n$ and $\widehat{\beta}_n$ in (A.2) with Λ_0 and β_0, we obtain

$$\widetilde{\Lambda}_n(t)=\int_0^t \frac{\sum_{i=1}^n dN_i(s)/n}{|\phi_n(s;\Lambda_0,\beta_0)|}. \qquad (A.3)$$

If follows from the Glivenko-Cantelli theorem together with simple algebra that the right-hand side of (A.3) uniformly converges to Λ_0 almost surely. By (A.2) and (A.3) and the lower bound of $|\phi_n|$, $\widehat{\Lambda}_n(t)$ is absolutely continuous respect to $\widetilde{\Lambda}_n(t)$ and $d\widehat{\Lambda}_n/d\widetilde{\Lambda}_n$ converges to a bounded measurable function $\psi(t)$. That is, $\Lambda^*(t)=\int_0^t\psi(s)d\Lambda_0(t)$. Thus, $\Lambda^*(t)$ is absolutely continuous with respect to the Lebsgue measure and we denote its derivative as $\lambda^*(t)$. In addition, $\psi(t)=\lambda^*(t)/\lambda_0(t)$. Finally, since $l_n(\Lambda,\beta)$ is maximized at $(\widehat{\Lambda}_n,\widehat{\beta}_n)$,

$$\frac{1}{n}\sum_{i=1}^n\left[\int_0^\tau \log\frac{\widehat{\Lambda}_n\{t\}}{\widetilde{\Lambda}_n(t)}dN_i(t)-G(\int_0^\tau Y_i(t)e^{\widehat{\beta}_n^T Z_i(t)}d\widehat{\Lambda}_n)+G(\int_0^\tau Y_i(t)e^{\beta_0^T Z_i(t)}d\widetilde{\Lambda}_n)\right.$$
$$+\int_0^\tau \log G'(\int_0^t Y_i(s)e^{\widehat{\beta}_n^T Z_i(s)}d\widehat{\Lambda}_n)dN_i(t)+\int_0^\tau e^{\widehat{\beta}_n^T Z_i(t)}dN_i(t)$$
$$\left.-\int_0^\tau \log G'(\int_0^t Y_i(s)e^{\beta_0^T Z_i(s)}d\widetilde{\Lambda}_n)dN_i(t)-\int_0^\tau e^{\beta_0^T Z_i(t)}dN_i(t)\right]\geq 0.$$

We take the limits on both sides. By the Glivenko-Cantelli theorem and the fact that $\widehat{\Lambda}_n\{t\}/\widetilde{\Lambda}\{t\}$ converges uniformly to $\lambda^*(t)/\lambda_0(t)$, the Kullback-Leibler information between the density indexed by (Λ^*, β^*) and the true density is negative. Therefore, with probability one,

$$\begin{aligned} & \int_0^\tau \log\left\{Y(t)\lambda^*(t)e^{\beta^{*T}Z(t)}G'(\int_0^t Y(s)e^{\beta^{*T}Z(s)}d\Lambda^*)\right\}dN(t) - G(\int_0^\tau Y(t)e^{\beta^{*T}Z(t)}d\Lambda) \\ = & \int_0^\tau \log\left\{Y(t)\lambda_0(t)e^{\beta_0^T Z(t)}G'(\int_0^t Y(s)e^{\beta_0^T Z(s)}d\Lambda_0)\right\}dN(t) - G(\int_0^\tau Y(t)e^{\beta_0^T Z(t)}d\Lambda_0). \end{aligned}$$

This equality holds for the case in which $\overline{Y}^*(\tau) = 1$, $N^*(\tau) = 0$ and $C \geq \tau$ and also holds for the case in which $\overline{Y}^*(\tau) = 1$, $N^*(t-) = 0$, $N^*(\tau) = 1$ and $C \geq \tau$. The difference between the equalities from these two cases entails that

$$\lambda^*(t)e^{\beta^{*T}Z(t)}G'(\int_0^t e^{\beta^{*T}Z(s)}d\Lambda^*) = \lambda_0(t)e^{\beta_0^T Z(t)}G'(\int_0^t e^{\beta_0^T Z(s)}d\Lambda_0).$$

Integrating from 0 to t yields

$$G(\int_0^t e^{\beta^{*T}Z(s)}d\Lambda^*) = G(\int_0^t e^{\beta_0^T Z(s)}d\Lambda_0).$$

Thus,

$$\int_0^t e^{\beta^{*T}Z(s)}d\Lambda^* = \int_0^t e^{\beta_0^T Z(s)}d\Lambda_0.$$

If then follows from Condition 2 that $\beta^* = \beta_0$ and $\Lambda^* = \Lambda_0$.

Hence, we have proved that $\widehat{\beta}_n \longrightarrow \beta_0$ and $\widehat{\Lambda}_n(t) \longrightarrow \Lambda_0(t)$ almost surely. The latter can be strengthened to uniform convergence in $t \in [0, \tau]$ by the continuity of Λ_0.

Asymptotic distribution. We denote the empirical measure determined by n i.i.d. observations as $\mathcal{P}_n$ and denote its expectation as $\mathcal{P}$. Let $\mathcal{G}_n$ be the empirical process given by $\sqrt{n}(\mathcal{P}_n - \mathcal{P})$. In addition, we define $l(\Lambda, \beta)$ as the logarithm of the observed likelihood function from a single subject and define its derivative with respect to Λ as

$$l_\Lambda(\Lambda, \beta)[\Delta\Lambda] = \lim_{\epsilon \longrightarrow 0} \frac{l(\Lambda + \epsilon\Delta\Lambda, \beta) - l(\Lambda, \beta)}{\epsilon}.$$

We also define

$$l_{\Lambda\Lambda}(\Lambda, \beta)[\Delta_1\Lambda, \Delta_2\Lambda] = \lim_{\epsilon \longrightarrow 0} \frac{l_\Lambda(\Lambda + \epsilon\Delta_2\Lambda, \beta)[\Delta_1\Lambda] - l_\Lambda(\Lambda, \beta)[\Delta_1\Lambda]}{\epsilon}.$$

Likewise, $l_\beta(\Lambda, \beta)$ denotes the score vector for β and $l_{\beta\beta}(\Lambda, \beta)$ the Hessian matrix of $l(\Lambda, \beta)$ with respect to β. For convenience, we define

$$\Psi(t; \Lambda, \beta) = G''(\int_0^t Y(s)e^{\beta^T Z(s)}d\Lambda)/G'(\int_0^t Y(s)e^{\beta^T Z(s)}d\Lambda),$$

$$\widetilde{\Psi}(\Lambda,\beta) = G''(\int_0^\tau Y(t)e^{\beta^T Z(t)}d\Lambda).$$

We choose ϵ_0 small enough and define a map $W_n := (W_{n1}, W_{n2})$ from $\{(\Lambda,\beta) : \|\Lambda - \Lambda_0\|_{l^\infty[0,\tau]} < \epsilon_0, |\beta-\beta_0| < \epsilon_0\} \subset l^\infty(\mathcal{Q}) \times \mathcal{R}^p$ to $l^\infty(\mathcal{Q}) \times \mathcal{R}^p$ as follows: for any $q(t) \in \mathcal{Q}$,

$$\begin{aligned} W_{n1}(\Lambda,\beta)[q] &= \frac{d}{d\delta}\mathcal{P}_n\left\{l(\Lambda(t) + \delta\int_0^t q(s)d\Lambda,\beta)\right\}\Big|_{\delta=0} \\ &= \mathcal{P}_n\left\{\int_0^\tau q(t)dN(t) + \int_0^\tau \Psi(t;\Lambda,\beta)\int_0^t Y(s)q(s)e^{\beta^T Z(s)}d\Lambda dN(t) \right. \\ &\quad \left. - \int_0^\tau Y(t)e^{\beta^T Z(t)}q(t)d\Lambda G'(\int_0^\tau Y(t)e^{\beta^T Z(t)}d\Lambda)\right\}, \end{aligned}$$

and

$$\begin{aligned} W_{n2}(\Lambda,\beta) &= \nabla_\beta \mathcal{P}_n\left\{l(\Lambda,\beta)\right\} \\ &= \mathcal{P}_n\left\{\int_0^\tau \Psi(t;\Lambda,\beta)\int_0^t Y(s)e^{\beta^T Z(s)}Z(s)d\Lambda dN(t) + \int_0^\tau Z(t)dN(t) \right. \\ &\quad \left. -G'(\int_0^\tau Y(t)e^{\beta^T Z(t)}d\Lambda)\int_0^\tau Y(t)e^{\beta^T Z(t)}Z(t)d\Lambda\right\}. \end{aligned}$$

Likewise, we can define the limit version of W_n as $W := (W_1, W_2)$ by replacing $\mathcal{P}_n$ with $\mathcal{P}$ in the above two definitions. Clearly, $W_n(\widehat{\Lambda}_n, \widehat{\beta}_n) = 0$ and $W(\Lambda_0,\beta_0) = 0$. By Conditions 1-2 and the Donsker theorem, $\sqrt{n}(W_n - W)(\widehat{\Lambda}_n,\widehat{\beta}_n) - \sqrt{n}(W_n - W)(\Lambda_0,\beta_0) = o_p(1)$ in the metric space $l^\infty(\mathcal{Q}) \times \mathcal{R}^p$. In light of Theorem 3.3.1 of van der Vaart and Wellner (1996), it remains to verify that W is Fréchet-differentiable at (Λ_0,β_0) and that the derivative is continuously invertible in the set $\mathcal{A} = \{(\Lambda - \Lambda_0, \beta-\beta_0) : \|\Lambda-\Lambda_0\|_{l^\infty[0,\tau]} < \epsilon_0, |\beta-\beta_0| < \epsilon_0\}$. The Fréchet-differentiability of W can be checked directly.

To verify the invertibility of the derivative, we note that the derivative of $\mathcal{W}$ maps $\mathcal{A}$ to $l^\infty(\mathcal{Q}) \times \mathcal{R}^p$ and has the form

$$\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}\begin{pmatrix} \Lambda-\Lambda_0 \\ \beta-\beta_0 \end{pmatrix}\left[\begin{pmatrix} q \\ b \end{pmatrix}\right] = \begin{pmatrix} W_{11}(\Lambda-\Lambda_0)[q] + W_{12}(\beta-\beta_0)[q] \\ W_{21}(\Lambda-\Lambda_0)^T b + W_{22}(\beta-\beta_0)^T b \end{pmatrix}.$$

In addition,

$$\begin{aligned} &W_{11}(\Lambda-\Lambda_0)[q] = \int(-p(t)I + K)[q]d(\Lambda-\Lambda_0), \\ &W_{12}(\beta-\beta_0)[q] = A[\int q d\Lambda_0](\beta-\beta_0), \\ &W_{21}(\Lambda-\Lambda_0) = A^*[\Lambda-\Lambda_0], \\ &W_{22}(\beta-\beta_0) = B(\beta-\beta_0), \end{aligned}$$

where $p(t) > 0$, I is identity operator, A and K are both linear operators, A^* is the dual operator of A, and B is $p \times p$ matrix. Specifically,

$$\begin{aligned}
p(t) &= E\left\{Y(t)e^{\beta_0^T Z(t)} G'(\int_0^\tau Y(s)e^{\beta_0^T Z(s)} d\Lambda_0)\right\} \\
&\quad -E\left\{Y(t)e^{\beta_0^T Z(t)} \int_t^{\tau\wedge C} \Psi(s;\Lambda_0,\beta_0) dN(s)\right\} \\
&= E\left\{Y(t)e^{\beta_0^T Z(t)} G'(\int_0^t Y(s)e^{\beta_0^T Z(s)} d\Lambda_0)\right\}, \\
K[q] &= -E\left\{Y(t)e^{\beta_0^T Z(t)} \widetilde{\Psi}(\Lambda_0,\beta_0) \int_0^\tau Y(s)e^{\beta_0^T Z(s)} q(s) d\Lambda_0\right\} \\
&\quad +E\left\{Y(t)e^{\beta_0^T Z(t)} \int_t^{\tau\wedge C} \Psi'(s;\Lambda_0,\beta_0) \int_0^s q(\widetilde{s})Y(\widetilde{s})e^{\beta_0^T Z(\widetilde{s})} d\Lambda_0 dN(s)\right\}, \\
A[\int q d\Lambda_0] &= E\left[\int_0^\tau \Psi'(t;\Lambda_0,\beta_0) \int_0^t Y(s)e^{\beta_0^T Z(s)} q(s) d\Lambda_0 \int_0^t Y(s)e^{\beta_0^T Z(s)} Z(s) d\Lambda_0\right] \\
&\quad +E\left[\int_0^\tau \Psi(t;\Lambda_0,\beta_0) \int_0^t Y(s)e^{\beta_0^T Z(s)} q(s)Z(s) d\Lambda_0\right] \\
&\quad -E\left[\int_0^\tau Y(t)e^{\beta_0^T Z(t)} q(t) d\Lambda_0 \int_0^\tau Y(t)e^{\beta_0^T Z(t)} Z(t) d\Lambda_0 \widetilde{\Psi}(\Lambda_0,\beta_0)\right] \\
&\quad -E\left[\int_0^\tau Y(t)e^{\beta_0^T Z(t)} q(t)Z(t) d\Lambda_0 G'(\int_0^\tau Y(t)e^{\beta_0^T Z(t)} d\Lambda_0)\right], \\
B &= E\left[\int_0^\tau \Psi(t;\Lambda_0,\beta_0) \int_0^t Y(s)e^{\beta_0{}^T Z(s)} Z(s)Z(s)^T d\Lambda_0 dN(t)\right] \\
&\quad +E\left[\int_0^\tau \Psi'(t;\Lambda_0,\beta_0) \left\{\int_0^t Y(s)e^{\beta_0{}^T Z(s)} Z(s) d\Lambda_0\right\}^{\otimes 2} dN(t)\right] \\
&\quad -E\left[G'(\int_0^\tau Y(t)e^{\beta_0^T Z(t)} d\Lambda_0) \int_0^\tau Y(t)e^{\beta_0{}^T Z(t)} Z(t)Z(t)^T d\Lambda_0\right] \\
&\quad -E\left[\widetilde{\Psi}(\Lambda_0,\beta_0) \left\{\int_0^\tau Y(t)e^{\beta_0{}^T Z(t)} Z(t) d\Lambda_0\right\}^{\otimes 2}\right].
\end{aligned}$$

Thus, to show the invertibility of $\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$, it suffices to show that W_{22} and $V := W_{11} - W_{12}W_{22}^{-1}W_{21}$ are continuously invertible.

We first show that W_{22} is invertible. Note that $-W_{22}$ is the information at β_0 for the densities with parameters (Λ_0, β), so that it is non-negative. If there exists some $b \in \mathcal{R}^p$ such that $b^T W_{22} b = 0$, then the score for β along the direction b should be zero with probability one, or

$$\begin{aligned}
0 &= \left\{\int_0^\tau \Psi(t;\Lambda_0,\beta_0) \int_0^t Y(s)e^{\beta_0{}^T Z(s)} Z(s) d\Lambda_0 dN(t)\right. \\
&\quad \left.+\int_0^\tau Z(t)dN(t) - G'(\int_0^\tau Y(t)e^{\beta_0{}^T Z(t)} d\Lambda_0) \int_0^\tau Y(t)e^{\beta_0{}^T Z(t)} Z(t) d\Lambda_0\right\}^T b.
\end{aligned}$$

The equality holds when $\overline{Y}^*(\tau) = 1$, $N(\tau) = 0$ and $C \geq \tau$, and also holds when $\overline{Y}^*(\tau) = 1$, $C \geq \tau$ and $N(\cdot)$ has only one jump at t. The comparison of the equalities from these two cases yields

$$Z(t)^T b = -\int_0^t e^{\beta_0^T Z(s)} Z(s)^T b d\Lambda_0 \Psi(t;\Lambda_0,\beta_0).$$

This can be regarded as a homogeneous integral equation for the function $Z(t)^T b$. We thus conclude $Z(t)^T b = 0$ for all $\in [0,\tau]$. Condition 2 then entails that $b = 0$.

Next we show that the operator V is invertible. Note that

$$V[\Lambda - \Lambda_0](h) = \int_0^\tau \left\{-p(t)I + \widetilde{K}\right\}[q] d(\Lambda - \Lambda_0),$$

where $\widetilde{K}$ is an integral operator of $q(t)$. If we can show that $\widetilde{\epsilon}\mathcal{Q} \subset \left\{-p(t)I + \widetilde{K}\right\}(\mathcal{Q})$ for some constant $\widetilde{\epsilon}$, then V is continuously invertible on its image in $l^\infty(\mathcal{Q})$. However, $\widetilde{K}$ is a compact operator, so that the previous condition is equivalent to that $-p(t)I + \widetilde{K}$ is one to one; that is, if some function $q \in \mathcal{Q}^p$ satisfies $\left\{-p(t)I + \widetilde{K}\right\}[q] = 0$, then $q = 0$. To prove this, we note that the following equality holds for any (Λ,β) and (q,b),

$$\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} \Lambda - \Lambda_0 \\ \beta - \beta_0 \end{pmatrix} \left[\begin{pmatrix} q \\ b \end{pmatrix}\right] = -\mathcal{P} \begin{pmatrix} l_{\Lambda\Lambda}(\Lambda_0,\beta_0) & l_{\Lambda\beta}(\Lambda_0,\beta_0) \\ l_{\beta\Lambda}(\Lambda_0,\beta_0) & l_{\beta\beta}(\Lambda_0,\beta_0) \end{pmatrix} \left[\begin{pmatrix} \Lambda - \Lambda_0 \\ \beta - \beta_0 \end{pmatrix}, \begin{pmatrix} \int q d\Lambda_0 \\ b \end{pmatrix}\right].$$

Thus, if there exists some q such that $\left\{-p(t)I + \widetilde{K}\right\}[q] = 0$, then in the above equation, we let

$$\Lambda(t) - \Lambda_0(t) = \int_0^t q d\Lambda_0, \quad b = \beta - \beta_0 = -W_{22}^{-1} W_{21}[\int_0^t q d\Lambda_0].$$

The left-hand side is equal to $V[\int q d\Lambda_0](q)$, which is zero. By the fundamental equality $E[l_{\theta\theta}] = -E[l_\theta l_\theta^T]$, the right-hand side is equal to

$$E\left[\left\{l_\Lambda(\Lambda_0,\beta_0)[\int q d\Lambda_0] + l_\beta(\Lambda_0,\beta_0)^T b\right\}^2\right].$$

Thus, there exists some $b \in \mathcal{R}^p$ such that the score function along the path $(\Lambda_0 + \delta \int q d\Lambda_0, \beta_0 + b)$ is zero. This gives that

$$\begin{aligned} 0 \;=\; & \Big[\int_0^\tau q(t) dN(t) + \int_0^\tau \Psi(t;\Lambda_0,\beta_0) \int_0^t Y(s) e^{\beta_0^T Z(s)} q(s) d\Lambda_0 dN(t) \\ & -G'(\int_0^\tau Y(t) e^{\beta_0^T Z(t)} d\Lambda_0) \int_0^\tau Y(t) e^{\beta_0^T Z(t)} q(t) d\Lambda_0\Big] \\ & + \Big[\int_0^\tau \Psi(t;\Lambda_0,\beta_0) \int_0^t Y(s) e^{\beta_0^T Z(s)} Z(s) d\Lambda_0 dN(t) + \int_0^\tau Z(t) dN(t) \\ & -G'(\int_0^\tau Y(t) e^{\beta_0^T Z(t)} d\Lambda_0) \int_0^\tau Y(t) e^{\beta_0^T Z(t)} Z(t) d\Lambda_0\Big]^T b. \end{aligned}$$

For the case of $\overline{Y}^*(\tau) = 1$, $N(\tau) = 0$ and $C \geq \tau$ and for the case of $\overline{Y}^*(\tau) = 1$, $N(t) = I(t \geq t_0)$ and $C \geq \tau$, we obtain two equalities. By taking the difference, we obtain that

$$\left\{q(t_0) + Z(t_0)^T b\right\} + \Psi(t_0;\Lambda_0,\beta_0) \int_0^{t_0} \left\{q(s) + Z(s)^T b\right\} e^{\beta_0^T Z(s)} d\Lambda_0 = 0.$$

Again, this is a homogeneous equation for $q(t) + Z(t)^T b$ with only trivial solutions. Thus, $q(t) + Z(t)^T b = 0$ for all $t \in [0, \tau]$. It follows from Condition 2 that $b = 0$ and $q(t) = 0$. Therefore, V is invertible.

It now follows from Theorem 3.3.1 of van der Vaart and Wellner (1996) that, in the metric space $l^\infty(\mathcal{Q}) \times \mathcal{R}^p$, $\sqrt{n}(\widehat{\Lambda}_n - \Lambda_0, \widehat{\beta}_n - \beta_0)$ weakly converges to some Gaussian process. Furthermore,

$$\sqrt{n}\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}\begin{pmatrix} \widehat{\Lambda}_n - \Lambda_0 \\ \beta_n - \beta_0 \end{pmatrix}\left[\begin{pmatrix} q \\ b \end{pmatrix}\right] = \mathcal{G}_n\left\{l_\Lambda[\int q d\Lambda_0] + l_\beta^T b\right\} + o_p(1).$$

The left-hand side of the equation can be written as

$$\sqrt{n}\left\{\int \sigma_1(q,b) d(\widehat{\Lambda}_n - \Lambda_0) + \sigma_2(q,b)^T(\widehat{\beta}_n - \beta_0)\right\},$$

where σ_1 is a linear map from $\mathcal{Q} \times \mathcal{R}^p$ to $l^\infty[0,\tau]$, and σ_2 is a linear map from $\mathcal{Q} \times \mathcal{R}^p$ to $\mathcal{R}^p$. The invertibility of $\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$ implies the invertibility of the map (σ_1, σ_2). Thus, if we choose q such that $\sigma_1(q,b) = 0$ and $\sigma_2(q,b) = b$, then

$$\sqrt{n}(\widehat{\beta}_n - \beta_0)^T b = \mathcal{G}_n\left\{l_\Lambda[\int q d\Lambda_0] + l_\beta^T b\right\} + o_p(1).$$

We conclude that $\widehat{\beta}_n$ is an asymptotically linear estimator for β_0 and that its influence function is on the space spanned by the score functions. Thus, $\widehat{\beta}_n$ is semiparametrically efficient.

Consistency of covariance estimators. The above proof implies that

$$-\mathcal{P}\begin{pmatrix} l_{\Lambda\Lambda}(\Lambda_0,\beta_0) & l_{\Lambda\beta}(\Lambda_0,\beta_0) \\ l_{\beta\Lambda}(\Lambda_0,\beta_0) & l_{\beta\beta} \end{pmatrix}\left[\begin{pmatrix} \sqrt{n}(\widehat{\Lambda}_n - \Lambda_0) \\ \sqrt{n}(\widehat{\beta}_n - \beta_0) \end{pmatrix}, \begin{pmatrix} \int_0^t q d\Lambda_0 \\ b \end{pmatrix}\right]$$

$$= \mathcal{G}_n\begin{pmatrix} l_\Lambda(\Lambda_0,\beta_0)[\int_0^t q d\Lambda_0] \\ l_\beta^T b \end{pmatrix} + o_p(1).$$

This approximation holds uniformly for q with bounded variation and b with bounded norm. We define a function $\widetilde{\Lambda}(t)$ as a step function with jumps at the observed event times X_{ij} and the jump size at X_{ij} is equal to $\Lambda_0(X_{ij}) - \max_{X_{kl} < X_{ij}} \Lambda_0(X_{kl})$. Clearly, $\widetilde{\Lambda}(X_{ij}) = \Lambda_0(X_{ij})$. For any bounded vector $\{p_{ij}, i = 1, \ldots, , n, j = 1, \ldots, , n_i\}$ and bounded vector $b \in \mathcal{R}^p$, we define a step function $p(t)$ such that it only jumps at X_{ij} and $p(X_{ij}) = p_{ij}$ and define $\vec{\Delta}$ as the vector consisting of $p_{ij}\widehat{\Lambda}_n\{X_{ij}\}$. By the definition of $\mathcal{I}_n$,

$$(\vec{\Delta}, b)'\mathcal{I}_n\begin{pmatrix} \vec{\Delta} \\ b \end{pmatrix} = -\mathcal{P}_n\begin{pmatrix} l_{\Lambda\Lambda}(\widehat{\Lambda}_n,\widehat{\beta}_n) & l_{\Lambda\beta}(\widehat{\Lambda}_n,\widehat{\beta}_n) \\ l_{\beta\Lambda}(\widehat{\Lambda}_n,\widehat{\beta}_n) & l_{\beta\beta} \end{pmatrix}\left[\begin{pmatrix} \int_0^t p d\widehat{\Lambda}_n \\ b \end{pmatrix}, \begin{pmatrix} \int_0^t p d\widehat{\Lambda}_n \\ b \end{pmatrix}\right].$$

The right-hand side approximates

$$-\mathcal{P}\begin{pmatrix} l_{\Lambda\Lambda}(\Lambda_0,\beta_0) & l_{\Lambda\beta}(\Lambda_0,\beta_0) \\ l_{\beta\Lambda}(\Lambda_0,\beta_0) & l_{\beta\beta} \end{pmatrix}\left[\begin{pmatrix} \int_0^t p d\Lambda_0 \\ b \end{pmatrix}, \begin{pmatrix} \int_0^t p d\Lambda_0 \\ b \end{pmatrix}\right] > 0$$

uniformly in any bounded function $p(t)$ and b. It follows immediately that $\mathcal{I}_n$ is positive definite when n is large.

On the other hand,

$$
\begin{aligned}
&-\sqrt{n}\begin{pmatrix} \left\{\widehat{\Lambda}_n\{X_{ij}\} - \widetilde{\Lambda}\{X_{ij}\}\right\} \\ \widehat{\beta}_n - \beta_0 \end{pmatrix} \mathcal{I}_n \begin{pmatrix} \vec{\Delta} \\ \underline{\mathrm{b}} \end{pmatrix} \\
=\; & -\sqrt{n}\mathcal{P}_n \begin{pmatrix} l_{\Lambda\Lambda}(\widehat{\Lambda}_n, \widehat{\beta}_n) & l_{\Lambda\beta}(\widehat{\Lambda}_n, \widehat{\beta}_n) \\ l_{\beta\Lambda}(\widehat{\Lambda}_n, \widehat{\beta}_n) & l_{\beta\beta} \end{pmatrix} \left[\begin{pmatrix} \widehat{\Lambda}_n(t) - \widetilde{\Lambda}(t) \\ \widehat{\beta}_n - \beta_0 \end{pmatrix}, \begin{pmatrix} \int_0^t p d\widehat{\Lambda}_n \\ b \end{pmatrix} \right] \\
=\; & -\sqrt{n}\mathcal{P}_n \begin{pmatrix} l_{\Lambda\Lambda}(\widehat{\Lambda}_n, \widehat{\beta}_n) & l_{\Lambda\beta}(\widehat{\Lambda}_n, \widehat{\beta}_n) \\ l_{\beta\Lambda}(\widehat{\Lambda}_n, \widehat{\beta}_n) & l_{\beta\beta} \end{pmatrix} \left[\begin{pmatrix} \widehat{\Lambda}_n(t) - \Lambda_0(t) \\ \widehat{\beta}_n - \beta_0 \end{pmatrix}, \begin{pmatrix} \int_0^t p d\widehat{\Lambda}_n \\ b \end{pmatrix} \right] \\
=\; & -\sqrt{n}\mathcal{P} \begin{pmatrix} l_{\Lambda\Lambda}(\Lambda_0, \beta_0) & l_{\Lambda\beta}(\Lambda_0, \beta_0) \\ l_{\beta\Lambda}(\Lambda_0, \beta_0) & l_{\beta\beta} \end{pmatrix} \left[\begin{pmatrix} \widehat{\Lambda}_n(t) - \Lambda_0(t) \\ \widehat{\beta}_n - \beta_0 \end{pmatrix}, \begin{pmatrix} \int_0^t p d\Lambda_0 \\ b \end{pmatrix} \right] + o_p(1) \\
=\; & \mathcal{G}_n \left\{ l_\Lambda(\Lambda_0, \beta_0)[\int_0^t p d\Lambda_0] + l_\beta^T b \right\} + o_p(1) \\
=\; & \mathcal{G}_n \left\{ l_\Lambda(\Lambda_0, \beta_0)[\int_0^t p d\widehat{\Lambda}_n] + l_\beta^T b \right\} + o_p(1). \qquad (A.4)
\end{aligned}
$$

In the above equations, $o_p(1)$ means convergence to zero in probability uniformly in p_{ij} and b.

Since $\mathcal{I}_n$ is invertible, for any bounded sequence $\{q_{ij}\}_{i=1,\ldots,n,j=1,\ldots,n_i}$ and $\widetilde{b}$, we can choose $\{p_{ij}\}_{i=1,\ldots,n,j=1,\ldots,n_i}$ and b such that $\mathcal{I}_n \begin{pmatrix} \vec{\Delta} \\ b \end{pmatrix} = \begin{pmatrix} \vec{q} \\ \widetilde{b} \end{pmatrix}$, where $\vec{\Delta} = \{p_{ij}\widehat{\Lambda}_n\{X_{ij}\}\}$ and $\vec{q}$ is the vector consisting of q_{ij}. With such choices, equation (A.4) yields

$$
\begin{aligned}
& \sum_{i=1}^n \sum_{j=1}^{n_i} \sqrt{n}(\widehat{\Lambda}_n\{X_{ij}\} - \widetilde{\Lambda}\{X_{ij}\})q_{ij} + \sqrt{n}(\widehat{\beta}_n - \beta_0)^T \widetilde{b} \\
=\; & \mathcal{G}_n \left\{ l_\Lambda(\Lambda_0, \beta_0)[\int_0^t p d\widehat{\Lambda}_n] + l_\beta^T b \right\} + o_p(1).
\end{aligned}
$$

The distribution of the right-hand side approximates a normal distribution with covariance matrix

$$
\begin{aligned}
& \mathcal{P}\left[\left\{ l_\Lambda(\Lambda_0, \beta_0)[\int_0^t p d\widehat{\Lambda}_n] + l_\beta(\Lambda_0, \beta_0)^T b \right\} \left\{ l_\Lambda(\Lambda_0, \beta_0)[\int_0^t p d\widehat{\Lambda}_n] + l_\beta(\Lambda_0, \beta_0)^T b \right\}^T \right] \\
& \quad = -\mathcal{P} \begin{pmatrix} l_{\Lambda\Lambda}(\Lambda_0, \beta_0) & l_{\Lambda\beta}(\Lambda_0, \beta_0) \\ l_{\beta\Lambda}(\Lambda_0, \beta_0) & l_{\beta\beta} \end{pmatrix} \left[\begin{pmatrix} \int_0^t p d\widehat{\Lambda}_n \\ b \end{pmatrix}, \begin{pmatrix} \int_0^t p d\widehat{\Lambda}_n \\ b \end{pmatrix} \right].
\end{aligned}
$$

This distribution can be approximated by

$$
-\mathcal{P}_n \begin{pmatrix} l_{\Lambda\Lambda}(\widehat{\Lambda}_n, \widehat{\beta}_n) & l_{\Lambda\beta}(\widehat{\Lambda}_n, \widehat{\beta}_n) \\ l_{\beta\Lambda}(\widehat{\Lambda}_n, \widehat{\beta}_n) & l_{\beta\beta} \end{pmatrix} \left[\begin{pmatrix} \int_0^t p d\widehat{\Lambda}_n \\ b \end{pmatrix}, \begin{pmatrix} \int_0^t p d\widehat{\Lambda}_n \\ b \end{pmatrix} \right],
$$

which is equal to $(\vec{\Delta}, b)\mathcal{I}_n \begin{pmatrix} \vec{\Delta} \\ b \end{pmatrix}$. Thus, the asymptotic variance for

$$
\sum_{i=1}^n \sum_{j=1}^{n_i} \sqrt{n}(\widehat{\Lambda}_n\{X_{ij}\} - \widetilde{\Lambda}\{X_{ij}\})q_{ij} + \sqrt{n}(\widehat{\beta}_n - \beta_0)^T \widetilde{b}
$$

can be approximated by

$$(\vec{\Delta}, \widetilde{b})\mathcal{I}_n \begin{pmatrix} \vec{\Delta} \\ \widetilde{b} \end{pmatrix} = (\vec{q}, \widetilde{b})\mathcal{I}_n^{-1} \begin{pmatrix} \vec{q} \\ \widetilde{b} \end{pmatrix}.$$

That is, for any vector $\widetilde{b}$ and any bounded function $q(t)$ such that $q(X_{ij}) = q_{ij}$, the asymptotic variance for $\sqrt{n}\int_0^\tau q(t)d(\widehat{\Lambda}_n - \Lambda_0) + \sqrt{n}(\widehat{\beta}_n - \beta_0)^T\widetilde{b}$ can be consistently estimated by $(\vec{q}, \widetilde{b})\mathcal{I}_n^{-1} \begin{pmatrix} \vec{q} \\ \widetilde{b} \end{pmatrix}$. This holds uniformly for any bounded function $q(t)$ and bounded vector $\widetilde{b}$.

Some other transformations. Condition 4 rules out such transformations as $G(x) = \log(1 + x)$. However, Condition 4 is only used in the first two steps of the consistency proof. Thus, if we can verify those two steps for the class of transformations $G(x) = \varrho\log(1 + rx)$, where ϱ and r are positive constants, then all the asymptotic results also hold for such transformations.

To prove Step 1, we rely on the explicit form of $G(x)$. It can be easily shown that the ith term of (4) is bounded from above. Condition 3 implies that, almost surely, there exist some subjects with $\overline{Y}_i^*(\tau) = 1, N_i^*(\tau) = 0$ and $C_i \geq \tau$. For such a subject, the corresponding term in (4) is equal to $-\varrho\log(1 + r\int_0^\tau e^{\beta^T Z_i(t)}d\Lambda)$, which is negative infinity if Λ has infinite jump sizes. Thus, Step 1 is proved.

To verify Step 2, it suffices to show $\widehat{\Lambda}_n(\tau) < \infty$. By equation (A.2) and the fact that $G'' < 0$,

$$\frac{1}{n\widehat{\Lambda}_n\{X_{ij}\}} \geq \frac{\varrho r}{n}\sum_{k=1}^n \int_0^\tau \frac{I(t \geq X_{ij})Y_k(X_{ij})e^{-M}}{1 + re^M\int_0^t Y_k(s)d\widehat{\Lambda}_n} dN_k(t).$$

Thus,

$$0 \leq \frac{1}{n}\left\{l_n(\widehat{\Lambda}_n, \widehat{\beta}_n) - l_n(\widehat{\Lambda}_n, \beta_0)\right\} \leq O(1)$$

$$-\frac{1}{n}\sum_{i=1}^n \int_0^\tau \log\left\{\frac{1}{n}\sum_{k=1}^n \int_0^\tau \frac{I(t \geq s)Y_k(s)e^{-M}}{1 + re^M\int_0^t Y_k(s)d\widehat{\Lambda}_n} dN_k(t)\right\} dN_i(s) - \frac{\varrho}{n}\sum_{i=1}^n \log(1 + re^{-M}\int_0^\tau Y_i(s)d\widehat{\Lambda}_n). \tag{A.5}$$

For simplicity, assume that $Y(\cdot)$ is non-increasing. We introduce a sequence $s_0 = \tau > s_1 > s_2 > \ldots, > s_Q = 0$. Then the right-hand side of the above inequality can be bounded from above by

$$O(1) - \frac{1}{n}\sum_{i=1}^n\sum_{q=1}^Q I(Y_i(s_{q-1}) = 0, Y_i(s_q) = 1)$$

$$\times \int_0^\tau \log\left\{\frac{1}{n}\sum_{k=1}^n \int_0^\tau \frac{I(t \geq s, t \in [s_q, s_{q-1}])}{1 + re^M\widehat{\Lambda}_n(s_{q-1})} dN_k(t)\right\} dN_i(s)$$

$$-\frac{\varrho}{n}\sum_{i=1}^n\sum_{q=1}^Q I(Y_i(s_{q-1}) = 0, Y_i(s_q) = 1)\log(1 + re^{-M}\widehat{\Lambda}_n(s_q)) - \frac{\varrho}{n}\sum_{i=1}^n I(Y_i(s_0) = 1)\log(1 + re^{-M}\widehat{\Lambda}_n(\tau)).$$

Rearranging this expression, we obtain that the right-hand side of (A.5) is bounded by

$$
\begin{aligned}
& O(1) - \frac{\varrho}{2n}\sum_{i=1}^{n} I(Y_i(s_0)=1)\log(1+re^{-M}\widehat{\Lambda}_n(\tau)) \\
&+ \left[\frac{1}{n}\sum_{i=1}^{n} I(Y_i(s_0)=0, Y_i(s_1)=1)N_i(\tau)\log(1+re^{M}\widehat{\Lambda}_n(\tau)) \right.\\
&\qquad \left. -\frac{\varrho}{2n}\sum_{i=1}^{n} I(Y_i(s_0=1))\log(1+re^{-M}\widehat{\Lambda}_n(\tau))\right] \\
&+ \sum_{q=1}^{Q-1}\left[\frac{1}{n}\sum_{i=1}^{n} I(Y_i(s_q)=0, Y_i(s_{q+1})=1)N_i(\tau)\log(1+re^{M}\widehat{\Lambda}_n(s_q)) \right.\\
&\qquad \left. -\frac{\varrho}{n}\sum_{i=1}^{n} I(Y_i(s_{q-1})=0, Y_i(s_q)=1)\log(1+re^{-M}\widehat{\Lambda}_n(s_q))\right]. \qquad (A.6)
\end{aligned}
$$

Therefore, if we can choose the sequence $s_0 = 0 > s_1 > \ldots, > s_Q = 0$ such that

$$\frac{1}{n}\sum_{i=1}^{n} I(Y_i(s_0)=0, Y_i(s_1)=1)N_i(\tau) < \frac{\varrho}{2n}\sum_{i=1}^{n} I(Y_i(s_0=1))$$

and

$$\frac{1}{n}\sum_{i=1}^{n} I(Y_i(s_q)=0, Y_i(s_{q+1})=1)N_i(\tau) < \frac{\varrho}{n}\sum_{i=1}^{n} I(Y_i(s_{q-1})=0, Y_i(s_q)=1),$$

then the first term in (A.6) diverges to negative infinity when $\widehat{\Lambda}_n(\tau) \to \infty$ but the second and third terms in (A.6) do not diverge. Thus, the right-hand side of (A.5) goes to negative infinity. This is a contradiction, so that Step 2 is verified.

The sequence $s_0 > s_1 > \ldots,$ can be chosen sequentially as follows: first, s_1 is defined as

$$s_1 = \inf_{0\le s<\tau}\left\{s : E[I(Y(s_0)=0, Y(s)=1)N(\tau)] < \epsilon_0 E[I(Y(s_0)=1)]\right\};$$

then given s_q, s_{q+1} is defined as

$$s_{q+1} = \inf_{0\le s<s_q}\left\{s : E[I(Y(s_q)=0, Y(s)=1)N(\tau)] < \epsilon_0 E[I(Y(s_{q-1})=0, Y(s_q)=1)]\right\},$$

where ϵ_0 is a constant less than $\varrho/2$ and is to be determined later. Clearly, such a sequence is well defined. We show that eventually $s_Q = 0$ for some finite Q. Otherwise, we obtain $s_0 > s_1 > \ldots, \to s^* \ge 0$. Since

$$E[I(Y(s_0)=0, Y(s_1)=1)N(\tau)] = \epsilon_0 E[I(Y(s_0)=1)],$$

$$E[I(Y(s_q)=0, Y(s_{q+1})=1)N(\tau)] = \epsilon_0 E[I(Y(s_{q-1})=0, Y(s_q=1))]$$

for $q \ge 1$, the summation of all these equalities yields

$$E[N(\tau)I(Y(s_0)=0, Y(s^*)=1)] = \epsilon_0 E[I(Y(s^*)=1)].$$

This cannot be true if we choose ϵ_0 small enough. Thus, s_Q must be zero for some finite Q.