The Annals of Statistics
2000, Vol. 28, No. 3, 779–814

TWO ESTIMATORS OF THE MEAN OF A COUNTING PROCESS WITH PANEL COUNT DATA

By Jon A. Wellner[1] and Ying Zhang[2]

University of Washington and University of Central Florida

We study two estimators of the mean function of a counting process based on "panel count data." The setting for "panel count data" is one in which n independent subjects, each with a counting process with common mean function, are observed at several possibly different times during a study. Following a model proposed by Schick and Yu, we allow the number of observation times, and the observation times themselves, to be random variables. Our goal is to estimate the mean function of the counting process.

We show that the estimator of the mean function proposed by Sun and Kalbfleisch can be viewed as a pseudo-maximum likelihood estimator when a non-homogeneous Poisson process model is assumed for the counting process. We establish consistency of both the nonparametric pseudo maximum likelihood estimator of Sun and Kalbfleisch and the full maximum likelihood estimator, even if the underlying counting process is not a Poisson process. We also derive the asymptotic distribution of both estimators at a fixed time t, and compare the resulting theoretical relative efficiency with finite sample relative efficiency by way of a limited Monte-Carlo study.

1. Introduction. Suppose that $N = \{N(t) : t \geq 0\}$ is a univariate counting process. In many applications, it is important to estimate the expected number of events which will occur by the time t, $\Lambda(t) = EN(t)$, the mean function of N.

In practice, a number of subjects are under study. Each subject is observed several times during the study. At the observation times, only the counts of events up to that time are observable; the exact times of events are unknown. The number of observation times and the observation times themselves are allowed to vary across the subjects. This kind of data often appear in demographic studies, system reliability, and clinical trials. Data of this type are sometimes referred to as *Panel Count Data*. Examples are given by Kalbfleisch and Lawless (1985), Gaver and O'Muircheataigh (1987), Thall and Lachin (1988), Thall (1988) and Sun and Kalbfleisch (1995).

In some situations, there is only one event recorded by the counting process for each subject; for example, death or onset of disease. Then panel count

Received February 1998; revised January 2000.

[1]Supported in part by NSF Grant DMS-95-32039 and NIAID Grant 2R01 AI291968-04.

[2]Supported in part by NSF Grant DMS-95-32039.

AMS 1991 *subject classifications*. Primary 62G05; secondary 62G20, 62N01.

Key words and phrases. Algorithm, asymptotic distributions, consistency, convex minorant, counting process, current status data, empirical processes, interval censoring, iterative, maximum likelihood, monte-carlo, pseudo likelihood, relative efficiency.

data are often referred to as *interval-censored data*. There is a large literature on statistical methods for interval-censored data and applications thereof; for examples of interval-censored data see, for example, Diamond and McDonald (1992), Diamond, McDonald and Shah (1986), Finkelstein (1986) and Finkelstein and Wolf (1985). Some theoretical results are also available. However, most of those results are limited to the special case in which each subject has the same number of observation times: see, for example, Groeneboom and Wellner (1992), Huang (1996) and Huang and Wellner (1995) for cases when each subject is only observed once or twice, while Wellner (1995) gives a preliminary study of the NPMLE for the case when each subject is observed k times. Schick and Yu (1999) show the consistency of the NPMLE for interval censoring when each subject is allowed to have different number of observation times and the counting process is a simple indicator (one jump) counting process. In this paper, we extend the formulation of the sample space given by Schick and Yu (1999) to study panel count data for general counting processes. We study both a pseudo-likelihood estimator $\widehat{\Lambda}_n^{ps}$ and the full maximum likelihood estimator $\widehat{\Lambda}_n$ of the mean function Λ_0 based on a non-homogeneous Poisson process model, that is,

$$P(N(t)=k)=\exp\{-\Lambda_0(t)\}\frac{\Lambda_0^k(t)}{k!},\qquad k=0,1,2,\dots, \tag{1.1}$$

where $\Lambda_0(t)=E(N(t))$, the mean function of the counting process N. We also establish asymptotic properties of the proposed estimators without assuming any model for the counting process N, thereby demonstrating an important robustness property of both estimators. Our proofs of the asymptotic results rely strongly on the generality in choice of sample space allowed by modern empirical process theory.

The pseudo-likelihood estimator we study was proposed by Sun and Kalbfleisch (1995), who constructed the estimator based on isotonic regression considerations. The pseudo-likelihood estimator ignores the dependence between counts in the counting process as successive observation times, treating these successive counts as if they were independent random variables to form a "pseudo likelihood." On the other hand, the full Non-Parametric Maximum Likelihood Estimator (NPMLE) of the mean function Λ_0 involves taking account of the dependence of the successive counts, and under the assumption that the counting process is a (nonhomogeneous) Poisson process, this is easily accomplished via independence of the increments of the process

$$\begin{aligned} P(N(t_1)=k_1,\dots,N(t_s)=k_s) = \prod_{j=1}^{s} & \frac{(\Lambda_0(t_j)-\Lambda_0(t_{j-1}))^{k_j-k_{j-1}}}{(k_j-k_{j-1})!} \\ & \times\exp[-(\Lambda_0(t_j)-\Lambda_0(t_{j-1}))] \end{aligned} \tag{1.2}$$

where $0<t_1<\cdots<t_s; 0\le k_1\le\cdots\le k_s$ and $t_0=k_0\equiv 0$ by convention. We show that the NPMLE $\widehat{\Lambda}_n$ of the mean function Λ_0 based on (1.2) is consistent

under mild restrictions even when the underlying true counting process is not a Poisson process. We also show that it is more efficient than the Sun-Kalbfleisch estimator, both when the nonhomogeneous Poisson assumption holds and when it fails.

When the counting process for each individual is observed at just one time (current status data), the estimator of Sun and Kalbfleisch (1995) is shown to be the NPMLE if the counting process has just one jump (a simple indicator counting process) or if it is a non-homogeneous Poisson process. Thus our results extend previous results for current status data (or interval censoring case 1) for interval censored data due to Groeneboom (1991) and Groeneboom and Wellner (1992). When the counting process for each individual is observed exactly twice, the estimator of Sun and Kalbfleisch (1995) differs from the NPMLE and provides an alternative to the NPMLE studied in the case of a simple counting process by Groeneboom (1991), Groeneboom and Wellner (1992), Wellner (1995) and Groeneboom (1996). When the counting process for each individual is observed a random number of times ["mixed case" interval censoring in the terminology of Schick and Yu (1999)], then our results generalize and extend those of Schick and Yu (1999) in several directions.

The outline of the rest of paper is as follows: In Section 2, we characterize the two estimators $\widehat{\Lambda}_n^{ps}$ and $\widehat{\Lambda}_n$. In Section 3, we provide algorithms for computation of the estimators: computation of the non-parametric maximum pseudo likelihood Estimator (NPMPLE) $\widehat{\Lambda}_n^{ps}$ can be accomplished in one step. On the other hand, computation of the NPMLE $\widehat{\Lambda}_n$ requires an iterative convex minorant algorithm (ICM). In Section 4, we state the main asymptotic results for both estimators: strong consistency and pointwise asymptotic distributions. These results are proved in Section 7 by use of tools from empirical process theory. In Section 5, we present an example involving data from a bladder tumor study in which we compute and compare both estimators. In Section 6, we present the results of simulation studies to compare the two estimators.

2. Characterization of two nonparametric estimators. Suppose that $N = \{N(t) : t \geq 0\}$ is a counting process with mean function $EN(t) = \Lambda_0(t)$, K is an integer-valued random variable and $\underline{T} = \{T_{k,j},\ j = 1, \dots, k,\ k = 1, 2, \dots\}$ is a triangular array of potential observation times. We assume throught that N and $(K, \underline{T})$ are independent, and $T_{k,j-1} \leq T_{k,j}$ for $j = 1, \dots, k$ and $k = 1, 2, \dots$. Let $X = (\mathbb{N}_K, \mathbb{T}_K, K)$, with a possible value $x = (\mathbf{n}_k, \mathbf{t}_k, k)$, where $\mathbb{N}_k = (N_{k,1}, \dots, N_{k,k})$ with $N_{k,j} = N(T_{k,j})$, $j = 1, 2, \dots, k$ and $\mathbb{T}_k$ is the kth row of the triangular array $\underline{T}$. Thus the sample space for one observation is $\mathscr{X} = \cup_{k=1}^{\infty} \{\mathbb{N}^k \times (0, \infty)^k \times \{k\}\}$. Suppose we observe n i.i.d. copies of X; $X_1, X_2, \dots, X_n$, where $X_i = (\mathbb{N}_{K_i}^{(i)}, \mathbb{T}_{K_i}^{(i)}, K_i)$, $i = 1, 2, \dots, n$. Here $(N^{(i)}, \underline{T}^{(i)}, K_i)$, $i = 1, 2, \dots$ are the underlying i.i.d. copies of $(N, \underline{T}, K)$.

Our goal is to construct nonparametric estimators of Λ_0 and study the properties of these estimators.

If we assume the non-homogeneous Poisson process model (1.1) and ignore the dependency of the events within a subject, we can form a "pseudo log-likelihood function" for Λ given by

$$l_n^{ps}(\Lambda|\underline{X}) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{N_{K_i,j}^{(i)}\log\Lambda(T_{K_i,j}^{(i)}) - \Lambda(T_{K_i,j}^{(i)})\right\}, \tag{2.1}$$

omitting the parts that are irrelevant in estimating Λ_0.

Let $s_1 < s_2 < \cdots < s_m$ denote the ordered distinct observation time points in the set of all observation time points $\{T_{K_i,j},\ j = 1, \ldots, K_i,\ i = 1, \ldots, n\}$. For $l \in \{1, \ldots, m\}$, we define

$$w_l = \sum_{i=1}^{n}\sum_{j=1}^{K_i} 1_{\{T_{K_i,j}^{(i)}=s_l\}}, \qquad \bar{N}_l = \frac{1}{w_l}\sum_{i=1}^{n}\sum_{j=1}^{K_i} N_{K_i,j}^{(i)} 1_{\{T_{K_i,j}^{(i)}=s_l\}},$$

and $\Lambda_l = \Lambda(s_l)$, and write $\underline{\Lambda} = (\Lambda_1, \ldots, \Lambda_m)$. Then (2.1) can be rewritten, with a slight abuse of notation, as

$$l_n^{ps}(\Lambda|\underline{X}) = l_n^{ps}(\underline{\Lambda}|\underline{X}) = \sum_{l=1}^{m} w_l\left[\bar{N}_l\log\Lambda_l - \Lambda_l\right], \tag{2.2}$$

and a nonparametric estimator $\widehat{\Lambda}_n^{ps}$ of Λ_0 can be defined to be a nondecreasing step function with possible jumps only occurring at s_i, $i = 1, \ldots, m$, that maximizes (2.2). (Of course only $\Lambda_1, \ldots, \Lambda_m$ are identifiable, and our choice of $\widehat{\Lambda}_n^{ps}$ as a step function with jumps at $s_1, \ldots, s_m$ is arbitrary; other conventions are also possible.) The following two lemmas characterize the nonparametric estimator $\widehat{\Lambda}_n^{ps}$ of Λ_0. The proofs rely on Theorem 2.1 below, directly follow the same lines as those of Propositions 1.1 and 1.2 of Groeneboom and Wellner (1992) and are therefore omitted. Let

$$\Omega^+ = \left\{\underline{y} \in R_+^m:\ 0 \le y_1 \le y_2 \le \cdots \le y_m\right\}. \tag{2.3}$$

LEMMA 2.1. *Suppose that* $\bar{N}_1 > 0$ *and* $\underline{\Lambda} = (\Lambda_1, \ldots, \Lambda_m) \in \Omega^+$. *Then* $\underline{\Lambda}$ *maximizes* (2.2) *over* Ω^+ *if and only if*

$$\sum_{j\ge l} w_j\left\{\frac{\bar{N}_j}{\Lambda_j} - 1\right\} \le 0 \qquad \text{for } l = 1, \ldots, m \tag{2.4}$$

and

$$\sum_{l=1}^{m} w_l\Lambda_l\left\{\frac{\bar{N}_l}{\Lambda_l} - 1\right\} = 0. \tag{2.5}$$

LEMMA 2.2. *Let* H^* *be the greatest convex minorant of the points*

$$\left(\sum_{j\le l} w_j, \sum_{j\le l} w_j\bar{N}_j\right)$$

on $[0, \sum_{l=1}^{m} w_l]$; *that is, for* $t \in [0, \sum_{l=1}^{m} w_l]$,

$$H^*(t) = \sup\left\{H(t):\ H\left(\sum_{j\le l} w_l\right) \le \sum_{j\le l} w_j \bar{N}_j \right.$$
$$\left. \textit{for each } l,\ 0 \le l \le m, H(0) = 0 \textit{ and } H \textit{ is convex}\right\}.$$

Moreover, let $\widehat{\Lambda}_i^{ps}$ *be the left derivative of* H^* *at* $\sum_{j\le i} w_j$. *Then* $\underline{\widehat{\Lambda}}^{ps} = (\widehat{\Lambda}_1^{ps}, \dots, \widehat{\Lambda}_m^{ps})$ *is the unique vector maximizing* (2.2) *over* Ω^+.

The left derivative of the greatest convex minorant in Lemma 2.2 can be explicitly solved by the "max-min" formula

$$\widehat{\Lambda}_l^{ps} = \widehat{\Lambda}_n^{ps}(s_l) = \max_{i\le l} \min_{j\ge l} \frac{\sum_{i\le p\le j} w_p \bar{N}_p}{\sum_{i\le p\le j} w_p}, \qquad l = 1, \dots, m. \tag{2.6}$$

The estimator $\widehat{\Lambda}_n^{ps}(s_l)$ given in (2.6) based on the "pseudo log-likelihood function" (2.2) is exactly that of Sun and Kalbfleisch (1995). They derived (2.6) based on isotonic regression considerations introduced in Barlow, Bartholomew, Bremner and Brunk (1972). The relationship between the NPMLE for case 1 interval-censored data and isotonic regression is also developed in Groeneboom and Wellner [(1992), page 43].

If we assume the nonhomogeneous Poisson process with joint distribution given by (1.2), it follows that the log likelihood function for Λ is given by

$$\begin{aligned} l_n(\Lambda|\underline{X}) = \sum_{i=1}^{n}\sum_{j=1}^{K_i} \left(N_{K_i,j}^{(i)} - N_{K_i,j-1}^{(i)}\right) \log\left[\Lambda(T_{K_i,j}^{(i)}) - \Lambda(T_{K_i,j-1}^{(i)})\right] \\ - \sum_{i=1}^{n} \Lambda(T_{K_i,K_i}^{(i)}), \end{aligned} \tag{2.7}$$

omitting the parts that do not depend on Λ.

Much as in the case of the pseudo-likelihood, for $1 \le l' < l \le m$ set

$$A_{l,l'} = \sum_{i=1}^{n}\sum_{j=1}^{K_i} \left(N_{K_i,j}^{(i)} - N_{K_i,j-1}^{(i)}\right) 1_{[T_{K_i,j}=s_l, T_{K_i,j-1}=s_{l'}]}$$

and, for $l \in \{1, \dots, m\}$,

$$B_l = \sum_{i=1}^{n} 1_{[T_{K_i,K_i}^{(i)}=s_l]}.$$

Then we can rewrite $l_n(\Lambda|\underline{X})$ in (2.7), with a slight abuse of notation, as

$$l_n(\Lambda|\underline{X}) = l_n(\underline{\Lambda}|\underline{X}) = \sum_{l'=1}^{m}\sum_{l=l'+1}^{m} A_{l,l'} \log(\Lambda_l - \Lambda_{l'}) - \sum_{l=1}^{m} B_l \Lambda_l \tag{2.8}$$

and the NPMLE of Λ_0 is defined to be the non-decreasing, non-negative step function with possible jumps only occurring at s_j, $j = 1, 2, \dots, m$, that maximizes (2.8). (Once again only $\Lambda_1, \dots, \Lambda_m$ are identifiable, and our choice of $\widehat{\Lambda}_n$ as a step function with jumps at $s_1, \dots, s_m$ is arbitrary; other conventions are also possible.)

Necessary and sufficient conditions characterizing the solution of the maximization problem $\widehat{\underline{\Lambda}} = \arg\max_{\underline{\Lambda}\in\Omega^+} l_n(\underline{\Lambda}|\underline{X})$ are easily formulated using the following version of the Fenchel duality theorem. The theorem below is stated in Groeneboom (1996); see also Lemma 3.1 of Wellner and Zhan (1997) and Jongbloed (1995, 1998).

THEOREM 2.1. *Let $\phi: \mathbb{R}^n \to \mathbb{R}\cup\{-\infty\}$ be a continuous concave function. Let $\mathscr{K} \subset R^n$ be a convex cone and let $\mathscr{K}_0 = \mathscr{K} \cap \phi^{-1}(\mathbb{R})$. Suppose $\mathscr{K}_0$ is nonempty and ϕ is differentiable on $\mathscr{K}_0$. Then $\hat{x} \in \mathscr{K}_0$ satisfies $\phi(\hat{x}) = \max_{x\in\mathscr{K}} \phi(x)$ if and only if*

$$\langle \hat{x}, \nabla\phi(\hat{x})\rangle = 0 \tag{2.9}$$

and

$$\langle x, \nabla\phi(\hat{x})\rangle \le 0, \tag{2.10}$$

for all $x \in \mathscr{K}$.

For notational ease and to match notation with the theorem, we define

$$\phi(\underline{u}) \equiv l_n(\underline{u}|\underline{X}), \qquad \underline{u} \in \Omega. \tag{2.11}$$

Since the convex cone Ω is generated by the finite subset $\{\mathbf{1}_j : \ j = 0, \dots m\}$, where $\mathbf{1}_j = (\underbrace{0\cdots 0}_{m-j}\underbrace{1\cdots 1}_{j})$, it follows that in our current problem (2.10) is equivalent to

$$\sum_{j=i}^{m} \frac{\partial\phi(\hat{x})}{\partial x_j} \le 0 \qquad \text{for } i = 1, 2, \dots, m. \tag{2.12}$$

The equivalence between (2.10) and (2.12) can be proved by following the same argument as given on pages 39–40 of Groeneboom and Wellner (1992).

Denote

$$\phi_l(\underline{u}) = \frac{\partial\phi(\underline{u}|\underline{X})}{\partial u_l} \qquad \text{for } l = 1, 2, \dots, m\,;$$

$\phi_l(\underline{u})$ can be computed explicitly from (2.11) and (2.8). Then the NPMLE of the mean function, $\widehat{\Lambda}_n$ can be characterized by the following:

$$\sum_{l=1}^{m} \phi_l(\widehat{\underline{\Lambda}})\widehat{\Lambda}_l = 0 \tag{2.13}$$

and

$$\sum_{l=p}^{m} \phi_l(\underline{\widehat{\Lambda}}) \leq 0 \qquad \text{for all } p = 1, 2, \dots, m. \tag{2.14}$$

3. Computation of the estimators. As noted in Section 2, the NPMPLE $\widehat{\Lambda}_n^{ps}$ can be computed in one step via the max-min formula (2.6). Our proposed method for computation of the NPMLE $\widehat{\Lambda}_n$ characterized by (2.13) – (2.14) involves the iterative convex minorant algorithm (ICM). This algorithm has proved, by experience, to be numerically stable and rapidly convergent in many similar problems; see, for example, Groeneboom and Wellner (1992), Jongbloed (1998) and Wellner and Zhan (1997).

For any $\underline{u} = (u_1, u_2, \dots, u_m) \in \Omega$, define

$$\phi_{ll}(\underline{u}) = \frac{\partial^2 \phi(\underline{u}|\underline{X})}{\partial u_l^2} \qquad \text{for } l = 1, 2, \dots, m\,;$$

$\phi_{ll}(\underline{u})$ can be computed explicitly from (2.11) and (2.8). We define two processes $G(\underline{u}, \cdot)$ and $V(\underline{u}, \cdot)$ by $G(\underline{u}, 0) = 0$,

$$G(\underline{u}, p) = \sum_{l=1}^{p} (-\phi_{ll}(\underline{u})) \qquad \text{for } p = 1, 2, \dots, m, \tag{3.1}$$

$V(\underline{u}, 0) = 0$ and

$$V(\underline{u}, p) = \sum_{l=1}^{p} [\phi_l(\underline{u}) + u_l(-\phi_{ll}(\underline{u}))] \qquad \text{for } p = 1, 2, \dots, m. \tag{3.2}$$

THEOREM 3.1. *A vector* $\underline{\widehat{\Lambda}}$ *satisfies* (2.13) *and* (2.14) *if and only if* $\underline{\widehat{\Lambda}}$ *is the left derivative of the convex minorant of the cumulative sum-diagram consisting of the following points*:

$$P_0 = (0, 0),$$
$$P_l = (G(\underline{\widehat{\Lambda}}, l), V(\underline{\widehat{\Lambda}}, l)), \qquad l = 1, 2, \dots, m.$$

This theorem is actually Theorem 4.3 of Wellner and Zhan [(1997), page 952]. In general, Theorem 3.1 does not give an explicit solution $\underline{\widehat{\Lambda}}$ as simple as that we have seen in Lemma 2.2 for the maximum pseudo likelihood estimator. Solving for $\underline{\hat{\Lambda}}$ typically requires an iterative procedure as follows.

ITERATIVE CONVEX MINORANT ALGORITHM. Let

$$\Delta_{i,j}(\underline{u}) = \frac{V(\underline{u}, i) - V(\underline{u}, j)}{G(\underline{u}, i) - G(\underline{u}, j)},$$

and η be the accuracy parameter.

Step 1. Select an initial guess $\underline{u}^0 \in \Omega$.

Step 2. Update the solution by

$$u_l^{k+1} = \max_{j \le l} \min_{i \ge l} \Delta_{i,j}(\underline{u}^k), \quad l = 1, 2 \dots, m; \; k = 0, 1, 2, \dots.$$

Step 3. If

$$\left| \sum_{l=1}^{m} \left[\sum_{i=1}^{n} \phi_{i,l}(\underline{u}^{k+1}) \right] u_l^{k+1} \right| > \eta$$

or

$$\max_{1 \le p \le m} \sum_{l=p}^{m} \sum_{i=1}^{n} \phi_{i,l} \left(\underline{u}^{k+1} \right) > \eta,$$

go back to Step 2, otherwise stop the iteration here.

Although the ICM works well in many applications, it is not guaranteed to be globally convergent. To avoid this potential problem, Jongbloed (1998) devised the "modified iterative convex minorant algorithm (MICM)" by inserting a binary line search procedure into the ICM. He proved that the MICM algorithm converges globally. The essence of the line search procedure is to keep the iterations in the feasible region. Our experience with the MICM algorithm is that inclusion of the line search (Armijo's rule) is very important. See Jongbloed (1998) for a proof of global convergence of the modified ICM algorithm.

4. Asymptotic theory: Results. Although the estimator $\widehat{\Lambda}_n^{ps}$ described by (2.6) is also given in Sun and Kalbfleisch (1995), the properties and behavior of this estimator are still unknown. For the consistency of the estimator $\widehat{\Lambda}_n^{ps}$, Sun and Kalbfleisch (1995) referred to some results for isotonic regression given by Brunk (1970). In the present problem, however, the average counts $\bar{N}_l, l = 1, \dots, m$ are not independent, and hence do not satisfy the conditions of Brunk's theorems. The estimator $\widehat{\Lambda}_n$ is described here for the first time.

Here we will use empirical process theory to study the properties of the estimators $\widehat{\Lambda}_n^{ps}$ and $\widehat{\Lambda}_n$. We establish consistency of both estimators in L_2-metrics related to the observation scheme, and establish the asymptotic distribution of the estimators at fixed time points under some mild conditions.

First some notation. Let $\mathscr{B}$ denote the collection of Borel sets in $\mathbb{R}$ and let $\mathscr{B}_{[0,\tau]} = \{B \cap [0, \tau] : \; B \in \mathscr{B}\}$. On $([0, \tau], \mathscr{B}_{[0,\tau]})$ we define measures μ, μ_2 and ν, as follows: for $B, B_1, B_2 \in \mathscr{B}_{[0,\tau]}$, set

$$\begin{aligned} \mu(B) &= \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(T_{k,j} \in B | K = k) \\ &= E\left\{ \sum_{j=1}^{K} 1_B(T_{K,j}) \right\}, \end{aligned} \tag{4.1}$$

$$\mu_2(B_1 \times B_2) = \sum_{k=1}^{\infty} P(K=k) \sum_{j=1}^{k} P(T_{k,j-1} \in B_1, T_{k,j} \in B_2 | K=k) \tag{4.2}$$
$$= E\left\{ \sum_{j=1}^{K} 1_{B_1}(T_{K,j-1}) 1_{B_2}(T_{K,j}) \right\}$$

and

$$\nu(B) = \sum_{k=1}^{\infty} P(K=k) P(T_{k,k} \in B | K=k) = E\left\{1_B(T_{K,K})\right\}. \tag{4.3}$$

Let $\mathscr{F}$ be the class of functions

$$\mathscr{F} \equiv \{\Lambda : [0,\infty) \to [0,\infty) \big| \Lambda \text{ is nondecreasing}, \Lambda(0) = 0\}, \tag{4.4}$$

and let d be the $L_2(\mu)$ metric on $\mathscr{F}$; thus for $\Lambda_1, \Lambda_2 \in \mathscr{F}$,

$$d(\Lambda_1, \Lambda_2) = \left[\int |\Lambda_1(t) - \Lambda_2(t)|^2 d\mu(t)\right]^{1/2}.$$

We also define a metric d_2 on $\mathscr{F}$ in terms of μ_2:

$$d_2(\Lambda_1, \Lambda_2) = \left[\int\int |\Lambda_1(v) - \Lambda_1(u) - (\Lambda_2(v) - \Lambda_2(u))|^2 d\mu_2(u,v)\right]^{1/2}.$$

The metrics d and d_2 are closely related and if $P(K \le k_0) = 1$ for some $k_0 < \infty$, then they are equivalent:

$$\tfrac{1}{2} d_2(\Lambda_1, \Lambda_2) \le d(\Lambda_1, \Lambda_2) \le k_0 d_2(\Lambda_1, \Lambda_2). \tag{4.5}$$

Moreover, with $\tilde{\mu}$ defined by

$$\tilde{\mu}(B) = \sum_{k=1}^{\infty} P(K=k) \frac{1}{k^2} \sum_{j=1}^{k} P(T_{k,j} \in B | K=k), \tag{4.6}$$

the corresponding metric $\tilde{d}$ defined by $\tilde{d}^2(\Lambda_1, \Lambda_2) = \int [\Lambda_1(t) - \Lambda_2(t)]^2 d\tilde{\mu}(t)$ satisfies

$$\tilde{d}(\Lambda_1, \Lambda_2) \le d_2(\Lambda_1, \Lambda_2). \tag{4.7}$$

[Proofs of (4.5) and (4.7) will be given along with the proof of Theorem 4.2 in Section 7.] The measure μ was introduced by Schick and Yu (1999); note that μ and μ_2 are finite measures if $E(K) < \infty$. Also note that $d^2(\Lambda_1, \Lambda_2)$ can also be written in terms of an expectation as

$$d^2(\Lambda_1, \Lambda_2) = E\left[\sum_{j=1}^{K} \left(\Lambda_1(T_{K,j}) - \Lambda_2(T_{K,j})\right)^2\right]. \tag{4.8}$$

To prove consistency of the maximum pseudo likelikehood estimator $\widehat{\Lambda}_n^{ps}$ and the maximum likelihood estimator $\widehat{\Lambda}_n$ we will use the following regularity conditions on the true mean function Λ_0 and the underlying distribution of observation times.

CONDITIONS.

A. The observation times $T_{k,j}$, $j = 1, \dots, k$, $k = 1, 2, \dots$ are random variables taking values in the bounded set $[0, \tau]$ where $\tau \in (0, \infty)$ and $E(K) < \infty$.

B. The true mean function Λ_0 satisfies $\Lambda_0(\tau) \le M$ for some $M \in (0, \infty)$ and $\Lambda_0(t) = EN(t)$, $0 \le t < \infty$.

C. The function M_0^{ps} defined by $M_0^{ps}(X) \equiv \sum_{j=1}^{K} N_{K,j} \log(N_{K,j})$ satisfies $PM_0^{ps}(X) < \infty$.

D. The function M_0 defined by $M_0(X) \equiv \sum_{j=1}^{K} \Delta N_{K,j} \log(\Delta N_{K,j})$ satisfies $PM_0(X) < \infty$.

Note that Condition C holds if $P(\sum_{j=1}^{K}(N_{K,j})^{1+\delta}) < \infty$ for some $\delta > 0$. Both Condition C and the sufficient condition can be expressed in terms of the measure μ since $PM_0^{ps} = \int E[N(t)) \log(N(t))] d\mu(t)$ and $P(\sum_{j=1}^{K}(N_{K,j})^{1+\delta}) = \int E[(N(t))^{1+\delta}] d\mu(t)$. Similarly, Condition D holds if $P(\sum_{j=1}^{K}(\Delta N_{K,j})^{1+\delta}) < \infty$ for some $\delta > 0$ and both Condition D and the sufficient condition can be expressed in terms of the measure μ_2 since we have $PM_0 = \int\int E[(N(v) - N(u)) \log(N(v) - N(u))] d\mu_2(u, v)$ and $P(\sum_{j=1}^{K}(\Delta N_{K,j})^{1+\delta}) = \int\int E[(N(v) - N(u))^{1+\delta}] d\mu_2(u, v)$.

In the current setting, our two consistency theorems for the pseudo likelihood estimator $\widehat{\Lambda}_n^{ps}$ and the full maximum likelihood estimator $\widehat{\Lambda}_n$ are as follows.

THEOREM 4.1 (Consistency of the NPMPLE). *Suppose that A, B and C hold. Then, for every $b < \tau$ for which $\mu([b, \tau]) > 0$,*

$$d(\widehat{\Lambda}_n^{ps} 1_{[0,b]}, \Lambda_0 1_{[0,b]}) \longrightarrow 0 \qquad a.s.\ as\ n \to \infty.$$

In particular, if $\mu(\{\tau\}) > 0$, then

$$d(\widehat{\Lambda}_n^{ps}, \Lambda_0) \longrightarrow 0 \qquad a.s.\ as\ n \to \infty.$$

This conclusion also holds if

$$\limsup_{n\to\infty} \widehat{\Lambda}_n^{ps}(\tau) < \infty \qquad a.s. \tag{4.9}$$

Note that by the max-min definition of $\widehat{\Lambda}_n^{ps}$, (4.9) holds if

$$\limsup_{n\to\infty} \max_{1\le i\le n} N^{(i)}(\tau) < \infty \qquad \text{a.s.},$$

and this is always true if the counting processes $N^{(i)}$ have at most finitely many jumps.

THEOREM 4.2 (Consistency of the NPMLE). *Suppose that* A, B *and* D *hold. Then, for every* $b < \tau$ *for which* $\nu([b, \tau]) > 0$,

$$d(\widehat{\Lambda}_n 1_{[0,b]}, \Lambda_0 1_{[0,b]}) \longrightarrow 0 \qquad a.s. \text{ as } n \to \infty.$$

In particular, if $\nu(\{\tau\}) > 0$, *then*

$$d(\widehat{\Lambda}_n, \Lambda_0) \longrightarrow_p 0 \qquad \text{as } n \to \infty.$$

These consistency theorems have a number of important corollaries concerning consistency of the estimators in stronger metrics under additional hypotheses on the mean function Λ_0 and the observation scheme just as in the treatment by Schick and Yu (1999) of "mixed case interval censoring," but we will forego those corollaries here.

We now turn to asymptotic distribution theory at a fixed point t_0.

FURTHER CONDITIONS E.

E1. There is an $\alpha > 0$ and $M_1 > 0$, such that $E(K^{2+\alpha}) < \infty$ and

$$E(N^{2+\alpha}(t)) \le M_1 \quad \text{for all } t \in S_\mu.$$

E2. For a fixed $t_0 \in S_\mu \equiv \text{supp}(\mu)$, there is a neighborhood of t_0, such that $G_{k,j}$ is differentiable, and $G'_{k,j}(s)$ is continuous in this neighborhood. Moreover, $G'_{k,j}(t)$ is positive and uniformly bounded for all $j = 1, 2, \dots, k$, $k = 1, 2, \dots$.

E3. In the neighborhood of t_0 described in E2, the true mean function Λ_0 is differentiable, and its derivative Λ'_0 is continuous and strictly positive.

E4. Suppose that $G_{k,i,j}(s, t) = P(T_{k,i} \le s, T_{k,j} \le t)$ is differentiable with respect to s and t, and in a neighborhood of (t_0, t_0)

$$g_{k,i,j}(s, t) = \frac{\partial^2}{\partial s \partial t} G_{k,i,j}(s, t)$$

exists and is uniformly bounded for all i, $j = 1, 2, \dots, k$, $k = 1, 2, \dots$.

E5. $\sigma^2(t) \equiv Var(N(t))$ is continuous in a neighborhood of t_0.

THEOREM 4.3 (Asymptotic distribution, NPMPLE). *Suppose that Conditions* A–C *and* E1–E5 *hold. Then*

$$n^{1/3}\left(\widehat{\Lambda}_n^{ps}(t_0) - \Lambda_0(t_0)\right) \longrightarrow_d \left[\frac{\sigma^2(t_0)\Lambda'_0(t_0)}{2G'(t_0)}\right]^{1/3} 2 \arg\max_h \{\mathbb{Z}(h) - h^2\}, \tag{4.10}$$

where $G'(t) = \sum_{k=1}^\infty P(K = k) \sum_{j=1}^k G'_{k,j}(t)$, *and* $\mathbb{Z}$ *is a two-sided Brownian motion process, starting from zero.*

Now we will study the "toy estimator" $\widehat{\Lambda}_n^{(0)}$ corresponding to $\widehat{\Lambda}_n$. As in Groeneboom and Wellner (1992), this is defined to be the result of carrying out one step of the iterative convex minorant algorithm starting with the truth, namely Λ_0. For progress in the study of the full NPMLE of the distribution function F in the case of interval censored data, see Groeneboom [(1996), Section 4.2]. While it seems clear that the toy version of the NPMLE and the full NPMLE will be asymptotically equivalent, at least under "strict separation hypotheses" on the observation time distributions, the complete proof of this will require more effort. We postpone this to future research. Moreover, for the moment we will establish the pointwise asymptotic distribution of $\widehat{\Lambda}_n^{(0)}$ only when the hypothesized non-homogeneous Poisson process model used to derive the estimator is valid.

Suppose that the distributions $G_{k,j,j'}$ have densities $g_{k,j,j'}$ with respect to Lebesgue measure on $\mathbb{R}^2$, and define functions $H_{i,k,j}$, $i = 1, 2$, $j = 1, \ldots, k$, $k = 1, 2, \ldots$ and $\overline{H}$ as follows:

$$H_{1,k,j}(t_{k,j}) = \int_0^{t_{k,j}} \frac{1}{\Lambda_0(t_{k,j}) - \Lambda_0(t_{k,j-1})} g_{k,j-1,j}(t_{k,j-1}, t_{k,j}) dt_{k,j-1},$$

$$H_{2,k,j}(t_{k,j}) = \int_{t_{k,j}}^{\infty} \frac{1}{\Lambda_0(t_{k,j+1}) - \Lambda_0(t_{k,j})} g_{k,j,j+1}(t_{k,j}, t_{k,j+1}) dt_{k,j+1}$$

and

$$\overline{H}(t) = \sum_{k=1}^{\infty} P(K = k) \left\{ \sum_{j=1}^{k-1} \left[H_{1,k,j}(t) + H_{2,k,j}(t) \right] + H_{1,k,k}(t) \right\}. \tag{4.11}$$

We will assume that all the $H_{i,k,j}$'s, and hence also $\overline{H}$, are finite. We also define

$$\begin{aligned} H_{1,k,j}(t_{k,j}, \varepsilon) \equiv \int_0^{t_{k,j}} & \frac{1}{\Lambda_0(t_{k,j}) - \Lambda_0(t_{k,j-1})} 1_{[1/(\Lambda_0(t_{k,j}) - \Lambda_0(t_{k,j-1})) > \varepsilon]} \\ & \times g_{k,j-1,j}(t_{k,j-1}, t_{k,j}) dt_{k,j-1}, \end{aligned}$$

$$\begin{aligned} H_{2,k,j}(t_{k,j}, \varepsilon) \equiv \int_{t_{k,j}}^{\infty} & \frac{1}{\Lambda_0(t_{k,j+1}) - \Lambda_0(t_{k,j})} 1_{[1/(\Lambda_0(t_{k,j+1}) - \Lambda_0(t_{k,j})) > \varepsilon]} \\ & \times g_{k,j,j+1}(t_{k,j}, t_{k,j+1}) dt_{k,j+1} \end{aligned}$$

and

$$\overline{H}(t, \varepsilon) \equiv \sum_{k=1}^{\infty} P(K = k) \left[\sum_{j=1}^{k-1} \left\{ H_{1,k,j}(t; \varepsilon) + H_{2,k,j}(t; \varepsilon) \right\} + H_{1,k,k}(t, \varepsilon) \right].$$

Then we will assume the following asymptotic negligibility condition:

$$\alpha \int_{(t_0, t_0 + t/\alpha]} \overline{H}(u, \varepsilon\alpha) du \to 0 \qquad \text{as } \alpha \to \infty \tag{4.12}$$

for each $\varepsilon > 0$ and $t > 0$.

FURTHER CONDITIONS F.

F1. $E(K^{2+\alpha}) < \infty$ for some $\alpha > 0$ and the asymptotic negligibility condition (4.12) holds.

F2. The counting process N is a nonhomogeneous Poisson process with mean function Λ_0.

F3. Suppose that the function $\overline{H}$ defined in (4.11) is continuous and strictly positive in a neighborhood of t_0.

THEOREM 4.4 (Asymptotic distribution, NPMLE toy version). *Suppose that* A, B, D, E2, E3 *and* F1–F3 *hold. Then*

$$n^{1/3}\left(\widehat{\Lambda}_n^{(0)}(t_0) - \Lambda_0(t_0)\right) \longrightarrow_d \left[\frac{\Lambda_0'(t_0)}{2\overline{H}(t_0)}\right]^{1/3} 2 \arg\max_h \left\{\mathbb{Z}(h) - h^2\right\}, \tag{4.13}$$

where $\mathbb{Z}$ *is a two-sided Brownian motion process, starting from zero.*

REMARK 1. It should be emphasized that the hypotheses of Theorems 4.1–4.3 *do not* require that the true counting process N be a (non-homogeneous) Poisson process. Thus these results yield a very strong robustness property of both the pseudo-likelihood estimator $\widehat{\Lambda}_n^{ps}$ and the NPMLE $\widehat{\Lambda}_n$.

REMARK 2. We call the distribution of $\operatorname{argmax}_h \left\{\mathbb{Z}(h) - h^2\right\}$, which appears in Theorems 4.3 and 4.4, *Chernoff's distribution*. It has recently been computed and tabled by Groeneboom and Wellner (1999). This enables construction of pointwise confidence sets for $\Lambda(t_0)$ given consistent estimators of $\sigma^2(t_0)$, $\Lambda'(t_0)$ and $G'(t_0)$.

REMARK 3. Suppose that for each $i = 1, 2, \ldots$, $G'_{k,i}(t)$ are all equal to G'_i for $k = i, i+1, \ldots$. Then $G'(t)$ in Theorem 4.3 can be rewritten as

$$G'(t) = \sum_{i=1}^{\infty} G'_i(t) \sum_{k=i}^{\infty} P(K = k) = \sum_{i=1}^{\infty} G'_i(t) P(K \geq i) \geq G'_1(t).$$

Hence $G'(t)$ is bigger than G'_1, the density function of observation time in the case of current status data, the case $K = 1$ with probability 1. Hence under the above condition the estimator $\widehat{\Lambda}_n^{ps}$ of the mean function Λ_0 based on panel count data has smaller variance than the estimator based only on current status data; that is, multiple observation times help in estimating Λ_0 under the above condition. Since the numerator of the constant appearing in (4.10) depends only on the the counting process N, whenever the denominator $G'(t_0)$ is bigger than the corresponding term in the denominator $g(t_0)$ in the case of current status data, the asymptotic variance is smaller.

The following examples illustrate the generality of Theorems 4.1–4.4 and relate them to previous results in several important special cases.

EXAMPLE 4.1 [Current status data, simple (one event) counting process]. Suppose that the counting process N is the simple counting process $N(t) = 1_{[Y\le t]}$ where Y is a non-negative random variable with distribution function F. Then $\Lambda_0(t) = F(t)$, and $\sigma^2(t) = F(t)(1-F(t))$. If $K = 1$ with probability one, then our model reduces to "current-status data" for Y [or "interval censoring case 1" in the terminology of Groeneboom and Wellner (1992)]. Moreover it is easily seen that the estimator $\widehat{\Lambda}_n^{ps}(t)$ of $\Lambda_0(t) = F(t)$ is exactly the NPMLE of F studied in Groeneboom and Wellner (1992), and the convergence in distribution in Theorem 4.3 agrees exactly with Theorem 5.1 of Groeneboom and Wellner [(1991), page 89] upon noting that

$$\frac{\sigma^2(t_0)\Lambda_0'(t_0)}{2G'(t_0)} = \frac{F(t_0)(1-F(t_0))f(t_0)}{2g(t_0)}$$

in this case (with $G' \equiv g$).

EXAMPLE 4.2 (Interval censoring case 2; simple counting process). Suppose that N is a simple indicator counting process (so that $\Lambda_0(t) = F(t)$ and $\sigma^2(t) = F(t)(1-F(t))$) as in Example 1, but now suppose that $K = 2$ with probability one so that $G(t) = G_{2,1}(t) + G_{2,2}(t)$, the sum of the two marginal distributions of $\underline{T}_2 = (T_{2,1}, T_{2,2})$. In this case the estimator $\widehat{\Lambda}_n^{ps}(t)$ of $\Lambda_0(t) = F(t)$ differs from the NPMLE of F studied by Groeneboom (1991), Groeneboom and Wellner (1992), Wellner (1995) and Groeneboom (1996), and is somewhat inefficient, even under the "separation type" hypotheses under which the NPMLE $\widehat{\mathbb{F}}_n(t)$ converges to $F(t)$ at rate $n^{1/3}$; cf. Wellner (1995) and Groeneboom (1996). [Under non-separation, 'positive density along the diagonal'-type hypotheses on the distribution G_2 of $\underline{T}_2$, the arguments of Groeneboom (1991) and Groeneboom and Wellner (1992) suggest that $\widehat{\mathbb{F}}_n(t)$ converges to $F(t)$ at rate $(n\log n)^{1/3}$.] For example, if

$$g_2(t_1, t_2) \equiv \frac{\partial^2}{\partial t_1 \partial t_2} G_2(t_1, t_2) = (\alpha+2)(\alpha+1)(t_2 - t_1)^{\alpha} 1_{[0\le t_1\le t_2\le 1]}$$

as in Example 2 of Wellner [(1995), page 276] and if $F(x) = x$, $0 \le x \le 1$, then the asymptotic relative efficiency of the estimator $\widehat{\Lambda}_n^{ps}(t)$ relative to the NPMLE based on the conjectured asymptotic distribution given in Wellner (1995) is, at a point $t_0 \in (0,1)$,

$$\begin{aligned}
&\mathrm{ARE}_{\widehat{\Lambda}^{ps},\widehat{\Lambda}}(t_0)\\
&\quad = \frac{t_0^{1+\alpha} + (1-t_0)^{1+\alpha}}{t_0^{2+\alpha} + (1-t_0)^{2+\alpha} + \alpha^{-1}(1+\alpha)\left\{t_0(1-t_0)^{1+\alpha} + t_0^{1+\alpha}(1-t_0)\right\}}\\
&\quad = \frac{2\alpha}{2\alpha+1} \qquad \text{when } t_0 = \frac{1}{2}
\end{aligned}$$

and the latter ranges from 0 to 1 as α goes from 0 to ∞. When $\alpha = 0$, the current density g_2 is uniform on the triangle $0 \le t_1 \le t_2 \le 1$, and this is exactly the "positive density along the diagonal" case in which Groeneboom and Wellner (1992) showed that the "toy" version of the NPMLE in interval censoring case 2 converges at the faster rate $(n \log n)^{1/3}$. See Wellner (1995) for further discussion. While this conjecture has not yet been proved for the NPMLE even for interval censoring case 2, we conjecture that the faster rate holds for the NPMLE in our current model under such positive density along the diagonal type hypotheses, while the NPMPLE will continue to converge at rate $n^{1/3}$. These conjectures are completely consistent with the $ARE = 0$ when $\alpha = 0$.

EXAMPLE 4.3 (Current status data, nonhomogeneous Poisson process). Suppose that the counting process N is a non-homogeneous Poisson process with mean function Λ_0. Furthermore, suppose that $K = 1$ with probability one. Then there is just one term in the inner sum in (2.1), and hence the pseudo log-likelihood $l_n(\Lambda|\underline{X})$ is actually the true log-likelihood. Thus the development in Section 2 shows that $\widehat{\Lambda}_n^{ps}$ is the NPMLE of $\Lambda_0(t)$ in this case. Our Theorems 4.1 and 4.2 show that $\widehat{\Lambda}_n^{ps}$ and $\widehat{\Lambda}_n$ are consistent estimators of Λ_0, even if the counting process N is *not* a Poisson process.

EXAMPLE 4.4 ("Mixed case" interval censored data, simple counting process). Suppose that N is a simple indicator (one event) counting process, $N(t) = 1_{[Y \le t]}$, as in examples 1 and 2, but now suppose that K is random so that we have a *variable number* of observation times for each individual. This is what Schick and Yu (1999) refer to as "mixed case" interval censored data. The estimators $\widehat{\Lambda}_n^{ps}$ and $\widehat{\Lambda}_n$ of $\Lambda_0 = F$ both differ from the NPMLE of F in this case (at least when $K_i \ge 2$ for some i), but Theorems 4.1 and 4.2 guarantee that they are at least consistent, and Theorems 4.3 and 4.4 give conditions under which their (pointwise) convergence rates are $n^{1/3}$.

5. An example. The data, extracted from Andrews and Herzberg (1985), are from a bladder tumor study, conducted by the Veterans Administration Cooperative Urological Research Group (VACURG). Previous studies of this data set can be found in Byar, Blackard and the VACURG (1977), Byar (1980) and Wei, Lin and Weissfeld (1989).

This was a randomized clinical trial. In the trial, all patients experienced superficial bladder tumors when they entered the trial. They were randomly divided into three groups: one placebo group and two treatment groups (one group was assigned pyridoxine pills; a second group received periodic instillation of a chemotherapeutic agent, thiotepa, into the bladder). The follow-up time periods varied among patients. At each follow-up visit, any tumors noticed were counted, measured and then removed transurethrally, and the treatment was continued. The purpose of the study was to determine the effect of treatment on the frequency of tumor recurrence.

We estimated the mean functions of tumor counts for the three groups by both the pseudo maximum likelihood method (i.e., the Sun-Kalbfleisch estimator) and the full maximum likelihood method as defined in Section 2. The three estimated mean functions using the pseudo maximum likelihood method are shown in Figure 1. We can see that the treatment using thiotepa seems to reduce the tumor recurrence, assuming that randomization has already adjusted the effects of initial tumor count and size.

The three estimated mean functions using the full maximum likelihood method are also shown in Figure 1. The estimated mean functions agree qualitatively with the estimated mean functions using maximum pseudo likelihood, but there are also substantial differences. As can be seen from the plot, the NPMLE $\widehat{\Lambda}_n$ tends to be smaller than the maximum pseudo likelihood estimator $\widehat{\Lambda}_n^{ps}$, though the shapes of the curves are basically the same. This makes sense because the maximum pseudo likelihood estimator $\widehat{\Lambda}_n^{ps}$ is essentially a "mean-type" estimator (it can be viewed as the "pool-adjacent-violators" estimator). The shapes of the mean functions based on the NPMLEs imply that treatment using thiotepa seems to reduce the tumor recurrence, assuming that randomization has adjusted the effects from initial tumor count and size.

This data was also considered by Wei, Lin and Weissfeld (1989). They treated the clinical visit times as the time of tumor recurrence and applied marginal Cox regression models adapted to the multivariate data to estimate the treatment effect on the marginal distribution of tumor recurrence times. Here we treat the clinical visit times as the observation times, and take the tumor count as the main response variable. Thus the data can be treated as

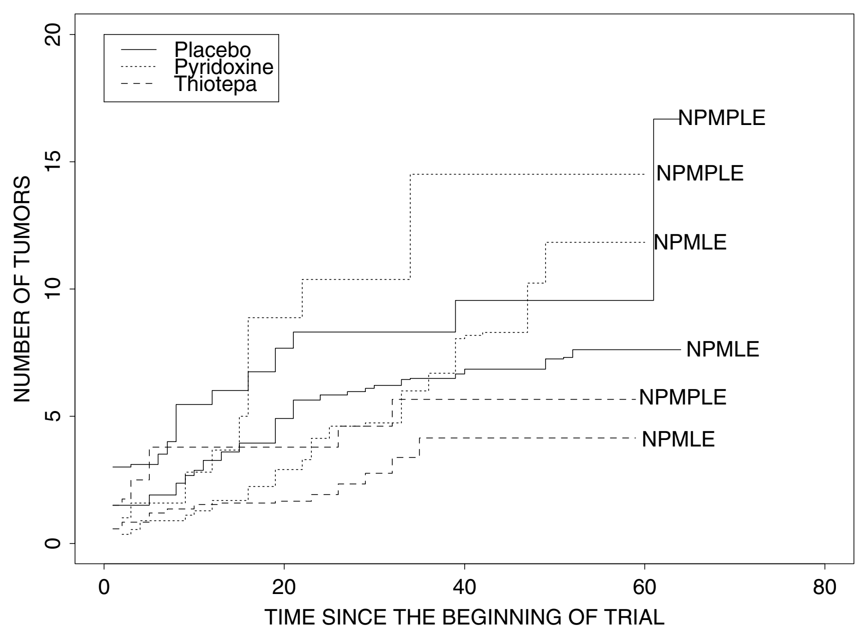

FIG. 1. *Estimators of the mean functions, bladder tumors data.*

panel count data, and fits quite naturally in the setting for panel count data introduced in Sections 1–3. Then the effect of treatment can be seen through the comparison of the mean functions of tumor counts across the treatment groups.

Another approach would involve development of a regression model for tumor counts, and could include the treatment, initial tumor count and initial tumor size as covariates. This would allow for semiparametric estimation of treatment effects on tumor counts with adjustments for the initial tumor count and size. Some work on models of this type has been carried out in Zhang (1998, 1999).

6. Simulations. To compare the properties of the NPMPLE $\widehat{\Lambda}_n^{ps}$ and the NPMLE $\widehat{\Lambda}_n$, we describe two Monte Carlo studies in this section.

Before describing the Monte Carlo experiments, we first explain our approach to estimation of the asymptotic relative efficiency of two estimators in this type of problem.

To study the asymptotic relative efficiency between two estimators, let us suppose that the estimators $\hat{a}_{n_a}$, and $\hat{b}_{n_b}$ have the same asymptotic distributions up to positive constants a and b: $n_a^{1/3}(\hat{a}_n - \bar{a}) \to_d a^{1/3}\mathbb{Z}$, and $n_b^{1/3}(\hat{b}_n - \bar{b}) \to_d b^{1/3}\mathbb{Z}$, respectively, where $\mathbb{Z}$ is a known random variable. Thus $\mathrm{Var}(\hat{a}_{n_a}) \cong n_a^{-2/3}a^{2/3}\mathrm{Var}(\mathbb{Z})$ and $\mathrm{Var}(\hat{b}_{n_b}) \cong n_b^{-2/3}b^{2/3}\mathrm{Var}(\mathbb{Z})$. If we ask two estimators to have the same variance, asymptotically, we find that this forces $(n_b/n_a)^{2/3} = (b/a)^{2/3}$. This implies that the asymptotic relative efficiency of two estimators with $n^{1/3}-$ convergence rate is simply the cube of the ratio of constants that appear in the asymptotic distributions: $\lim n_b/n_a = b/a$. On the other hand, we note that $\mathrm{Var}(\hat{a}_n) \cong n^{-2/3}a^{2/3}\mathrm{Var}(\mathbb{Z})$, and $\mathrm{Var}(\hat{b}_n) \cong n^{-2/3}b^{2/3}\mathrm{Var}(\mathbb{Z})$, and hence $\lim(n_b/n_a) = b/a \cong (\mathrm{Var}(\hat{b}_n)/\mathrm{Var}(\hat{a}_n))^{3/2}$. Based on the above arguments, we plot the 3/2 power of the ratio of sample variance of the NPMLE to the variance of the NPMPLE to obtain an approximation to the asymptotic relative efficiency.

SIMULATION 1 (Data from the Poisson process model). Let

$$\{(\underline{N}_i, \underline{T}_i, K_i) : i = 1, 2, \ldots, n\}$$

be a random sample, where

$$K_i \in \{1, 2, 3, 4, 5, 6\} \quad \text{for each } i = 1, 2, \ldots, n,$$

and $P(K_i = k) = 1/6$ for $k = 1, 2, 3, 4, 5, 6$. Then $\underline{T}_i = (T_{K_i,1}^{(i)}, T_{K_i,2}^{(i)}, \ldots, T_{K_i,K_i}^{(i)})$ are made from the order statistics of K_i random observations, generated from the distribution, **Unif(0,10)**. To make observation time points being possibly tied, and separated, $T_{K_i,j}^{(i)}$, $j = 1, 2, \ldots, K_i$, are rounded to 2^{nd} decimal point.

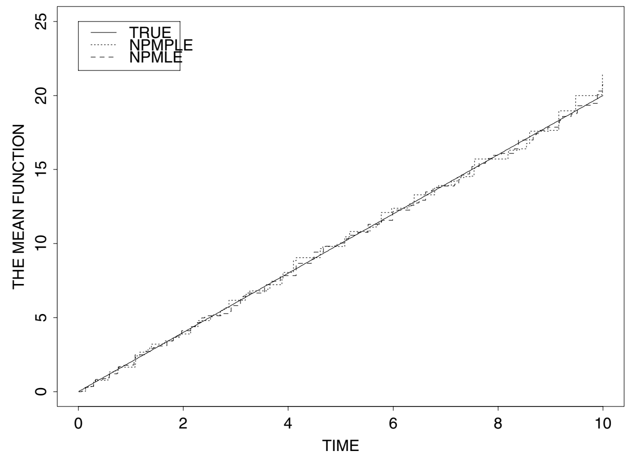

FIG. 2. *The NPMPLE and the NPMLE of the mean function based on panel count data, generated from a Poisson process.*

Panel counts $\underline{N}_i = (N^{(i)}_{K_i,1}, N^{(i)}_{K_i,2}, \dots, N^{(i)}_{K_i,K_i})$ are generated from Poisson$(2t)$, that is,

$$N^{(i)}_{K_i,j} - N^{(i)}_{K_i,j-1} \sim \text{Poisson}\left[2(T^{(i)}_{K_i,j} - T^{(i)}_{K_i,j-1})\right].$$

We chose $n = 100$ and 1000, $\eta = 0.000001$ and $\varepsilon = 0.2$ (ε is the line search parameter) to run the simulations. Since the NPMPLE $\widehat{\Lambda}^{ps}_n$ is consistent by Theorem 4.1, it is a good choice as a starting point for the Modified Iterative Convex Minorant (MICM) algorithm to compute $\widehat{\Lambda}_n$. However, the NPMPLE may yield a value of negative infinity for log likelihood function $\phi(\underline{u})$. To avoid this difficulty, we used a piecewise linear interpolation between points of jump so that the starting point for the MICM is strictly increasing. Based on these preparations, the MICM stops at step 195. Figure 2 displays both the NPMPLE and NPMLE of the mean function, along with the true mean function $\Lambda_0(t) = 2t$. Apparently, both the NPMPLE and NPMLE converge to the true mean function. Since the NPMLE appears to have more jumps than the NPMPLE, it suggests that the NPMLE has less variability than the NPMPLE, in agreement with Theorems 4.3 and 4.4.

To assess the variability of the estimators, we carried out a Monte Carlo study by repeating the simulation 1000 times for $n = 100$ and 100 times for $n = 1000$. The pointwise variances at time points $t = 1.5, 2.0, 2.5, 3.0, \dots, 9.5$ are calculated via the sample variances from 1000 and 100 runs respectively for the two sample sizes. Tables 1 and 3 list the bias and standard errors at those points for both the NPMPLE and NPMLE, respectively. Asymptotic unbiasedness for both estimators appears to be true, and the NPMLE does indeed have smaller variances at all these points, as expected. The plot of

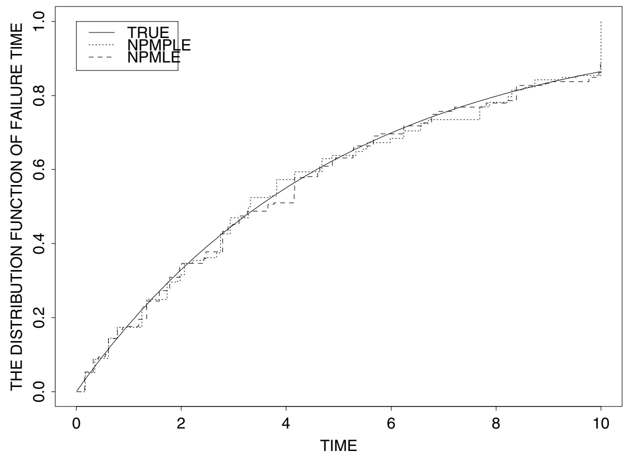

FIG. 3. *The NPMPLE and the NPMLE of the mean function based on panel count data, not generated from a Poisson process.*

estimated relative efficiency NPMPLE with respect to the NPMLE is plotted in Figures 4 and 5 (solid lines). As can be seen from Figures 4 and 5, the NPMLE tends to be more efficient than the NPMPLE: at most of those time points, the estimated relative efficiency of the NPMPLE is about 20% for $n = 1000$ and below 50% for $n = 100$.

SIMULATION 2 (Data from a one-jump counting process). Let

$$\{\underline{N}_i, \underline{T}_i, K_i) : i = 1, 2, \dots, n\}$$

be a random sample, where K_i, $\underline{T}_i$ are generated following the same schemes as those in simulation 1.

Let V_i denote the failure time, generated from the distribution Exp(−0.2). Panel counts $\underline{N}_i = (N_{K_i,1}^{(i)}, N_{K_i,2}^{(i)}, \dots, N_{K_i,K_i}^{(i)})$, $i = 1, 2, \dots, n$ are given by

$$N_{K_i,j}^{(i)} = 1_{[V_i \le T_{K_i,j}^{(i)}]} \qquad \text{for } j = 1, 2, \dots, K_i.$$

Because the counting process in this simulation always has just one jump, this situation can be viewed as "mixed case interval censored data," as described by Schick and Yu (1999). Because of this, at most two terms are needed in forming the likelihood function, and this means that a data reduction procedure analogous to that described in Definition 1.1 of Groeneboom and Wellner (1992) is necessary in the coding to simplify the algorithm.

We chose $n = 100$ and 3000, $\eta = 0.000001$ and $\varepsilon = 0.2$ to run the simulations. Again, the linearly interpolated NPMPLE $\widehat{\Lambda}_n^{ps}$ was chosen as the starting point for the MICM algorithm. Under these tolerances, the MICM procedure stopped after 174 iterative steps for a simulated sample. Figure 3

TABLE 1
Comparison of bias and standard deviation between the NPMPLE and the NPMLE of the mean function, based on the data generated from a Poisson process, $n = 1000$

	NPMPLE			NPMLE	
Time	**Monte-Carlo Bias**	**Monte-Carlo s.d.**	**Theoretical s.d.**	**Monte-Carlo Bias**	**Monte-Carlo s.d.**
1.5	3.036×10^{-2}	0.228	0.203	-2.359×10^{-2}	0.181
2.0	5.424×10^{-2}	0.225	0.223	3.410×10^{-2}	0.181
2.5	7.207×10^{-3}	0.246	0.241	2.748×10^{-2}	0.173
3.0	1.017×10^{-2}	0.252	0.256	3.891×10^{-2}	0.184
3.5	3.764×10^{-2}	0.267	0.269	1.625×10^{-2}	0.155
4.0	-7.724×10^{-3}	0.276	0.281	1.905×10^{-2}	0.186
4.5	1.229×10^{-1}	0.327	0.291	5.283×10^{-2}	0.170
5.0	7.545×10^{-2}	0.329	0.303	2.787×10^{-2}	0.171
5.5	4.978×10^{-3}	0.326	0.313	2.106×10^{-2}	0.172
6.0	-1.871×10^{-2}	0.321	0.322	2.176×10^{-2}	0.193
6.5	-6.816×10^{-2}	0.312	0.331	3.857×10^{-2}	0.209
7.0	-1.110×10^{-1}	0.392	0.339	-5.941×10^{-5}	0.194
7.5	2.308×10^{-2}	0.401	0.347	3.575×10^{-2}	0.209
8.0	-3.528×10^{-4}	0.364	0.354	4.599×10^{-3}	0.201
8.5	1.041×10^{-2}	0.417	0.362	4.738×10^{-2}	0.217
9.0	4.519×10^{-2}	0.423	0.369	4.556×10^{-2}	0.230
9.5	1.201×10^{-1}	0.389	0.375	5.382×10^{-2}	0.215

TABLE 2
Comparison of bias and standard deviation between the NPMPLE and the NPMLE of the mean function, based on the data not generated from a Poisson process, $n = 3000$

	NPMPLE			NPMLE	
Time	**Monte-Carlo Bias**	**Monte-Carlo s.d.**	**Theoretical s.d.**	**Monte-Carlo Bias**	**Monte-Carlo s.d.**
1.5	-3.233×10^{-3}	0.0253	0.0236	-1.218×10^{-4}	0.0216
2.0	-4.290×10^{-3}	0.0229	0.0240	-4.877×10^{-3}	0.0207
2.5	2.040×10^{-3}	0.0242	0.0238	2.285×10^{-3}	0.0168
3.0	-7.834×10^{-4}	0.0279	0.0233	1.596×10^{-3}	0.0198
3.5	-1.851×10^{-3}	0.0250	0.0226	8.456×10^{-4}	0.0172
4.0	2.008×10^{-3}	0.0246	0.0218	8.642×10^{-4}	0.0171
4.5	3.180×10^{-3}	0.0254	0.0209	1.437×10^{-3}	0.0172
5.0	1.141×10^{-3}	0.0229	0.0199	-3.361×10^{-3}	0.0159
5.5	-7.662×10^{-4}	0.0187	0.0190	-1.944×10^{-3}	0.0148
6.0	-7.295×10^{-3}	0.0192	0.0181	-2.074×10^{-3}	0.0141
6.5	-7.083×10^{-4}	0.0181	0.0171	-1.502×10^{-4}	0.0142
7.0	-4.248×10^{-3}	0.0178	0.0162	1.495×10^{-3}	0.0136
7.5	5.800×10^{-6}	0.0180	0.0153	-1.770×10^{-3}	0.0125
8.0	5.580×10^{-4}	0.0153	0.0145	2.525×10^{-3}	0.0109
8.5	-1.938×10^{-3}	0.0155	0.0136	-1.702×10^{-3}	0.0120
9.0	-2.494×10^{-3}	0.0147	0.0128	-2.194×10^{-3}	0.0112
9.5	1.889×10^{-3}	0.0149	0.0121	-6.706×10^{-5}	0.0112

TABLE 3
Comparison of bias and standard deviation between the NPMPLE and the NPMLE of the mean function, based on the data generated from a Poisson process, $n = 100$

	NPMPLE		NPMLE	
Time	**Monte-Carlo Bias**	**Monte-Carlo s.d.**	**Monte-Carlo Bias**	**Monte-Carlo s.d.**
1.5	−0.04776	0.46826	−0.02716	0.39500
2.0	−0.06374	0.51756	−0.03656	0.42282
2.5	−0.04266	0.54970	−0.04196	0.43752
3.0	−0.05531	0.59206	−0.02538	0.45651
3.5	0.00699	0.64312	0.00043	0.46550
4.0	−0.05989	0.67109	−0.02262	0.46200
4.5	0.02752	0.69124	0.00656	0.48900
5.0	−0.02633	0.69978	−0.00998	0.49201
5.5	−0.04738	0.71900	−0.01602	0.50250
6.0	−0.03009	0.74609	−0.01610	0.50899
6.5	−0.08780	0.77180	−0.00761	0.53868
7.0	−0.08836	0.83721	−0.02580	0.54304
7.5	−0.00534	0.90201	0.01187	0.55320
8.0	0.02755	0.85949	−0.02862	0.57569
8.5	−0.01033	0.92738	−0.01616	0.59424
9.0	−0.01438	0.94491	−0.01021	0.62486
9.5	0.11517	0.99226	−0.00334	0.65618

TABLE 4
Comparison of bias and standard deviation between the NPMPLE and the NPMLE of the mean function, based on the data not generated from a Poisson process, $n = 100$

	NPMPLE		NPMLE	
Time	**Monte-Carlo Bias**	**Monte-Carlo s.d.**	**Monte-Carlo Bias**	**Monte-Carlo s.d.**
1.5	−0.01917	0.08820	0.00043	0.08309
2.0	−0.01345	0.08761	−0.00023	0.08104
2.5	−0.00726	0.08941	0.00356	0.08076
3.0	−0.00630	0.08615	0.00324	0.07862
3.5	−0.00549	0.08445	0.00344	0.07564
4.0	−0.00148	0.08227	0.00645	0.07632
4.5	0.00200	0.08198	0.00417	0.07161
5.0	0.00300	0.07850	0.00139	0.07005
5.5	−0.00024	0.07248	0.00385	0.06766
6.0	−0.00079	0.07192	0.00425	0.06524
6.5	0.00270	0.06944	0.00250	0.06361
7.0	0.00445	0.06671	0.00073	0.05944
7.5	0.00137	0.06480	0.00137	0.05855
8.0	0.00371	0.06258	−0.00162	0.05836
8.5	0.00686	0.06042	−0.00637	0.05723
9.0	0.01403	0.06202	−0.00706	0.05599
9.5	0.02581	0.06572	−0.00260	0.06305

displays both the NPMPLE and NPMLE of the distribution function of failure time $F(t) = P(T \leq t) = 1 - \exp(-0.2t)$ (the mean function is just the distribution function of failure time in a one jump process).

The same phenomenon as in simulation 1 happens in this study that both the NPMPLE and the NPMLE appear to converge to the true target function $F(t)$, and the NPMLE tends to have less variation than the NPMPLE. This study serves as a numerical evidence to verify that both the NPMPLE $\widehat{\Lambda}_n^{ps}$ and the NPMLE $\widehat{\Lambda}_n$ based on the Poisson process assumption are robust with respect to the actual distribution of the underlying counting process.

A Monte Carlo study based on 1000 runs for $n = 100$, 100 runs for $n = 3000$ of simulation 2 was also implemented to compare the NPMPLE and NPMLE. Tables 2 and 4 show the results of the study by listing the bias and standard errors of the estimators at points $t = 1.5, 2.0, 2.5, 3.0, \dots, 9.5$. Apparently, these estimators are both asymptotically unbiased with the NPMLE being less variable. The plot of estimated relative efficiency NPMPLE with respect to the NPMLE is plotted in Figures 4 and 5 (dotted lines). As can be seen from Figures 4 and 5, the NPMLE tends to be more efficient than the NPMPLE: at most of those time points, the estimated relative efficiency of the NPMPLE is about 30% - 40% in this second situation for $n = 3000$, and about 80% for $n = 100$.

While the NPMLE is apparently more efficient than the NPMPLE, it should be noted that the implementation of the NPMLE is much more involved than that of the NPMPLE, requires relatively delicate programming structure, and also much more computing time. Here all these implementations were coded in the *C language* and the plots were produced using S-plus. For executable versions of the above algorithms described in Section 3 and the

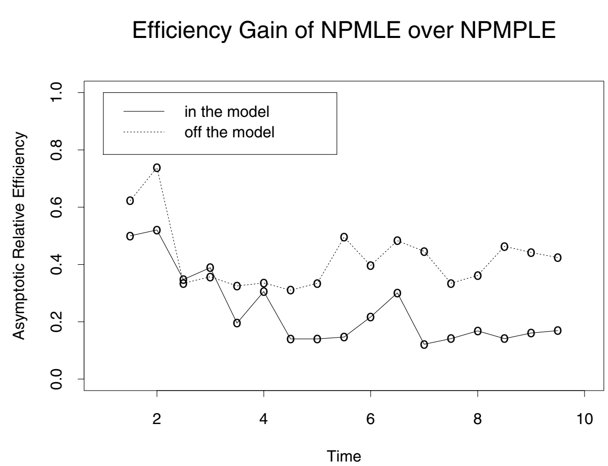

FIG. 4. *The asymptotic relative efficiency of the NPMLE vs. the NPMPLE,* $n = 1000$ (*in the model*) *and* $n = 3000$ (*off the model*).

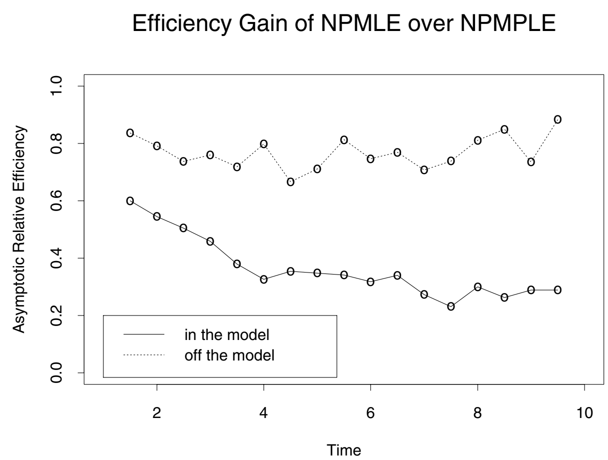

FIG. 5. *The asymptotic relative efficiency of the NPMLE vs. the NPMPLE,* $n = 100$.

above simulations (implemented in C), see Wellner's web site for software, `http://www.stat.washington.edu/jaw/RESEARCH/SOFTWARE`.

7. Asymptotic results: Proofs. One of the strengths of modern empirical process theory is the generality allowed in the choice of the sample space. Here we exploit that generality. We will use several results from Van der Vaart and Wellner (1996) to prove Theorems 3.1 and 3.2.

Throughout this section we let C stand for a generic constant which may change from line to line in the proofs.

PROOF OF THEOREM 4.1. We begin by outlining the main steps in the proof. The basic idea is to show that for almost all ω the sequence $\{\widehat{\Lambda}_n^{ps}(\cdot,\omega)\}$ is sequentially compact for the topology of pointwise convergence and that every pointwise limit $\Lambda^\dagger(\cdot,\omega)$ of $\{\widehat{\Lambda}_n^{ps}(\cdot,\omega)\}$ must satisfy $\mathbb{M}^{ps}(\Lambda_0) \le \mathbb{M}^{ps}(\Lambda^\dagger)$ where $\mathbb{M}^{ps}$ is the population log-pseudolikelihood function defined below. Since $\mathbb{M}^{ps}$ has Λ_0 as its unique maximum, this then yields that every limit equals Λ_0 a.e. μ. This then yields $\widehat{\Lambda}_n^{ps} \to \Lambda_0$ a.e. μ almost surely. By the finiteness of μ and the dominated convergence theorem, this convergence yields $d(\widehat{\Lambda}_n^{ps}1_{[0,b]}, \Lambda_0 1_{[0,b]}) \to 0$ a.s. for every $b \le \tau$ for which $\limsup_{n\to\infty}\widehat{\Lambda}_n^{ps}(b) < \infty$ almost surely. This last conclusion and sequential compactness will follow from an argument which shows that $\limsup_{n\to\infty}\widehat{\Lambda}_n^{ps}(b) \le C/\mu([b,\tau])$ for every $b < \tau$.

Now for the detailed argument. Let $S_\mu \equiv \operatorname{supp}(\mu)$ and set $\tau_r = \sup(S_\mu)$. It follows from assumption A that $\tau_r < \infty$ and we can take $\tau = \tau_r$. Let

$$\mathbb{M}_n^{ps}(\Lambda) = \frac{1}{n} l_n(\Lambda|\underline{X}) = \mathbb{P}_n m_\Lambda^{ps} \qquad \text{and} \qquad \mathbb{M}^{ps}(\Lambda) = P m_\Lambda^{ps}$$

where

$$m_\Lambda^{ps}(X) = \sum_{j=1}^{K} \left[N_{K,j} \log \Lambda(T_{K,j}) - \Lambda(T_{K,j}) \right]. \tag{7.1}$$

Our proof of consistency will use the one-sided Glivenko-Cantelli Theorem A.1. To apply Theorem A.1, we first need an upper envelope for the class of functions $\mathscr{M}^{ps} \equiv \{m_\Lambda^{ps} : \Lambda \in \mathscr{F}\}$. Noting that $g(x) = a \log x - x$, $x > 0$, $a \geq 0$, is maximized by $x = a$ (with $0 \log 0 \equiv 0$), we find that

$$m_\Lambda^{ps}(X) \leq \sum_{j=1}^{K} \left[N_{K,j} \log(N_{K,j}) - N_{K,j} \right] \leq \sum_{j=1}^{K} N_{K,j} \log(N_{K,j}) = M_0^{ps}(X),$$

where $PM_0^{ps}(X) < \infty$ by Assumption C.

The NPMPLE $\widehat{\Lambda}_n^{ps}$ is given by

$$\widehat{\Lambda}_n^{ps} = \arg\max_{\Lambda \in \mathscr{F}} \mathbb{M}_n^{ps}(\Lambda).$$

Since

$$\mathbb{M}_n^{ps}(\widehat{\Lambda}_n^{ps}) \geq \mathbb{M}_n^{ps}((1-\varepsilon)\widehat{\Lambda}_n^{ps} + \varepsilon \Lambda_0),$$

it follows that

$$\lim_{\varepsilon \downarrow 0} \frac{\mathbb{M}_n^{ps}((1-\varepsilon)\widehat{\Lambda}_n^{ps} + \varepsilon\Lambda_0) - \mathbb{M}_n^{ps}(\widehat{\Lambda}_n^{ps})}{\varepsilon} \leq 0.$$

Evaluating this limit we find that

$$\mathbb{P}_n \left\{ \sum_{j=1}^{K} \left(\frac{N_{K,j}}{\widehat{\Lambda}_{n,K,j}} - 1 \right) \left(\Lambda_{0,K,j} - \widehat{\Lambda}_{n,K,j} \right) \right\} \leq 0 \tag{7.2}$$

where $\Lambda_{0,K,j} \equiv \Lambda_0(T_{K,j})$, $\widehat{\Lambda}_{n,K,j} \equiv \widehat{\Lambda}_n^{ps}(T_{K,j})$. Now (7.2) can be rewritten as

$$\mathbb{P}_n \left\{ \sum_{j=1}^{K} \left(N_{K,j} \frac{\Lambda_{0,K,j}}{\widehat{\Lambda}_{n,K,j}} + \widehat{\Lambda}_{n,K,j} \right) \right\} - \mathbb{P}_n \left\{ \sum_{j=1}^{K} \left(\Lambda_{0,K,j} + N_{K,j} \right) \right\} \leq 0,$$

where the second term on the left side converges a.s. by the strong law of large numbers and the Assumptions A and B to

$$P\left(\sum_{j=1}^{K} \left(\Lambda_{0,K,j} + N_{K,j} \right) \right) = 2E\left(\sum_{j=1}^{K} \Lambda_0(T_{K,j}) \right) = 2 \int \Lambda_0(t) d\mu(t) \equiv C < \infty.$$

Hence it follows that for every $b \in (0, \tau]$ one has, almost surely,

$$\begin{aligned} C &= \limsup_{n\to\infty} \mathbb{P}_n \left\{ \sum_{j=1}^{K} \left(\Lambda_{0,K,j} + N_{K,j} \right) \right\} \\ &\geq \limsup_{n\to\infty} \mathbb{P}_n \left\{ \sum_{j=1}^{K} \left(N_{K,j} \frac{\Lambda_{0,K,j}}{\widehat{\Lambda}_{n,K,j}} + \widehat{\Lambda}_{n,K,j} \right) \right\} \end{aligned}$$

$$
\begin{aligned}
&\geq \limsup_{n\to\infty} \mathbb{P}_n \left\{ \sum_{j=1}^{K} 1_{[b,\tau]}(T_{K,j}) \left(N_{K,j} \frac{\Lambda_{0,K,j}}{\widehat{\Lambda}_{n,K,j}} + \widehat{\Lambda}_{n,K,j} \right) \right\} \\
&\geq \limsup_{n\to\infty} \widehat{\Lambda}_n^{ps}(b) \mathbb{P}_n \left\{ \sum_{j=1}^{K} 1_{[b,\tau]}(T_{K,j}) \right\} \\
&= \limsup_{n\to\infty} \widehat{\Lambda}_n^{ps}(b) \mu([b,\tau]).
\end{aligned}
$$

Note that either $\mu(\{\tau\}) > 0$, or there exists $0 < b < \tau$ arbitrarily close to τ such that $\mu([b,\tau]) > 0$. We first complete the proof in the case when $\mu(\{\tau\}) > 0$. In this case we conclude that on a set with probability one we have

$$
\limsup_{n\to\infty} \widehat{\Lambda}_n^{ps}(b) \leq \frac{C}{\mu(\{\tau\})} \equiv M_\tau < \infty \tag{7.3}
$$

and hence the functions $\{\widehat{\Lambda}_n^{ps}\}$ are a.s. bounded on $[0,\tau]$. Since the sequence of functions $\{\widehat{\Lambda}_n^{ps}(t,\omega) : t \in [0,\tau]\}$ is uniformly bounded (for all $n \geq$ some N_ω), it follows from the Helly selection theorem that the sequence $\{\widehat{\Lambda}_n^{ps}(\cdot,\omega)\}$ has a subsequence $\{\widehat{\Lambda}_{n'}^{ps}(\cdot)\} \equiv \{\widehat{\Lambda}_{n'(\omega)}^{ps}(\cdot,\omega)\}$ which converges to a nondecreasing function $\Lambda^\dagger = \Lambda_\omega^\dagger$, defined on $[0,\tau]$ and taking values in $[0, M_\tau]$. Consider the class of functions

$$
\mathscr{M}_\tau^{ps} = \{ m_\Lambda^{ps} : \Lambda \in \mathscr{F}_\tau \}
$$

where

$$
\mathscr{F}_\tau \equiv \{ \Lambda \in \mathscr{F} : \Lambda(\tau) \leq M_\tau + 1 \}. \tag{7.4}
$$

Note that $\mathscr{F}_\tau$ is compact for the (pseudo-)metric d. Moreover, the function $\Lambda \to m_\Lambda^{ps}(x)$ is upper semicontinuous in Λ for P almost all x, and $m_\Lambda^{ps}(x) \leq M_0^{ps}(x)$ for all x and $\Lambda \in \mathscr{F}_\tau$ where M_0^{ps} is integrable by Assumption C. [To see the upper semi-continuity, suppose that $\{\Lambda_m\}$ is a sequence with $d(\Lambda_m, \Lambda) \to 0$; then $\Lambda_m(t) \to \Lambda(t)$ for μ almost every t, and this implies that $m_{\Lambda_m}^{ps}(x) \to m_\Lambda^{ps}(x)$ for P almost all x.] Thus Theorem A.1 yields

$$
\limsup_{n\to\infty} \sup_{\Lambda\in\mathscr{F}_\tau} (\mathbb{P}_n - P)(m_\Lambda^{ps}(X)) \leq 0 \qquad \text{a.s.} \tag{7.5}
$$

Since $\mathbb{M}_n^{ps}(\Lambda_0) \to \mathbb{M}^{ps}(\Lambda_0)$ a.s. by the strong law of large numbers and $\mathbb{M}_n^{ps}(\Lambda_0) \leq \mathbb{M}_n^{ps}(\widehat{\Lambda}_n^{ps})$, it follows that

$$
\mathbb{M}^{ps}(\Lambda_0) \leq \liminf_{n\to\infty} \mathbb{M}_n^{ps}(\widehat{\Lambda}_n^{ps}) \qquad \text{a.s.} \tag{7.6}
$$

Hence it follows from (7.5) and (7.3) that

$$
\limsup_{n'\to\infty} \mathbb{M}_{n'}^{ps}(\widehat{\Lambda}_{n'}^{ps}) \leq P(m_{\Lambda^\dagger}^{ps}) \tag{7.7}
$$

almost surely, where the last inequality follows from (7.5) and

$$
\limsup_{n'\to\infty} P(m_{\widehat{\Lambda}_{n'}^{ps}}^{ps}) \leq P(m_{\Lambda^\dagger}^{ps})
$$

for any subsequence $\widehat{\Lambda}_{n'}^{ps}$ converging almost surely to $\Lambda^\dagger$ on $[0,\tau]$; this follows from the upper semicontinuity of $\Lambda \to P(m_\Lambda^{ps})$ which is a consequence of Theorem A.1 together with pointwise convergence of $\widehat{\Lambda}_{n'}^{ps}$ on $[0,\tau]$. Combining (7.6) and (7.7) yields

$$
\begin{aligned}
0 &\le \mathbb{M}^{ps}(\Lambda^\dagger) - \mathbb{M}^{ps}(\Lambda_0) \\
&= -(Pm_{\Lambda_0}^{ps} - Pm_{\Lambda^\dagger}^{ps}) \\
&= -P\left(\sum_{j=1}^{K} \Lambda_{0,K,j} \log \frac{\Lambda_{0,K,j}}{\Lambda_{K,j}^\dagger} - \left(\frac{\Lambda_{0,K,j}}{\Lambda_{K,j}^\dagger} - 1\right)\Lambda_{K,j}^\dagger\right) \\
&= -P\left(\sum_{j=1}^{K} \Lambda_{K,j}^\dagger h\left(\frac{\Lambda_{0,K,j}}{\Lambda_{K,j}^\dagger}\right)\right) \\
&= -\int \Lambda^\dagger(t) h\left(\frac{\Lambda_0(t)}{\Lambda^\dagger(t)}\right) d\mu(t) \le 0
\end{aligned}
\tag{7.8}
$$

since $h(x) \equiv x(\log x - 1) + 1 \ge 0$ with equality if and only if $x = 1$. This implies that $\Lambda^\dagger(t) = \Lambda_0(t)$ a.e. μ. Since this is true for any convergent subsequence, we conclude that all the limits $\Lambda^\dagger$ of subsequences of $\{\widehat{\Lambda}_n^{ps}\}$ are equal to Λ_0 a.e. μ. Since $\Lambda_0(t) \le \Lambda_0(\tau) \le M$ for $t \le \tau$, this implies that $\lim_{n\to\infty} \widehat{\Lambda}_n^{ps}(t) = \Lambda_0(t) \le \Lambda_0(\tau)$ almost surely for μ almost all $t \in [0,\tau]$. It follows by the dominated convergence theorem [with dominating functions $\Lambda_0(\tau)+1$ since μ is a finite measure] that $d(\widehat{\Lambda}_n^{ps}, \Lambda_0) \to 0$ a.s.

Now suppose that $\mu(\{\tau\}) = 0$. By the definition of τ_r it then follows that $\mu([b,\tau]) > 0$ for every $b < \tau$. Then we conclude that on a set with probability one we have

$$
\limsup_{n\to\infty} \widehat{\Lambda}_n^{ps}(b) \le \frac{C}{\mu([b,\tau])} \equiv M_b < \infty \tag{7.9}
$$

and hence the functions $\{\widehat{\Lambda}_n^{ps}\}$ are a.s. bounded on $[0,b]$. Since the sequence of functions $\{\widehat{\Lambda}_n^{ps}(t,\omega) : t \in [0,b]\}$ is uniformly bounded (for all $n \ge$ some N_ω), it follows from the Helly selection theorem that the sequence $\{\widehat{\Lambda}_n^{ps}(\cdot,\omega)\}$ has a subsequence $\{\widehat{\Lambda}_{n'}^{ps}(\cdot)\} \equiv \{\widehat{\Lambda}_{n'(\omega)}^{ps}(\cdot,\omega)\}$ which converges to a nondecreasing function $\Lambda^\dagger = \Lambda_{\omega,b}^\dagger$, defined on $[0,b]$ for each $b < \tau$ taking values in $[0, M_b]$. By considering a sequence of b's converging up to τ we get a function $\Lambda^\dagger = \Lambda_\omega^\dagger$ which is well-defined on $[0,\tau)$. Set $\Psi_b(X) \equiv 1_{[0,b]}(T_{K,K})$, and for $b < \tau$ consider the class of functions

$$
\mathscr{M}_b^{ps} = \{\Psi_b(X) m_\Lambda^{ps}(X) : \Lambda \in \mathscr{F}_b\}
$$

where

$$
\mathscr{F}_b \equiv \left\{\Lambda \in \mathscr{F} \,\middle|\, \Lambda(b) \le M_b + 1,\ \Lambda(t) = \Lambda(b),\ t \ge b\right\}. \tag{7.10}
$$

Note that $\mathscr{F}_b$ is compact for every $b < \tau$. Moreover, the functions $\Lambda \to \Psi_b(x) m_\Lambda^{ps}(x)$ are upper semicontinuous in Λ for each fixed x and $\Lambda \to$

$\Psi_b(x)m_\Lambda^{ps}(x) \le M_0^{ps}(x)$ for all x and $\Lambda \in \mathcal{F}_b$ where M_0^{ps} is integrable by Assumption C. Thus Theorem A.1 yields

$$\limsup_{n\to\infty} \sup_{\Lambda\in\mathcal{F}_b} (\mathbb{P}_n - P)(\Psi_b(X)m_\Lambda^{ps}(X)) \le 0 \qquad \text{a.s.} \tag{7.11}$$

Since $\mathbb{M}_n^{ps}(\Lambda_0) \to \mathbb{M}^{ps}(\Lambda_0)$ a.s. by the strong law of large numbers and $\mathbb{M}_n^{ps}(\Lambda_0) \le \mathbb{M}_n^{ps}(\widehat{\Lambda}_n^{ps})$, it follows that

$$\mathbb{M}^{ps}(\Lambda_0) \le \liminf_{n\to\infty} \mathbb{M}_n^{ps}(\widehat{\Lambda}_n^{ps}) \qquad \text{a.s.} \tag{7.12}$$

Now we can write, for any $\Lambda \in \mathcal{F}$,

$$\begin{aligned}\mathbb{M}_n^{ps}(\Lambda) &= \mathbb{P}_n(\Psi_b m_\Lambda^{ps}) + \mathbb{P}_n((1-\Psi_b)m_\Lambda^{ps})\\ &\le (\mathbb{P}_n - P)(\Psi_b m_\Lambda^{ps}) + P(\Psi_b m_\Lambda^{ps}) + \mathbb{P}_n((1-\Psi_b)M_0^{ps}).\end{aligned}$$

Hence it follows from (7.11) and (7.9) that

$$\limsup_{n'\to\infty} \mathbb{M}_n^{ps}(\widehat{\Lambda}_{n'}^{ps}) \le P(\Psi_b m_{\Lambda^\dagger}^{ps}) + P((1-\Psi_b)M_0^{ps}(X)) \quad \text{a.s.} \tag{7.13}$$

where the last inequality follows from (7.11) and

$$\limsup_{n'\to\infty} P\left(\Psi_b m_{\widehat{\Lambda}_{n'}^{ps}}^{ps}\right) \le P\left(\Psi_b m_{\Lambda^\dagger}^{ps}\right)$$

for any subsequence $\widehat{\Lambda}_{n'}^{ps}$ converging almost surely to $\Lambda^\dagger$ on $[0,\tau)$; this follows from the upper semicontinuity of $\Lambda \to P(\Psi_b m_\Lambda^{ps})$ which is a consequence of Theorem 8.1 together with pointwise convergence of $\widehat{\Lambda}_{n'}^{ps}$ on $[0,\tau)$. Letting $b \uparrow \tau$ in (7.13) yields, using assumptions A and C, on a set with probability one,

$$\limsup_{n'\to\infty} \mathbb{M}_{n'}^{ps}(\widehat{\Lambda}_{n'}^{ps}) \le P(m_{\Lambda^\dagger}^{ps}) = \mathbb{M}^{ps}(\Lambda^\dagger). \tag{7.14}$$

Combining (7.12) and (7.14) yields (7.8) and hence $\Lambda^\dagger(t) = \Lambda_0(t)$ a.e. μ as before. Since this is true for any convergent subsequence, we conclude that all the limits $\Lambda^\dagger$ of subsequences of $\{\widehat{\Lambda}_n^{ps}\}$ are equal to Λ_0 a.e. μ. Since $\Lambda_0(t) \le \Lambda_0(\tau) \le M$ for $t \le \tau$, this implies that $\lim_{n\to\infty} \widehat{\Lambda}_n^{ps}(t) = \Lambda_0(t) \le \Lambda_0(\tau)$ almost surely for μ almost all $t \in [0,\tau)$. It follows by the dominated convergence theorem [with dominating functions $\Lambda_0(b)+1$ since μ is a finite measure] that $d(\widehat{\Lambda}_n^{ps} 1_{[0,b]}, \Lambda_0 1_{[0,b]}) \to_{\text{a.s.}} 0$. □

PROOF OF THEOREM 4.2. For proofs of inequalities (4.5) and (4.7), see Wellner and Zhang (1998).

Now let

$$\mathbb{M}_n(\Lambda) = \frac{1}{n} l_n(\Lambda|\underline{X}) = \mathbb{P}_n m_\Lambda \quad \text{and} \quad \mathbb{M}(\Lambda) = P m_\Lambda \tag{7.15}$$

where

$$
\begin{aligned}
m_\Lambda(X) &= \sum_{j=1}^{K}[(N_{K,j} - N_{K,j-1})\log(\Lambda(T_{K,j}) - \Lambda(T_{K,j-1})) \\
&\qquad\qquad\qquad\qquad -(\Lambda(T_{K,j}) - \Lambda(T_{K,j-1}))] \\
&\equiv \sum_{j=1}^{K}\left[\Delta N_{K,j}\log(\Delta\Lambda_{K,j}) - \Delta\Lambda_{K,j}\right].
\end{aligned}
$$

Our consistency proof will again use the one-sided Glivenko-Cantelli Theorem A.1. To apply Theorem A.1 we first need an upper envelope for the class of functions $\mathscr{M} \equiv \{m_\Lambda : \Lambda \in \mathscr{F}\}$. Noting that $g(x) = a\log x - x$ is maximized by $x = a$, we find that

$$
\begin{aligned}
m_\Lambda(X) &\le \sum_{j=1}^{K}\left[\Delta N_{K,j}\log(\Delta N_{K,j}) - \Delta N_{K,j}\right] \\
&\le \sum_{j=1}^{K}\Delta N_{K,j}\log(\Delta N_{K,j}) \equiv M_0(X),
\end{aligned}
$$

where $PM_0(X) < \infty$ by Assumption D. Now $\widehat{\Lambda}_n = \operatorname{argmax}_{\Lambda\in\mathscr{F}}\mathbb{M}_n(\Lambda)$ where $\mathscr{F}$ is as defined in (4.4). Then, since $\mathbb{M}_n(\widehat{\Lambda}_n) \ge \mathbb{M}_n((1-\varepsilon)\widehat{\Lambda}_n + \varepsilon\Lambda_0)$, it follows that

$$
\lim_{\varepsilon\downarrow 0}\frac{\mathbb{M}_n((1-\varepsilon)\widehat{\Lambda}_n + \varepsilon\Lambda_0) - \mathbb{M}_n(\widehat{\Lambda}_n)}{\varepsilon} \le 0.
$$

Evaluating this limit we find that

$$
\mathbb{P}_n\left\{\sum_{j=1}^{K}\left(\frac{\Delta N_{K,j}}{\Delta\widehat{\Lambda}_{n,K,j}} - 1\right)\left(\Delta\Lambda_{0,K,j} - \Delta\widehat{\Lambda}_{n,K,j}\right)\right\} \le 0 \tag{7.16}
$$

where $\Delta\Lambda_{0,K,j} \equiv \Lambda_0(T_{K,j}) - \Lambda_0(T_{K,j-1})$, $\Delta\widehat{\Lambda}_{n,K,j} \equiv \widehat{\Lambda}_n(T_{K,j}) - \widehat{\Lambda}_n(T_{K,j-1})$. Now (7.16) can be rewritten as

$$
\begin{aligned}
0 &\ge \mathbb{P}_n\left\{\sum_{j=1}^{K}\left(\Delta N_{K,j}\frac{\Delta\Lambda_{0,K,j}}{\Delta\widehat{\Lambda}_{n,K,j}} + \Delta\widehat{\Lambda}_{n,K,j}\right)\right\} - \mathbb{P}_n\left\{\sum_{j=1}^{K}\left(\Delta\Lambda_{0,K,j} + \Delta N_{K,j}\right)\right\} \\
&= \mathbb{P}_n\left\{\sum_{j=1}^{K}\left(\Delta N_{K,j}\frac{\Delta\Lambda_{0,K,j}}{\Delta\widehat{\Lambda}_{n,K,j}} + \Delta\widehat{\Lambda}_{n,K,j}\right)\right\} - \mathbb{P}_n\left\{\Lambda_{0,K,K} + N_{K,K}\right\}
\end{aligned}
$$

where the second term on the left side converges a.s. by the strong law of large numbers and the Assumptions A and B, to

$$
\begin{aligned}
P\left(\Lambda_{0,K,K} + N_{K,K}\right) &= 2P\left(\Lambda_0(T_{K,K})\right) \\
&= 2\sum_{k=1}^{\infty}P(K=k)\int\Lambda_0(t)\,dG_{k,k}(t) \equiv C < \infty.
\end{aligned}
$$

Hence it follows that for any subset A of $\{t \in R^k : 0 \le t_1 \le t_2 \le \cdots \le t_k\}$ and $\underline{T}_K = (T_{K,1}, \ldots, T_{K,K})$ we have, almost surely,

$$
\begin{aligned}
C &= \limsup_{n\to\infty} \mathbb{P}_n \{\Lambda_{0,K,K} + N_{K,K}\} \\
&\ge \limsup_{n\to\infty} \mathbb{P}_n \left\{ \sum_{j=1}^{K} \Delta N_{K,j} \frac{\Delta\Lambda_{0,K,j}}{\Delta\widehat{\Lambda}_{n,K,j}} + \widehat{\Lambda}_{n,K,K} \right\} \qquad (7.17)\\
&\ge \limsup_{n\to\infty} \mathbb{P}_n \left\{ 1_A(\underline{T}_K) \left(\sum_{j=1}^{K} \Delta N_{K,j} \frac{\Lambda_{0,K,j}}{\Delta\widehat{\Lambda}_{n,K,j}} + \widehat{\Lambda}_{n,K,K} \right) \right\}.
\end{aligned}
$$

Let $S_\nu \equiv \mathrm{supp}(\nu)$ where ν is the measure defined by (4.3) and set $\tau \equiv \sup(S_\nu)$. It follows from (7.18) with $A = R^{k-1} \times [b, \tau]$ that for every $b \in (0, \tau]$

$$
\begin{aligned}
C &\ge \limsup_{n\to\infty} \mathbb{P}_n \left\{ 1_{[b,\tau]}(T_{K,K}) \widehat{\Lambda}_{n,K,K} \right\} \\
&\ge \limsup_{n\to\infty} \widehat{\Lambda}_n(b) \mathbb{P}_n 1_{[b,\tau]}(T_{K,K}) = \limsup_{n\to\infty} \widehat{\Lambda}_n(b)\nu([b,\tau])
\end{aligned}
$$

almost surely by the strong law of large numbers. Henceforth the proof is almost exactly the same as for the NPMPLE $\widehat{\Lambda}_n^{ps}$. By similar arguments we conclude that

$$
\mathbb{M}(\Lambda_0) \le \liminf_{n\to\infty} \mathbb{M}_n(\widehat{\Lambda}_n) \qquad \text{a.s.} \tag{7.18}
$$

and

$$
\limsup_{n'\to\infty} \mathbb{M}_n(\widehat{\Lambda}_{n'}) \le P(\Psi_b m_{\Lambda^\dagger}) + P((1-\Psi_b)M_0(X)) \tag{7.19}
$$

almost surely. Letting $b \uparrow \tau$ in (7.19) yields, using assumptions A and D, on a set with probability one,

$$
\limsup_{n'\to\infty} \mathbb{M}_{n'}(\widehat{\Lambda}_{n'}) \le P(m_{\Lambda^\dagger}) = \mathbb{M}(\Lambda^\dagger). \tag{7.20}
$$

Combining (7.18) and (7.20) yields

$$
\begin{aligned}
0 &\le \mathbb{M}(\Lambda^\dagger) - \mathbb{M}(\Lambda_0) = -(Pm_{\Lambda_0} - Pm_{\Lambda^\dagger}) \\
&= -P\left(\sum_{j=1}^{K} \Delta\Lambda_{0,K,j} \log \frac{\Delta\Lambda_{0,K,j}}{\Delta\Lambda^\dagger_{K,j}} - \left(\frac{\Delta\Lambda_{0,K,j}}{\Delta\Lambda^\dagger_{K,j}} - 1 \right) \Delta\Lambda^\dagger_{K,j} \right) \\
&= -P\left(\sum_{j=1}^{K} \Delta\Lambda^\dagger_{K,j} h\left(\frac{\Delta\Lambda_{0,K,j}}{\Delta\Lambda^\dagger_{K,j}} \right) \right) \\
&= -\int\int (\Lambda^\dagger(v) - \Lambda^\dagger(u)) h\left(\frac{\Lambda_0(v) - \Lambda_0(u)}{\Lambda^\dagger(v) - \Lambda^\dagger(u)} \right) d\mu_2(u,v) \le 0
\end{aligned}
$$

since $h(x) \equiv x(\log x - 1) + 1 \ge 0$ with equality if and only if $x = 1$. This implies that $\Delta\Lambda^\dagger \equiv \Lambda^\dagger(v) - \Lambda^\dagger(u) = \Lambda_0(v) - \Lambda_0(u) \equiv \Delta\Lambda_0$ a.e. μ_2, and since, by (4.7) the

$L_2(\mu_2)$ metric d_2 dominates the $L_2(\tilde{\mu})$ metric $\tilde{d}$ where μ and $\tilde{\mu}$ are mutually absolutely continuous, we conclude that $\Lambda^\dagger(t) = \Lambda_0(t)$ a.e. μ. Since this is true for any convergent subsequence, we conclude that all the limits $\Lambda^\dagger$ of subsequences of $\{\widehat{\Lambda}_n\}$ are equal to Λ_0 a.e. μ. Since $\Lambda_0(t) \le \Lambda_0(\tau) \le M$ for $t \le \tau$, this implies that $\lim_{n\to\infty} \widehat{\Lambda}_n(t) = \Lambda_0(t) \le \Lambda_0(\tau)$ almost surely for μ almost all $t \in [0, \tau)$. It follows by the dominated convergence theorem [with dominating functions $\Lambda_0(b) + 1$ since μ is a finite measure] that $d(\widehat{\Lambda}_n 1_{[0,b]}, \Lambda_0 1_{[0,b]}) \to_{\text{a.s.}} 0$. □

PROOF OF THEOREM 4.3. Here we adopt the notation in Section 2. Let $c(s) : [0, \sum_{j=1}^m w_j] \to \mathbb{R}^+$ be the cumulative-sum diagram; that is,

$$c(s) = \sum_{j\le i} w_j \bar{N}_j \quad \text{for } s \in \left(\sum_{j\le i-1} w_j, \sum_{j\le i} w_j\right], \quad c(0) = 0.$$

From the characteriztion of the estimator $\widehat{\Lambda}_n^{ps}$ given in Lemma 2.2, we know that $\widehat{\Lambda}_n^{ps}(s_i)$ can be expressed as the slope H^* in the interval $(\sum_{j\le i-1} w_j, \sum_{j\le i} w_j]$. Similar to the case of current status data [interval censoring case 1, Groeneboom and Wellner (1992), pages 91–95], it follows that

$$\left\{\widehat{\Lambda}_n^{ps}(s_i) \le a\right\} = \left\{\sup_s\{s : c(s) - as \text{ is minimal}\} \ge \sum_{j\le i} w_j\right\}. \tag{7.21}$$

Define two stochastic processes

$$V_n(t) = \sum_{i=1}^m w_i \bar{N}_i 1_{\{s_i \le t\}} \quad \text{and} \quad U_n(t) = \sum_{i=1}^m w_i 1_{\{s_i \le t\}}.$$

It follows easily that $c \equiv V_n \circ U_n^{-1}$. Hence (7.21) can be written as

$$\begin{aligned}
&\left\{\widehat{\Lambda}_n^{ps}(s_i) \le a\right\} \\
&\quad \equiv \left\{\sup_s\{s \in [0, \sum_{j=1}^m w_j] : V_n \circ U_n^{-1}(s) - as \ \ge \sum_{j\le i} w_j\right\} \\
&\quad = \left\{\sup_s\{s \in [0, s_m] : V_n(s) - aU_n(s) \text{ is minimal}\} \ge s_i\right\}.
\end{aligned}$$

Let

$$\widehat{S}_n(a) = \arg\min_s \{V_n(s) - aU_n(s)\},$$

where the largest value is chosen when multiple maximizers exist. By the change of variable $s = t_0 + hn^{-1/3}$ in the definition of $\widehat{S}_n$, one gets

$$\begin{aligned}
&\widehat{S}_n(\Lambda_0(t_0) + n^{-1/3}x) - t_0 \\
&\quad = n^{-1/3} \arg\min_h \left\{V_n(t_0 + hn^{-1/3}) - (\Lambda_0(t_0) + n^{-1/3}x)U_n(t_0 + hn^{-1/3})\right\},
\end{aligned}$$

and hence

$$
\begin{aligned}
&P\left(n^{1/3}(\widehat{\Lambda}_n^{ps}(t_0) - \Lambda_0(t_0)) \le x\right) \\
&\quad = P\left(\widehat{\Lambda}_n^{ps}(t_0) \le \Lambda_0(t_0) + n^{-1/3}x\right) \\
&\quad = P\left(\widehat{S}_n(\Lambda_0(t_0) + n^{-1/3}x) \ge t_0\right) \\
&\quad = P\Big(\arg\min_h \{V_n(t_0 + n^{-1/3}h) \\
&\qquad\qquad - (\Lambda_0(t_0) + n^{-1/3}x)U_n(t_0 + n^{-1/3}h)\} \ge 0\Big).
\end{aligned} \tag{7.22}
$$

Now we rewrite U_n and V_n in terms of $N_{K_i,j}^{(i)}$ and $T_{K_i,j}^{(i)}$ rather than the s_l's as follows:

$$
V_n(t) = \sum_{i=1}^n \sum_{j=1}^{K_i} N_{K_i,j}^{(i)} 1_{\left\{T_{K_i,j}^{(i)} \le t\right\}} = n\mathbb{P}_n\left(\sum_{j=1}^K N_{K,j} 1_{\{T_{K,j} \le t\}}\right)
$$

and

$$
U_n(t) = \sum_{i=1}^n \sum_{j=1}^{K_i} 1_{\left\{T_{K_i,j}^{(i)} \le t\right\}} = n\mathbb{P}_n\left(\sum_{j=1}^K 1_{\{T_{K,j} \le t\}}\right).
$$

Then the "argmin" term inside the probability in the right side of (7.22) can be rewritten as follows:

$$
\begin{aligned}
&\arg\min_h \{V_n(t_0 + n^{-1/3}h) - (\Lambda_0(t_0) + n^{-1/3}x)U_n(t_0 + n^{-1/3}h)\} \\
&\quad = \arg\min_h \Bigg\{\mathbb{P}_n\left[\sum_{j=1}^K N_{K,j} 1_{\{T_{K,j} \le t_0 + n^{-1/3}h\}}\right] \\
&\qquad\qquad - (\Lambda_0(t_0) + n^{-1/3}x)\mathbb{P}_n\left[\sum_{j=1}^K 1_{\{T_{K,j} \le t_0 + n^{-1/3}h\}}\right]\Bigg\} \\
&\quad = \arg\min_h \Bigg\{n^{2/3}(\mathbb{P}_n - P)\Bigg[\sum_{j=1}^K (N_{K,j} - \Lambda_0(t_0)) \\
&\qquad\qquad \times \left(1_{\{T_{K,j} \le t_0 + n^{-1/3}h\}} - 1_{\{T_{K,j} \le t_0\}}\right)\Bigg] \\
&\qquad + n^{2/3}P\left[\sum_{j=1}^K (N_{K,j} - \Lambda_0(t_0))\left(1_{\{T_{K,j} \le t_0 + n^{-1/3}h\}} - 1_{\{T_{K,j} \le t_0\}}\right)\right] \\
&\qquad - n^{1/3}x\mathbb{P}_n\left[\sum_{j=1}^K \left(1_{\{T_{K,j} \le t_0 + n^{-1/3}h\}} - 1_{\{T_{K,j} \le t_0\}}\right)\right]\Bigg\}.
\end{aligned} \tag{7.23}
$$

Now we derive the limit process corresponding to the process on the right hand side of (7.23). First, let

$$f_{n,h}(X) = n^{1/6} \sum_{j=1}^{K} (N_{K,j} - \Lambda_0(t_0)) \left(1_{\{T_{K,j} \le t_0 + n^{-1/3} h\}} - 1_{\{T_{K,j} \le t_0\}} \right),$$

where, as always, $X = (N_K, T_K, K)$. Then

$$n^{2/3}(\mathbb{P}_n - P)\left[\sum_{j=1}^{K} (N_{K,j} - \Lambda_0(t_0)) \left(1_{\{T_{K,j} \le t_0 + n^{-1/3} h\}} - 1_{\{T_{K,j} \le t_0\}} \right) \right] = \mathbb{G}_n f_{n,h}.$$

Fix $B > 0$ and let

$$\mathscr{F}_{B,n} = \{ f_{n,h} : \; h \in [-B, B] \}.$$

The proof now proceeds by showing that $\mathscr{F}_{B,n}$ satisfies the conditions of Theorem 2.11.23 of Van der Vaart and Wellner [(1996), page 221]; for full details of these arguments, see Wellner and Zhang (1998). It then follows from Theorem 2.11.23, Van der Vaart and Wellner [(1996), page 221], that

$$\mathbb{G}_n f_{n,h} \longrightarrow_d \sqrt{\sigma^2(t_0) G'(t_0)} \mathbb{Z}(h) \quad \text{in } l^\infty[-B, B].$$

The limit of the second term in (7.23) is easily obtained by use of conditions E2, E3 and the mean value theorem. For simplicity of notation, we give the proof for the case when $h \ge 0$:

$$\begin{aligned}
& n^{2/3} P\left[\sum_{j=1}^{K} (N_{K,j} - \Lambda_0(t_0)) \left(1_{\{T_{K,j} \le t_0 + h n^{-1/3}\}} - 1_{\{T_{K,j} \le t_0\}} \right) \right] \\
&\quad = n^{2/3} E\left[\sum_{j=1}^{K} (\Lambda_0(T_{K,j}) - \Lambda_0(t_0)) 1_{\{t_0 < T_{K,j} \le t_0 + h n^{-1/3}\}} \right] \\
&\quad = n^{2/3} \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} \int_{t_0}^{t_0 + h n^{-1/3}} (\Lambda_0(s) - \Lambda_0(t_0)) \, dG_{k,j}(s) \\
&\quad = n^{2/3} \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} \int_{t_0}^{t_0 + h n^{-1/3}} \Lambda_0'(\xi(s))(s - t_0) G_{k,j}'(s) \, ds \\
&\quad \to \tfrac{1}{2} \Lambda_0'(t) \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} G_{k,j}'(t_0) h^2 = \tfrac{1}{2} \Lambda_0'(t_0) G'(t_0) h^2.
\end{aligned}$$

Almost identical to the proofs of the first two terms of (7.23), one can show that

$$
\begin{aligned}
&n^{1/3} x \mathbb{P}_n \left[\sum_{j=1}^{K} \left(1_{\{T_{K,j} \le t_0 + hn^{-1/3}\}} - 1_{\{T_{K,j} \le t_0\}} \right) \right] \\
&\quad = n^{-1/3} n^{2/3} x (\mathbb{P}_n - P) \left[\sum_{j=1}^{K} \left(1_{\{T_{K,j} \le t_0 + hn^{-1/3}\}} - 1_{\{T_{K,j} \le t_0\}} \right) \right] \\
&\qquad + n^{1/3} x P \left[\sum_{j=1}^{K} \left(1_{\{T_{K,j} \le t_0 + hn^{-1/3}\}} - 1_{\{T_{K,j} \le t_0\}} \right) \right] \\
&\quad \to_d xh \sum_{j=1}^{K} P(K = k) \sum_{j=1}^{k} G'_{k,j}(t_0) = G'(t_0) x h \qquad \text{in } l^\infty[-B, B].
\end{aligned}
$$

Putting the above results together, we finally arrive at the following limit process:

$$
\begin{aligned}
&V_n \left(t_0 + hn^{-1/3} \right) - \left(\Lambda_0(t_0) + n^{-1/3} x \right) U_n \left(t_0 + hn^{-1/3} \right) \\
&\qquad \longrightarrow_d \sqrt{\sigma^2(t_0) G'(t_0)} \mathbb{Z}(h) + \tfrac{1}{2} \Lambda_0'(t_0) G'(t_0) h^2 - G'(t_0) x h
\end{aligned}
$$

in $l^\infty[-B, B]$. Then the *Argmax Continuous Mapping Theorem* [Theorem 3.2.2, Van der Vaart and Wellner (1996), page 286] implies that

$$
\begin{aligned}
\widehat{h}_n &= \arg\min_h \left\{ V_n(t_0 + hn^{-1/3}) - \left(\Lambda_0(t_0) + n^{-1/3} x \right) U_n \left(t_0 + hn^{-1/3} \right) \right\} \\
&\to_d \arg\min_h \left\{ \sqrt{\sigma^2(t_0) G'(t_0)} \mathbb{Z}(h) + \tfrac{1}{2} \Lambda_0'(t_0) G'(t_0) h^2 - G'(t_0) x h \right\},
\end{aligned}
$$

provided that

$$
\widehat{h}_n = O_p(1). \tag{7.24}
$$

The proof of (7.24) proceeds via Theorems 2.14.2 and 3.2.5 of Van der Vaart and Wellner [(1996), pages 240 and 289]; for the complete proof, see Wellner and Zhang (1998).

Along the same lines as in Groeneboom and Wellner [(1992), pages 99–100], or using Exercise 3.2.8 of Van der Vaart and Wellner [(1996), page 308],

$$
\begin{aligned}
&\arg\min_h \left\{ \sqrt{\sigma^2(t_0) G'(t_0)} \mathbb{Z}(h) + \frac{1}{2} \Lambda_0'(t_0) G'(t_0) h^2 - x G'(t_0) h \right\} \\
&\qquad =_d \left[\frac{4 \sigma^2(t_0)}{\Lambda_0'^2(t_0) G'(t_0)} \right]^{1/3} \arg\min_g \left\{ \mathbb{Z}(g) + g^2 \right\} + \frac{x}{\Lambda_0'(t_0)},
\end{aligned}
$$

and it yields

$$
\begin{aligned}
P & \left(n^{1/3}(\widehat{\Lambda}_n^{ps}(t_0) - \Lambda_0(t_0)) \le x\right) \\
& = P(\widehat{h}_n \ge 0) \\
& \to P\left(\left[\frac{4\sigma^2(t_0)}{(\Lambda_0'(t_0))^2 G'(t_0)}\right]^{1/3} \arg\min_g \{\mathbb{Z}(g) + g^2\} \ge -\frac{x}{\Lambda_0'(t_0)}\right) \\
& = P\left(\left[\frac{\sigma^2(t_0)\Lambda_0'(t_0)}{2G'(t_0)}\right]^{1/3} 2 \cdot \arg\max_g \{\mathbb{Z}(g) - g^2\} \le x\right),
\end{aligned}
$$

and the proof is complete. □

For a sketch of the proof of Theorem 4.4, see Wellner and Zhang (1998).

APPENDIX

This section provides a statement of one of the empirical process results which we use in our proofs. It is a one-sided Glivenko-Cantelli theorem which is Le Cam's [(1953), page 301] recasting of a result of Wald (1949); see also Ferguson [(1996), pages 107–111]. Huber (1967) and Dudley (1998) give further results in this direction.

THEOREM A.1. [(One-sided Glivenko-Cantelli theorem)] *Suppose that* $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ *is a class of measurable functions defined on a probability space* $(\mathcal{X}, \mathcal{A}, P)$, *where* Θ *is compact, and* $f(x, \theta)$ *is upper semicontinuous in* θ *for* P *almost every* x. *Moreover, suppose that there exists a function* $F(x)$ *such that* $EF(X) < \infty$ *and* $f(x, \theta) \le F(x)$ *for all* $x \in \mathcal{X}$, $\theta \in \Theta$, *and for all* θ *and all sufficiently small* $\rho > 0$,

$$
\sup_{\{\theta' : d(\theta', \theta) < \rho\}} f(x, \theta')
$$

is measurable in x. *Then, if* $X_1, \ldots, X_n$ *are i.i.d.* P *with values in* $\mathcal{X}$, *and* $\mathbb{P}_n$ *is the empirical measure of the* X_i*'s.*

$$
\limsup_{n\to\infty} \sup_{\theta\in\Theta} \mathbb{P}_n f(\cdot, \theta) \le \sup_{\theta\in\Theta} Pf(\cdot, \theta) \quad a.s.
$$

and

$$
\limsup_{n\to\infty} \sup_{\theta\in\Theta} (\mathbb{P}_n - P) f(\cdot, \theta) \le 0 \quad a.s.
$$

Moreover, $\mu(\theta) = Pf(\cdot, \theta)$ *is upper semicontinuous:*

$$
\limsup_{\theta'\to\theta} \mu(\theta') \le \mu(\theta).
$$

Acknowledgments. We owe thanks to Piet Groeneboom for discussions leading to improvements and corrections of our consistency proofs, to Danyu Lin for pointing out the data set analyzed in Section 5, to Norman Breslow for discussions and suggestions about other possible applications, and to Anton Schick for pointing out errors in our first consistency proofs and for several helpful conversations. We also thank the two referees for providing several additional corrections and useful suggestions.

REFERENCES

ANDREWS, D. F. and HERZBERG, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer, New York.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

BRUNK, H. D. (1969). Estimation of isotonic regression. In *Proceedings of the International Symposium on Nonparametric Techniques in Statistical Inference* 177–195. Cambridge Univ. Press.

BYAR, D. P., BLACKARD, C. and THE VACURG. (1977). Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology* **10** 556–561.

BYAR, D. P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology* (M. Pavone-Macaluso, P. H. Smith and F. Edsmyr, eds.) 363–370. Plenum, New York

DIAMOND, I. D. and MCDONALD, J. W. (1992). Analysis of current status data. In *Demographic Applications of Event History Data Analysis* (J. Trussell, R. Hankinson and J. Tilton, eds.) 231–252. Clarendon Press, Oxford.

DIAMOND, I. D., MCDONALD, J. W. and SHAH, I. H. (1986). Proportional hazard models for current status data: Applications to the study of differentials in age at weaning in Pakistan. *Demography* **28** 607–620.

DUDLEY, R. M. (1998). Consistency of M-estimators and one-sided bracketing. In *Proceedings of the First International Conference on High-Dimensional Probability* (E. Eberlein, M. Hahn and M. Talagrand, eds.) 33–58. Birkhäuser, Basel.

FERGUSON, T. (1996). *A Course in Large Sample Theory.* Chapman and Hall, London.

FINKELSTEIN, D. M. and WOLF, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41** 933–945.

FINKELSTEIN, D. M. (1986). A proportional hazard model for interval-censored failure time data. *Biometrics* **42** 584–854.

GAVER, D. P. and O'MUIRCHEARTAIGH, I. G. (1987). Robust empirical Bayes analysis of event rates. *Technometrics* **29** 1–15.

GROENEBOOM, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. Technical Report 378, Dept. Statistics, Stanford Univ.

GROENEBOOM, P. (1996). Inverse problems in statistics. *Proceedings of the St. Flour Summer School in Probability. Lecture Notes in Math.* **1648** 67–164. Springer, Berlin.

GROENEBOOM, P. and WELLNER, J. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, Basel.

GROENEBOOM, P. and WELLNER, J. (2000). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **9**. To appear.

HUANG, J. (1996). Efficient estimation for the Cox model with interval censoring. *Ann. Statist.* **24** 540–568.

HUANG, J. and WELLNER, J. A. (1995). Efficient estimation for the Cox model with case 2 interval censoring. Unpublished manuscript.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

JONGBLOED, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.* **7** 310–321.

KALBFLEISCH, J. D. and LAWLESS, J. F. (1985). The analysis of panel count data under a Markov assumption. *J. Amer. Statist. Assoc.* **80** 863–871.

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related estimates. *Univ. Calif. Publ. Statist.* **1** 277–330.

SCHICK, A. and YU, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.* **27** 45–55.

SUN, J. and KALBFLEISCH, J. D. (1993). The analysis of current status data on point processes. *J. Amer. Statist. Assoc.* **88** 1449–1454.

SUN, J. and KALBFLEISCH, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5** 279–290.

THALL, P. F. and LACHIN, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *J. Amer. Statist. Assoc.* **83** 339–347.

THALL, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44** 197–209.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York.

WALD, A. (1949). Note on the consistency of the maximum likelihood estimator. *Ann. Math. Statist.* **20** 595–601.

WEI, L. J., LIN, D. Y. and WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84** 1065–1073.

WELLNER, J. A. (1995). Interval censoring, case 2: Alternative hypotheses. In *Analysis of Censored Data* 272–291. IMS, Hayward, CA.

WELLNER, J. A. and ZHAN, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J. Amer. Statist. Assoc.* **92** 945–959

WELLNER, J. A. and ZHANG, Y. (1998). Two estimators of the mean of a counting process with panel count data. Technical Report 341, Dept. Statistics, Univ. Washington. `www.stat.washington.edu/www/research/reports/`.

ZHANG, Y. (1998). Estimation for counting processes based on incomplete data. Ph. D. dissertation, Univ. Washington.

ZHANG, Y. (1999). A semiparametric pseudolikelihood estimation for panel count data. Technical report, Univ. Central Florida.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
BOX 354322
SEATTLE, WASHINGTON 98195–4322
E-MAIL: jaw@stat.washington.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF CENTRAL FLORIDA
BOX 162370
ORLANDO, FLORIDA 32816–2370
E-MAIL: zhang@pegasus.cc.ucf.edu