Statistica Sinica **14**(2004), 269-282

ESTIMATION WITH UNIVARIATE "MIXED CASE" INTERVAL CENSORED DATA

Shuguang Song

The Boeing Company

Abstract: In this paper, we study the Nonparametric Maximum Likelihood Estimator (NPMLE) of univariate "Mixed Case" interval-censored data in which the number of observation times, and the observation times themselves are random variables. We provide a characterization of the NPMLE, then use the ICM algorithm to compute the NPMLE. We also study the asymptotic properties of the NPMLE: consistency, global rates of convergence with and without a separation condition, and an asymptotic minimax lower bound.

Key words and phrases: Consistency, convex minorant algorithm, empirical processes, interval censored data, maximum likelihood, rate of convergence.

1. Introduction and Formulation of the Problem

Univariate interval censoring problem arises when a random variable, such as failure time or onset of disease, cannot be directly observed, but is only known to be in an interval determined by several observation times. Interval censored data and the corresponding application of statistical methods can be found in animal carcinogenicity, epidemiology, HIV and AIDS studies, see Finkelstein and Wolfe (1985), Finkelstein (1986), Self and Grossman (1986), Becker and Melbye (1991) and Aragón and Eberly (1992). There are three types of univariate interval censorship models: "Case 1" interval censoring or current status data; "Case 2" and "Case k" interval censoring; "Mixed Case" interval censoring.

Suppose that Y is a random variable with distribution function $F \in \mathcal{F} \equiv \{$ all distribution functions on $R_+\}$. Unfortunately, we are unable to directly observe Y itself, but we can observe a random vector (δ, T), where T is the observation time independent of Y, $\delta = 1_{[Y \le T]}$. So the only knowledge about the event of Y is whether it occurred before T or after T. This model is called the "Case 1" interval censoring model in Groeneboom and Wellner (1992). In "Case 2" interval censoring, Y falls into one of three time intervals formed by two observation times T_1 and T_2. The data observed are: $(\delta_1, \delta_2, T_1, T_2) = (1_{[Y \le T_1]}, 1_{[T_1 < Y \le T_2]}, T_1, T_2)$. The nonparametric maximum likelihood estimation (NPMLE) of "Case 1" and "Case 2" interval censored data are summarized in Groeneboom and Wellner

(1992), Groeneboom (1996) and Huang and Wellner (1997). Wellner (1995) studied the NPMLE of the "Case k" interval censoring in which each subject has exactly k examination times.

To compute the NPMLE, Turnbull (1976) derived self-consistency equations for a very general censoring scheme and proposed solving the equations by the EM algorithm. Groeneboom (1991) developed an iterative convex minorant algorithm (ICM). The ICM algorithm is considerably faster than the EM algorithm especially when the sample size is large; see Groeneboom and Wellner (1992). Aragón and Eberly (1992) and Jongbloed (1995, 1998) proposed modifications of the ICM algorithm, and Jongbloed (1995, 1998) showed that his modified algorithm always converges.

Very often in clinical trials, each patient has several follow-ups and the number of follow-ups differs from patient to patient. This motivates the study of the following model. Let $\underline{T} = T_{k,j}$, $j = 0, 1, \ldots, k, k+1$, $k = 1, \ldots$, be a triangular array of "potential observation times" with $T_{k,0} \equiv 0$ and $T_{k,k+1} \equiv +\infty$, and let K, the number of observation times, be an integer-valued random variable such that Y and $(K, \underline{T})$ are independent. What we can observe is a vector $X = (\Delta_K, T_K, K)$, with possible value $x = (\delta_k, t_k, k)$, where T_k is the kth row of the triangular array $\underline{T}$, $\Delta_k = (\Delta_{k,1}, \ldots, \Delta_{k,k})$ with $\Delta_{k,j} = 1_{(T_{k,j-1}, T_{k,j}]}(Y), j = 1, \ldots, k+1$. Suppose we observe n i.i.d. copies of X: $X_1, \ldots, X_n$, where $X_i = (\Delta^{(i)}_{K^{(i)}}, T^{(i)}_{K^{(i)}}, K^{(i)})$, $i = 1, \ldots, n$. Here $(Y^{(i)}, \underline{T}^{(i)}, K_i), i = 1, 2, \ldots$, are the underlying i.i.d. copies of $(Y, \underline{T}, K)$. Schick and Yu (2000) referred to this model as the "Mixed Case" interval censoring model. They proved strong consistency in the $L_1(\mu)$-topology of the NPMLE for a measure μ which is derived from the distribution of observation times.

Van der Vaart and Wellner (2000) gave a different formulation of "Mixed Case" interval censoring. They noted that conditional on K and T_K, the vector Δ_K has a multinomial distribution: $(\Delta_K | K, T_K) \sim \text{Multinomial}_{K+1}(1, \Delta F_K)$, where $\Delta \underline{F}_K \equiv (F(T_{K,1}), F(T_{K,2}) - F(T_{K,1}), \ldots, 1 - F(T_{K,K}))$. Suppose for the moment that the distribution G_k of $(T_K | K = k)$ has density g_k, and $p_k \equiv P(K = k)$. Then the density of X is given by

$$p_F(x) \equiv p_F(\delta, t_k, k) = \sum_{j=1}^{k+1} \delta_{k,j}(F(t_{k,j}) - F(t_{k,j-1})) \tag{1.1}$$

with respect to the dominating measure ν which is determined by the joint distribution of $(K, \underline{T})$. Thus the normalized log-likelihood function for F of $X_1, \ldots, X_n$ is given by

$$\frac{1}{n} l_n(F|\underline{X}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i+1} \Delta_{K_i,j} \log(F(T^{(i)}_{K_i,j}) - F(T^{(i)}_{K_i,j-1})) = \mathbb{P}_n m_F, \tag{1.2}$$

where $m_F(\underline{X}) = \sum_{j=1}^{K_i+1} \Delta_{K_i,j} \log(F(T^{(i)}_{K_i,j}) - F(T^{(i)}_{K_i,j-1}))$
$\equiv \sum_{j=1}^{K+1} \Delta_{K_i,j} \log(\Delta F(T^{(i)}_{K_i,j}))$.

Van der Vaart and Wellner (2000) proved the consistency of the NPMLE of the “Mixed Case” interval censoring in Hellinger distance, and also recovered the results of Schick and Yu (2000) by using preservation theorems for Glivenko-Cantelli classes. Here, we use the above formulation of the “Mixed Case” interval censoring problem. In Section 2, we give a characterization of the NPMLE. We then use the ICM algorithm to compute the NPMLE. In Section 3, we state the main asymptotic properties of the NPMLE: consistency, global rates of convergence with and without a separation condition, and an asymptotic minimax lower bound. The results in Section 3 are proved in Section 5 by using empirical process theory. There are other two estimators for the univariate “mixed case” interval censored data that are based on the non-homogeneous Poisson process model: nonparametric maximum likelihood estimator and nonparametric maximum pseudo-likelihood estimator (NPMPLE), see Wellner and Zhang (2000). We denote these two estimators as NPMLE^{WZ} and NPMPLE^{WZ}. In Section 4, we present simulation studies to compare the asymptotic relative efficiency of the above three estimators.

2. Characterization and Computation of the NPMLE

Let $t_1 < \cdots < t_m$ denote the ordered distinct observation time points in the set of all observation time points: $\{T_{K_i,j}, j = 1, \ldots, K_i,\ i = 1, \ldots, n\}$. Define the rank function R: $\{T_{K_i,j},\ j = 1, \ldots, K_i,\ i = 1, \ldots, n\} \to \{1, \ldots, m\}$ such that $R(T_{K_i,j}) = s$, if $T_{K_i,j} = t_s$ for $s = 1, \ldots, m$. Let $\Omega = \{(F(t_1), \ldots, F(t_m)) : 0 \leq F(t_1) \leq \cdots \leq F(t_m)$, for all $F \in \mathcal{F}\}$, and $\underline{F} = (F(t_1), \ldots, F(t_m))$. Then the normalized log-likelihood (1.2) can be rewritten as:

$$\frac{1}{n} l_n(F|\underline{X}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i+1} \Delta_{K_i,j}\ \log(F(t_{R(T^{(i)}_{K_i,j})}) - F(t_{R(T^{(i)}_{K_i,j-1})})),$$

where $\Delta_{K_i,j} = 1_{(t_{R(T_{K_i,j-1})},\, t_{R(T_{K_i,j})}]}(Y_i)$.

Note that Ω is a convex set, $l_n(\underline{F}|\underline{X})$ is a concave function due to the linear combination of the concave function “log”. Let $l_k(\underline{F}|\underline{X}) \equiv \partial l_n(\underline{F}|\underline{X})/\partial F(t_k) = \sum_{i=1}^{n} l_{i,k}(\underline{F}|\underline{X})$, where

$$l_{i,k}(\underline{F}|\underline{X}) = \sum_{j=1}^{K_i} \left[\frac{\Delta_{K_i,j}}{F(t_{R(T_{K_i,j})}) - F(t_{R(T_{K_i,j-1})})} - \frac{\Delta_{K_i,j+1}}{F(t_{R(T_{K_i,j+1})}) - F(t_{R(T_{K_i,j})})} \right] \delta_{K_i,j}(k) \tag{2.1}$$

and $\delta_{K_i,j}(k) = 1_{[R(T_{K_i,j})=k]}$. The following characterization of the NPMLE $\hat{F}_n$ follows from Fenchel duality theorem.

Theorem 2.1. *The unique NPMLE $\underline{\hat{F}}_n$ maximizes $l_n(F|\underline{X})$ over all $F \in \mathcal{F}$ if and only if*

$$\sum_{k=l}^{m} \sum_{i=1}^{n} l_{i,k}(\underline{\hat{F}}_n|\underline{X}) \leq 0, \qquad \textit{for } l = 1, \ldots, m, \tag{2.2}$$

$$\sum_{k=1}^{m} \left[\sum_{i=1}^{n} l_{i,k}(\underline{\hat{F}}_n|\underline{X}) \right] \underline{\hat{F}}_n(t_k) = 0. \tag{2.3}$$

Denote $l_{kk}(\underline{F}|\underline{X}) \equiv \partial^2 l_n(\underline{F}|\underline{X})/\partial F^2(t_k) = \sum_{i=1}^n l_{i,kk}(\underline{F}|\underline{X})$, where

$$l_{i,kk}(\underline{F}|\underline{X}) = -\sum_{j=1}^{K_i} \left\{ \frac{\Delta_{K_i,j}}{[F(t_{R(T_{K_i,j})}) - F(t_{R(T_{K_i,j-1})})]^2} + \frac{\Delta_{K_i,j+1}}{[F(t_{R(T_{K_i,j+1})}) - F(t_{R(T_{K_i,j})})]^2} \right\} \delta_{K_i,j}(k).$$

Then define the $G(\underline{F}, \cdot)$ and $V(\underline{F}, \cdot)$ processes by

$$\begin{aligned}
G(\underline{F}, 0) &= 0, \qquad V(\underline{F}, 0) = 0, \\
G(\underline{F}, p) &= \sum_{k=1}^{p} (-l_{kk}(\underline{F}|\underline{X})) = -\sum_{k=1}^{p} \sum_{i=1}^{n} l_{i,kk}(\underline{F}|\underline{X}), \\
V(\underline{F}, p) &= \sum_{k=1}^{p} (l_k(\underline{F}|\underline{X}) + F(t_k)(-l_{kk}(\underline{F}|\underline{X}))) \\
&= \sum_{k=1}^{p} \sum_{i=1}^{n} [\, l_{i,k}(\underline{F}|\underline{X}) - F(t_k)\, l_{i,kk}(\underline{F}|\underline{X})] \,,
\end{aligned}$$

where $p = 1, \ldots, m$. Since

$$\begin{aligned}
G(\underline{F}, p|\underline{X}) &= \sum_{k=1}^{p} -\frac{\partial^2 l_n(\underline{F}|\underline{X})}{\partial F^2(t_k)}, \\
V(\underline{F}, p|\underline{X}) &= \sum_{k=1}^{p} \left[F(t_k)(-\frac{\partial^2 l_n(\underline{F}|\underline{X})}{\partial F^2(t_k)}) + \frac{\partial l_n(\underline{F}|\underline{X})}{\partial F(t_k)} \right],
\end{aligned}$$

by using Theorem 4.3 of Wellner and Zhan (1997), we can also characterize the NPMLE $\hat{F}_n$ as the slope of the convex minorant of a self-induced cumulative diagram.

Theorem 2.2. *$\hat{F}_n$ is the NPMLE of F_0 if and only if $\hat{F}_n$ is the left derivative of the convex minorant of the cumulative sum diagram consisting the points $P_p = (G(\underline{\hat{F}}_n, p), V(\underline{\hat{F}}_n, p))$, where $P_0 = (0, 0)$, $p = 1, \ldots, m$.*

The above left derivative of the convex minorant of the cumulative sum diagram can be calculated by the "max-min" formula

$$\hat{F}_n^{l+1}(t_p) = \max_{j \le p} \min_{i \ge p} \frac{V(\underline{\hat{F}}_n^l, i) - V(\underline{\hat{F}}_n^l, j)}{G(\underline{\hat{F}}_n^l, i) - G(\underline{\hat{F}}_n^l, j)},$$

where $p = 1, \ldots, m$, and l is the index of iteration. Only one $\Delta_{K_i,j}$, for $j = 1, \ldots, K_i$, is equal to one for each subject, so Theorem 2.1 and Theorem 2.2 can be reduced to Proposition 1.3 and Proposition 1.4 of Groeneboom and Wellner (1992). Thus, the computation of the NPMLE of "Mixed Case" interval censoring can be reduced to the "Case 2" interval censoring as noted by Huang and Wellner (1997) and Van der Vaart and Wellner (2000).

The ICM algorithm has been applied in many problems: see Groeneboom and Wellner (1992), Jongbloed (1995, 1998), Wellner and Zhan (1997), Wellner and Zhang (2000). To prove global convergence, Jongbloed (1995) designed a modified iterative convex minorant algorithm (MICM) by inserting a binary line search procedure. In this case, the NPMLE of the "Mixed Case" interval censoring can be calculated by the MICM algorithm as for the "Case 2" interval censoring.

3. Asymptotic Properties of the NPMLE: Results

In this section, we study the asymptotic properties of the NPMLE: consistency, global rate of convergence in Hellinger distance, and a local asymptotic minimax lower bound. The following regularity conditions are needed.

A. $EK < \infty$.

B. Separation condition: there exists a $\delta > 0$ such that $P(T_{K,j} - T_{K,j-1} \ge \delta) = 1$, for every $j = 1, \ldots, K, K = 1, 2, \ldots$.

C. For $0 \equiv t_{k,0} < t_{k,1} < \cdots < t_{k,k} < t_{k,k+1} \equiv \infty$, there exists a $\eta > 0$ such that the underlying distribution function F_0 satisfies $F_0(t_{k,j}) - F_0(t_{k,j-1}) > \eta$, for all $j = 1, \ldots, k+1$.

D. If $q_{k,j,j-1}$ is the density of the joint distribution $Q(t_{k,j}, t_{k,j-1}|K = k)$, and f_0 is the density of F_0 with respect to Lebesgue measure on R, there exist pointwise constants c_0 and c_1 such that $q_{k,j,j-1}(s,t) \le c_0$ for all t, s and $j = 1, \ldots, k+1$, $k = 1, 2, \ldots$, and $1/c_1 \le f_0(y) \le c_1$ for $y \in R$.

3.1. Consistency

Van der Vaart and Wellner (2000) have proved the consistency of the NPMLE $\hat{F}_n$ by using preservation theorems for Glivenko-Cantelli classes. Here we give another approach to the proof. Then, following this approach, we study the rate of convergence of the NPMLE.

In the probability space $(\Omega, \mathcal{A}, P_F)$, where Ω is $\cup_{k=1}^{\infty}\{0,1\}^{k+1} \times R_+^k \times \{k\}$; $\mathcal{A}$ is a σ-field generated by $\mathcal{A}_{k1} \times \mathcal{B}_k \times \mathcal{A}_{k3}$, $\mathcal{A}_{k1}$ is the class of all subsets of $\{e_j^k \equiv (0,\ldots,\underbrace{1}_{j\text{th}},0,\ldots,0) : j = 1,\ldots,k+1\}$, $\mathcal{B}_k$ is a σ-field generated by the set of all k-dimensional rectangles on R_+^k, $\mathcal{A}_{k3} = \sigma\{\emptyset, \{k\}\}$; P_F is a probability measure with density defined in (1.1) and dominating measure ν defined as

$$\nu(\{e_j^k\} \times B_k \times \{k\}) = \text{counting measure on } \{e_j^k\} \times P(K=k) \times P(\underline{T} \in B_k | K=k).$$

Let $Q(t_{k,1},\ldots,t_{k,k}|K=k)$ be the conditional distribution function of $\underline{T}_k$. Let $Q_{k,j}$ denote the marginal distribution of $T_{k,j}$, for $j = 1,\ldots,k$, $k = 1,2\ldots$, conditional on $K = k$. Then the Hellinger distance is

$$\begin{aligned} &h^2(p_F, p_{F_0}) \\ &= \frac{1}{2}\sum_{k=1}^{\infty} P(K=k) \sum_{j=1}^{k+1} \int \Big\{[F(t_{k,j}) - F(t_{k,j-1})]^{1/2} - [F_0(t_{k,j}) - F_0(t_{k,j-1})]^{1/2}\Big\}^2 dQ(\underline{t}_k). \end{aligned}$$

Note that $\int p_F d\nu = 1$. For fixed $F_0 \in \mathcal{F}$ and p_F as defined in (1.1), let $\mathcal{P} = \{p_F : F \in \mathcal{F}\}$, $m_F = (p_F - p_{F_0})/(p_F + p_{F_0}) = 2p_F/(p_F + p_{F_0}) - 1$, $\mathcal{M} \equiv \{m_F : F \in \mathcal{F}\}$, $\mathcal{M}_\delta \equiv \{m_F - m_{F_0} : h(p_F, p_{F_0}) < \delta, F \in \mathcal{F}\}$. Then $\mathcal{P}$ is convex. As shown by Van der Vaart and Wellner (2000, p.123), the Hellinger distance $h(p_{\hat{F}_n}, p_{F_0})$ is less than or equal to $\|\mathbb{P}_n - P\|_{\mathcal{M}}$ for a density p_F with respect to a dominating measure ν. To show that $h(p_{\hat{F}_n}, p_{F_0}) \to_{a.s} 0$, we need only prove that the class $\mathcal{M}$ is a Glivenko-Cantelli class by showing that its bracketing numbers are finite for each $\epsilon > 0$, see Theorem 2.4.1 in Van der Vaart and Wellner (1996). It follows from Lemma 3.1 below that the bracketing number $N_{[\,]}(\epsilon, \mathcal{M}, L_r(P))$ is bounded by $N_{[\,]}(\epsilon, \mathcal{P}, L_r(Q_\sigma))$, where $dQ_\sigma = 1_{[p_{F_0} > \sigma]} p_{F_0}^{1-r} d\nu$ for $\sigma > 0$.

Lemma 3.1. *For any integer* $r \geq 1$, *let* $\sigma_0(\epsilon) \equiv \sup\{\sigma \geq 0 : \int p_{F_0} 1_{[p_{F_0} \leq \sigma]} d\nu \leq \epsilon^r\}$, $\mathcal{G}_\sigma \equiv \{2p_F 1_{[p_{F_0} > \sigma]}/(p_F + p_{F_0}) : p_F \in \mathcal{P}\}$. *Then* $N_{[\,]}((2^r+1)^{1/r}\,\epsilon,\, \mathcal{M},\, L_r(P)) \leq N_{[\,]}(\epsilon,\, \mathcal{G}_{\sigma_0(\epsilon)},\, L_r(P))$ *and* $N_{[\,]}(2\epsilon,\, \mathcal{G}_{\sigma_0(\epsilon)},\, L_r(P)) \leq N_{[\,]}(\epsilon,\, \mathcal{P},\, L_r(Q_\sigma))$, *where* $dQ_\sigma = 1_{[p_{F_0} > \sigma]} p_{F_0}^{1-r} d\nu$.

Using Theorem 2.7.5 in Van der Vaart and Wellner (1996), it is not hard to show that if $EK < \infty$, then $\log N_{[\,]}(\epsilon, \mathcal{P}, L_1(P)) = O(1/\epsilon)$ and consequently $\log N_{[\,]}(\epsilon, \mathcal{M}, L_1(P)) = O(1/\epsilon)$ for univariate "mixed case" interval censored data. Thus we conclude that the NPMLE $\hat{F}_n$ of the univariate "mixed case" interval censored data is consistent in Hellinger distance when $EK < \infty$. Consistency of the NPMLE in L_1-norm can also be derived; see Van der Vaart and Wellner (2000). Under additional hypotheses, Schick and Yu (2000) proved that $\hat{F}_n$ converges pointwise and even uniformly.

3.2. A local asymptotic minimax lower bound

Let $\mathcal{G}$ be a set of probability densities on a measurable space $(\Omega, \mathcal{A})$ with respect to a σ-finite dominating measure μ, T be a real-valued functional on $\mathcal{G}$, and let T_n, $n \geq 1$, be a sequence of estimators of Tg based on samples of size n from a density g known to be contained in $\mathcal{G}$. The minimax risk of the estimator T_n, which measures the difficulty of the problem of estimating Tg, is $\inf_{T_n} \max_{g \in \mathcal{G}} E_{n,g}\, l(|T_n - Tg|)$, where l is an increasing convex loss function on $[0, +\infty)$ with $l(0) = 0$. As $n \to +\infty$, the asymptotic lower bound for the minimax risk follows from an inequality in Lemma 4.1 of Groeneboom (1996):

$$\inf_{T_n} \max_{g \in \mathcal{G}} \{E_{n,g_1} l(|T_n - Tg_1|), E_{n,g_2} l(|T_n - Tg_2|)\} \geq l(\frac{1}{4}|Tg_1 - Tg_2|\{1 - h^2(g_1, g_2)\}^{2n}), \tag{3.1}$$

where $h(g_1, g_2)$ is the Hellinger distance, g_1, $g_2 \in \mathcal{G}$. If we consider a subset of $\mathcal{G}$, a perturbed sequence of one fixed $g \in \mathcal{G}$, then the asymptotic lower bound for the corresponding minimax risk, called local minimax risk, gives the best possible local convergence rate.

The local rates of convergence for "Case 1" and "Case 2" of univariate interval censored data have been well studied; see Groeneboom (1996). Here we follow the same approach to extend the minimax results to univariate "mixed case" interval-censored data. Consider the following perturbations F_n of the underlying distribution function F_0:

$$F_n(t) = \begin{cases} F_0(t) + \theta\, f_0(t_0)\, (t - t_0 + cn^{-1/3}), & \text{if } t \in [t_0 - cn^{-1/3}, t_0), \\ F_0(t) + \theta\, f_0(t_0)\, (t_0 + cn^{-1/3} - t), & \text{if } t \in [t_0, t_0 + cn^{-1/3}], \\ F_0(t), & \text{otherwise}, \end{cases} \tag{3.2}$$

where $0 < \theta < 1$ is a constant, c is a constant to be determined by F_0 and Q. Then the density function f_n of F_n is positive on $[t_0, t_0 + cn^{-1/3}]$ for $0 < \theta < 1$ and large n.

For "Mixed Case" univariate interval-censored data, an asymptotic minimax result is given by the following theorem.

Theorem 3.1. *Let $F_0(t)$ be a function with density $f_0(t)$, let F_n be the sequence of perturbations of F_0 given by* (3.2), *let $q_{k0} \equiv 0$, $q_{k,k+1} \equiv 0$, $q_{k,j}(t)$ be the density of the distribution function $Q(t_{k,j}|K = k)$, and let $q_{k,j,j-1}$ be the joint density of $Q(t_{k,j}, t_{k,j-1}|K = k)$, for $j = 1, \ldots, k$, $k = 1, \ldots$. Suppose that condition B holds and $\sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k+1} (q_{k,j}(t_0) + q_{k,j-1}(t_0)) < \infty$. Then*

$$\liminf_{n \to \infty} n^{1/3} \max \{E_{n,p_{F_n}}\ |T_n - F_n(t_0)|,\ E_{n,p_{F_0}}\ |T_n - F_0(t_0)|\} \geq c_0 \Big\{ \frac{f_0(t_0)}{a(t_0)} \Big\}^{1/3},$$

where $c_0 \equiv (1/4)(2\theta/3)^{-1/3}\, e^{-1/3}$ *is a constant depending on* θ, *and*

$$a(t_0) \equiv \sum_{k=1}^{\infty} P(K=k) \left\{ \frac{q_{k,1}(t_0)}{F_0(t_0)} + \sum_{j=2}^{k} \int_{0 \le t_{k,j-1} < t_0} \frac{q_{k,j-1,j}(t_{k,j-1}, t_0)}{F_0(t_0) - F_0(t_{k,j-1})}\, dt_{k,j-1} \right.$$
$$\left. + \sum_{j=1}^{k-1} \int_{t_0 \le t_{k,j+1} < \infty} \frac{q_{k,j,j+1}(t_0, t_{k,j+1})}{F_0(t_{k,j+1}) - F_0(t_0)}\, dt_{k,j+1} + \frac{q_{k,k}(t_0)}{1 - F_0(t_0)} \right\}.$$

Note that in the cases of $P(K=1)=1$ and $P(K=2)=1$, $a(t_0)$ reduces to the known lower bound for univariate current status data and univariate interval censored data, case 2; see pp.137-138 in Groeneboom (1996).

3.3. Global rate of convergence in Hellinger distance

We apply empirical processes theory to study the rate of convergence of the NPMLE with univariate "mixed case" interval censored data. Consider a deterministic function $\mathbb{M} : \Theta \mapsto R$, and stochastic processes $\mathbb{M}_n$ indexed by a semimetric space Θ. Let θ_0 be a point of maximum of the map $\theta \mapsto \mathbb{M}(\theta)$. Let $\hat{\theta}_n$ be estimators that (nearly) maximize the maps $\theta \mapsto \mathbb{M}_n(\theta)$. An upper bound for the rate of convergence of $\hat{\theta}_n$ can be obtained from the continuity modulus of the difference $\mathbb{M}_n - \mathbb{M}$, see Theorem 3.2.5 in Van der Vaart and Wellner (1996). In the case of i.i.d. data, $\mathbb{M}_n(\theta) = \mathbb{P}_n(\theta)$ and $\mathbb{M}(\theta) = P(\theta)$, the centered process $\sqrt{n}(\mathbb{M}_n - \mathbb{M}) = \mathbb{G}_n(m_\theta)$ is an empirical process at m_θ. Define $\mathbb{M}_\delta = \{m_\theta - m_{\theta_0} : d(\theta, \theta_0) < \delta\}$. Corollary 3.2.6 in Van der Vaart and Wellner (1996) is used to derive the rates of convergence in this section.

The key step to apply Corollary 3.2.6 in Van der Vaart and Wellner (1996) is to derive sharp bounds on the modulus of continuity of the empirical process. Define the bracketing integral $J_{[\,]}(\eta, \mathbb{M}_\delta, L_2(P)) = \int_0^\eta \{1 + \log N_{[\,]}(\epsilon \|M_\delta\|_{P,2}, \mathbb{M}_\delta, L_2(P))\}^{1/2} d\epsilon$, where M_δ is an envelope function of the class $\mathbb{M}_\delta$. Then the bounds can be obtained using Lemma 3.4.2 in Van der Vaart and Wellner (1996), $E_P^* \|\mathbb{G}_n\|_{\mathbb{M}_\delta} \lesssim J_{[\,]}(\delta, \mathbb{M}_\delta, L_2(P))(1 + M\, J_{[\,]}(\delta, \mathbb{M}_\delta, L_2(P))/(\delta^2 \sqrt{n}))$, where each element in $\mathbb{M}_\delta$ is uniformly bounded by M. The rates are obtained with and without a separation condition. We first obtain the entropy with bracketing for the class $\mathcal{M}$ with density defined in (1.1).

Lemma 3.2. *Under conditions A, B and C,* $\log N_{[\,]}(\epsilon, \mathcal{M}, L_2(P)) = O(\epsilon^{-1})$.

Under the separation condition, the following theorem gives an $n^{1/3}$ global rate of convergence in Hellinger distance for the NPMLE of the univariate "mixed case" interval censored data.

Theorem 3.2. *Under conditions A, B and C,* $h(p_{\hat{F}_n}, p_{F_0}) = O_p(n^{-1/3})$.

Without the separation condition B, the rate result is weaker.

Theorem 3.3. *Under conditions A and D,* $h(p_{\hat{F}_n}, p_{F_0}) = O_P(n^{-1/3} \log^{1/6} n)$.

4. Efficiency of the NPMLE

Let $\hat{F}_n$ denote the NPMLE in the current model. Let $\hat{\Lambda}_n$ and $\hat{\Lambda}_n^{ps}$ denote the NPMLE^{WZ} and NPMPLE^{WZ} in the non-homogeneous Poisson process model, see Wellner and Zhang (2000). Note that the one-jump counting process $N(t) = 1_{[Y \le t]}$ has mean function $\Lambda(t) = E\{N(t)\} = P(Y \le t) = F(t)$; see Example 4.1, p.792, Wellner and Zhang (2000). The asymptotic distribution of the "toy estimator" for "mixed case" interval censored data, obtained by taking one step in the iterative convex minorant algorithm starting with the real underlying function F_0, has been established, see Theorem 3.5 in Song (2001) or Song (2002). It is conjectured that this "toy estimator" and the NPMLE $\hat{F}_n$ have the same asymptotic distribution. Considering the pointwise asymptotic distribution of the $\hat{\Lambda}_n$ and $\hat{\Lambda}_n^{ps}$, see Theorems 4.3 and 4.4 in Wellner and Zhang (2000), we suppose that the estimators $\hat{F}_n$, $\hat{\Lambda}_n$ and $\hat{\Lambda}_n^{ps}$ have the same asymptotic distribution up to positive constants L_1, L_2 and L_3: $n^{1/3}(\hat{F}_n(t_0) - F_0(t_0)) \to_d L_1^{2/3}\mathbb{Y}$, $n^{1/3}(\hat{\Lambda}_n(t_0) - \Lambda_0(t_0)) \to_d L_2^{2/3}\mathbb{Y}$, and $n^{1/3}(\Lambda\hat{^{ps}}_n(t_0) - \Lambda_0(t_0)) \to_d L_3^{2/3}\mathbb{Y}$, where $\mathbb{Y}$ is a known random variable. Note that $\mathrm{Var}(\hat{F}_n) \cong n^{-2/3}L_1^{2/3}$ $\mathrm{Var}(\mathbb{Y})$, $\mathrm{Var}(\hat{\Lambda}_n) \cong n^{-2/3}L_2^{2/3}$ $\mathrm{Var}(\mathbb{Y})$, and $\mathrm{Var}(\hat{\Lambda}_n^{ps}) \cong n^{-2/3}L_3^{2/3}$ $\mathrm{Var}(\mathbb{Y})$. To study the asymptotic relative efficiency of two estimators, we ask the two estimators to have the same variance, asymptotically. Thus, we can plot the 3/2 power of the ratio of sample variances of the two estimators to obtain an approximation to the asymptotic relative efficiency. Wellner and Zhang (2000) showed the asymptotic efficiency gain of $\hat{\Lambda}_n$ over $\hat{\Lambda}_n^{ps}$ in their Figures 3 and 4. In this simulation study, we study the efficiency gain of the $\hat{F}_n$ over $\hat{\Lambda}_n$ by plotting $(\mathrm{Var}(\hat{F}_n)/\mathrm{Var}(\hat{\Lambda}_n))^{3/2}$.

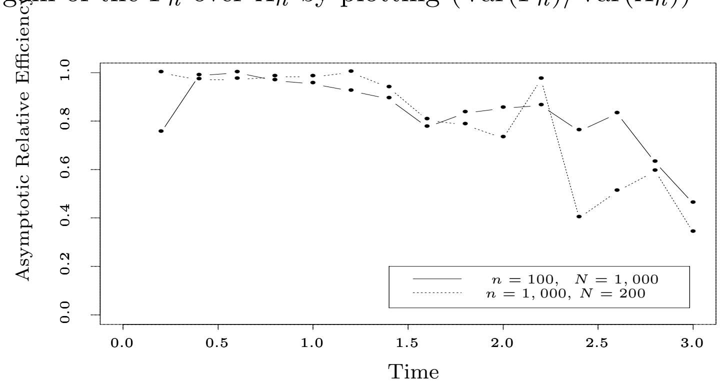

Figure 1. The asymptotic relative efficiency of $\hat{F}_n$ versus $\hat{\Lambda}_n$. N is the number of simulation runs; n is the number of subjects.

Random samples of univariate "mixed case" interval censored data $X_i = (\Delta^{(i)}_{K_i}, T^{(i)}_{K_i}, K_i)$, where $i = 1, \ldots, n$, were generated as follows: $K_i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$ with $P(K_i = k) = 1/8$ for $k = 1, \ldots, 8$; $T^{(i)}_{K_i}$ are ordered sequence of K_i random variables generated from the unit exponential distribution; Each failure time $Y^{(i)}$ was generated from the Weibull distribution with shape parameter equal to 2 and scale parameter equal to 2; the $\Delta^{(i)}_{K_i}$ follow from $\Delta_{k_i,j} = 1_{(T^{(i)}_{k_i,j-1}, T^{(i)}_{k_i,j}]}(Y^{(i)})$ for $j = 1, \ldots, k_i + 1$. The panel count data $\left\{(N^{(i)}_{K_i}, T^{(i)}_{K_i}, K_i) : i = 1, \ldots, n\right\}$ with panel counts $N^{(i)}_{K_i} = (N^{(i)}_{K_i,1}, N^{(i)}_{K_i,2}, \ldots, N^{(i)}_{K_i,K_i})$ are given by $N^{(i)}_{K_i,j} = 1_{[Y^{(i)} \le T^{(i)}_{K_i,j}]}$ for $j = 1, \ldots, K_i$. We ran Monte-Carlo simulations 1,000 times for the number of subjects $n = 100$, and 200 times for $n = 1,000$. The estimated efficiency gain of $\hat{F}_n$ over $\hat{\Lambda}_n$ is plotted in Figure 1. As shown there, the NPMLE $\hat{F}_n$ is more efficient than the NPMLE $\hat{\Lambda}_n$: for most of the points where $\hat{F}_n$ and $\hat{\Lambda}_n$ were estimated, the estimated asymptotic relative efficiency is above 40% for both simulations. When time is between 0.2 and 1.6, the estimated asymptotic relative efficiency for the simulation with $n = 100$ is close to that for the simulation with $n = 1,000$; when the time varies from 1.8 to 3.0, the estimated asymptotic relative efficiency for the simulation with $n = 100$ is generally higher than that for the simulation case with $n = 1,000$. Although Theorem 3.5 in Song (2001) or Song (2002) and Theorem 4.4 in Wellner and Zhang (2000) show that the asymptotic relative efficiency for the toy estimators of $\hat{F}_n$ and $\hat{\Lambda}_n$ is equal to 1 for a one jump process, the log-likelihood function for a one jump process in Wellner and Zhang (2000) is different from that of (1.2).

5. Asymptotic Properties of the NPMLE: Proof

In this section, we give the proofs of the asymptotic properties of the NPMLE in Section 3. More technical details can be found in Song (2001) and Song (2002).

Proof of Lemma 3.1. Construct the following brackets for the class $\mathcal{M}$: $m^l_i \equiv 2p^l_{F,i}/(p^l_{F,i}+p_{F_0})-1$, and $m^r_i \equiv 2p^r_{F,i}/(p^r_{F,i}+p_{F_0})-1$, where $[p^l_{F,i}, p^r_{F,i}]$, $i = 1, \ldots, I$, are brackets for class $\mathcal{P}$. Since $1/(1+t)$ is decreasing in t, and $p^l_{F,i} \le p^r_{F,i}$, so for all $m_F \in \mathcal{M}$, there exists $i \in \{1, \ldots, I\}$ such that $m^l_{F,i} \le m_F \le m^r_{F,i}$, and $|m^r_i - m^l_i|/2 = |p^r_{F,i}(p^r_{F,i} + p_{F_0}) - p^l_{F,i}/(p^l_{F,i} + p_{F_0})| = |p_{F_0}(p^r_{F,i} - p^l_{F,i})/[(p^r_{F,i} + p_{F_0})(p^l_{F,i} + p_{F_0})]| \le 1$. Thus, $|m^r_i - m^l_i| \le 2\,1_{[p_{F_0} \le \sigma]} + 2\,|m^r_i - m^l_i|\,1_{[p_{F_0} > \sigma]}$, and $\|m^r_i - m^l_i\|_{P,r} \equiv \left\{\int |m^r_i - m^l_i|^r dP\right\}^{1/r} \le (2^r\,\epsilon^r + \epsilon^r)^{1/r}$, if $\|(m^r_i - m^l_i)1_{p_{F_0} > \sigma}\|_{P,r} \le \epsilon$. This implies that $N_{[\,]}((2^r+1)^{1/r}\epsilon,\, \mathcal{M},\, L_r(P)) \le N_{[\,]}(\epsilon,\, \mathcal{G}_{\sigma_0},\, L_r(P))$. Also, $\|(m^r_i - m^l_i)\,1_{[p_{F_0} \ge \sigma]}\|^r_{P,r} \le 2^r \int |p^r_{F,i} - p^l_{F,i}|^r\,1_{[p_{F_0} > \sigma]}\,p^{1-r}_{F_0}\,d\nu$, since $p_{F_0}/(p^l_{F,i} + p_{F_0}) \le 1$, and $p_{F_0}/(p^r_{F,i} + p_{F_0}) \le 1$. Define $dQ_\sigma = (1_{[p_{F_0} > \sigma]}/p^{r-1}_{F_0})d\nu$. Then $N_{[\,]}(2\epsilon,\, \mathcal{G}_{\sigma_0(\epsilon)},\, L_r(P)) \le N_{[\,]}(\epsilon,\, \mathcal{P},\, L_r(Q_\sigma))$.

Proof of Theorem 3.1. Define $\mathcal{J}_n \equiv [t_0 - cn^{-1/3}, t_0 + cn^{-1/3}]$, and

$$H_n(t) \equiv \begin{cases} \theta\, f_0(t_0)\,(t - t_0 + cn^{-1/3}), & \text{if } t \in [t_0 - cn^{-1/3}, t_0), \\ \theta\, f_0(t_0)\,(t_0 + cn^{-1/3} - t), & \text{if } t \in [t_0, t_0 + cn^{-1/3}], \\ 0, & \text{otherwise.} \end{cases}$$

Note that $p_{F_n}(\underline{\delta}, \underline{t}) = p_{F_0}(\underline{\delta}, \underline{t}) + (\theta\, f_0(t_0)) \sum_{j=1}^{k+1} \delta_{k,j}\,(H_n(t_{k,j}) - H_n(t_{k,j-1}))$, and $|F_n(t_0) - F_0(t_0)| = \theta\, f_0(t_0)\, c\, n^{-1/3}$. Under the separation condition B, and for large n, $t_{k,j}$ and $t_{k,j-1}$ will not be in $\mathcal{J}_n$ at the same time. This indicates that $|H_n(t_{k,j}) - H_n(t_{k,j-1})| = H_n(t_{k,j})$ or $H_n(t_{k,j-1}) \le \theta\, f_0(t_0)\, c\, n^{-1/3}$. Thus $[1/(\sqrt{p_{F_n}} + \sqrt{p_{F_0}})]^2 = 1/(4\, p_{F_0}) + O(n^{-1/3})\, 1_{\mathcal{J}_n}(t)$, and $\int (p_{F_n} - p_{F_0})^2 d\nu = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k+1} \int (H_n(t_{k,j}) - H_n(t_{k,j-1}))^2\, dQ(\underline{t}_k | K = k) = (2/3)(\theta f_0(t_0))^2$ $c^3 n^{-1} (\sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k+1} (q_{k,j}(t_0) + q_{k,j-1}(t_0))) + o(n^{-1}) = O(n^{-1})$, if $\sum_{k=1}^{\infty}$ $P(K = k) \sum_{j=1}^{k+1} (q_{k,j}(t_0) + q_{k,j-1}(t_0)) < \infty$. The above condition is satisfied if $EK < \infty$ and $q_{k,j}$ is bounded above for all $j = 1, \ldots, k$, $k = 1, \ldots$. The Hellinger distance becomes $h^2(p_{F_n}, p_{F_0}) = (1/8) \int [(p_{F_n} - p_{F_0})^2 / p_{F_0}] d\nu + o(n^{-1})$. Now,

$$\int \frac{(p_{F_n} - p_{F_0})^2}{p_{F_0}} d\nu = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k+1} \int \frac{(H_n(t_{k,j}) - H_n(t_{k,j-1}))^2}{F_0(t_{k,j}) - F_0(t_{k,j-1})} dQ(\underline{t}_k | K = k)$$
$$= 2\, a(t_0)\, (\theta f_0(t_0))^2 c^3 n^{-1} + o(n^{-1}).$$

Thus, $h^2(p_{F_n}, p_{F_0}) = (1/4)\, a(t_0)\, (\theta f_0(t_0))^2 c^3 n^{-1} + o(n^{-1})$.

By (3.1), we have $n^{1/3} \inf_{T_n} \max\{E_{n,p_{F_n}} \mid T_n - F_n(\underline{t}_0)|,\ E_{n,p_{F_0}} \mid T_n - F_0(\underline{t}_0)|\} \ge 0.25 n^{1/3} |F_n(t_0) - F_0(t_0)| \{1 - h^2(p_{F_n}, p_{F_0})\}^{2n} \ge 0.25\theta c f_0(t_0)(1 - 0.25 a(t_0)(\theta f_0(t_0))^2 c^3 n^{-1} + o(n^{-1}))^{2n} \to 0.25\theta c f_0(t_0) e^{-a(t_0)(\theta f_0(t_0))^2 c^3/2}$. The last expression is maximized by $c \equiv \{1.5 a(t_0)\theta^2 f_0^2(t_0)\}^{-1/3}$, and the maximum value is $0.25\,(2\theta/3)^{1/3}\, e^{-1/3}\ (f_0(t_0)/a(t_0))^{1/3}$. Thus,

$$\liminf_{n\to\infty} n^{1/3} \max\{E_{n,p_{F_n}}\ |T_n - F_n(\underline{t}_0)|,\ E_{n,p_{F_0}}\ |T_n - F_0(\underline{t}_0)|\} \ge c_0 \left\{\frac{f_0(t_0)}{a(t_0)}\right\}^{1/3},$$

where $c_0 \equiv 0.25\,(2\theta/3)^{1/3}\, e^{-1/3}$ is a constant depending on θ.

Proof of Lemma 3.2. By Theorem 2.7.5 in Van der Vaart and Wellner (1996), for all $\epsilon > 0$ and for any probability measure Q, there exists a constant L such that $\log N_{[\,]}(\epsilon, \mathcal{F}, L_2(Q)) \le L\,(1/\epsilon)$. This implies that for all $F \in \mathcal{F}$, there exists a bracket $[F_i^l(t), F_i^r(t)]$ such that $F_i^l(t) \le F(t) \le F_i^r(t)$ for all t and some $i \in \{1, \ldots, I\}$, and $\|F_i^r(t_{K,j}) - F_i^l(t_{K,j})\|_{P_{K,j},2} \le \epsilon$ for $j = 1, \ldots, K$, $K = 1, 2, \ldots$. Then we have $F_i^l(t_{K,j}) - F_i^r(t_{K,j-1}) \le F(t_{K,j}) - F(t_{K,j-1}) \le F_i^r(t_{K,j}) - F_i^l(t_{K,j-1})$.

Note that $m_F = 2 \sum_{j=1}^{K+1} \delta_{K,j} [F(T_{K,j}) - F(T_{K,j-1})] / [(F(T_{K,j}) - F(T_{K,j-1})) + (F_0(T_{K,j}) - F_0(T_{K,j-1}))] - 1 \equiv 2 \sum_{j=1}^{K+1} \delta_{K,j} \Delta F_{K,j} / (\Delta F_{K,j} + \Delta F_{0,K,j}) - 1$. Also,

$\frac{d}{dx}(x/(x+a)) = a/(x+a)^2 > 0$ for $a > 0$ and, under conditions B and C, there exists a $\eta > 0$ such that $F_0(T_{K,j}) - F_0(T_{K,j-1}) > \eta$ for all $j = 1, \ldots, K$, $K = 1, 2, \ldots$. Thus, $0 < \Delta F_{0,K,j} \leq 1$, and $\Delta F_{K,j}/(\Delta F_{K,j} + \Delta F_{0,K,j})$ is increasing in $\Delta F_{K,j} \geq 0$. Choose the brackets of $\mathcal{M}$ as

$$m_i^l(\underline{X}) = 2\sum_{j=1}^{K+1} \delta_{K,j} \frac{F_i^l(T_{K,j}) - F_i^r(T_{K,j-1})}{[F_i^l(T_{K,j}) - F_i^r(T_{K,j-1})] + \Delta F_{0,K,j}} - 1,$$

$$m_i^r(\underline{X}) = 2\sum_{j=1}^{K+1} \delta_{K,j} \frac{F_i^r(T_{K,j}) - F_i^l(T_{K,j-1})}{[F_i^r(T_{K,j}) - F_i^l(T_{K,j-1})] + \Delta F_{0,K,j}} - 1,$$

where $i = 1, \ldots, I$. Then, it follows that $0.25P(|m_i^r - m_i^l|^2) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k+1} \int \{(F_i^r(t_{k,j}) - F_i^l(t_{k,j-1}))/[(F_i^r(t_{k,j}) - F_i^l(t_{k,j-1})) + \Delta F_{0,k,j}] - (F_i^l(T_{K,j}) - F_i^r(T_{K,j-1}))/[F_i^l(T_{K,j}) - F_i^r(T_{K,j-1}) + \Delta F_{0,K,j}]\}^2 \Delta F_{0,k,j}\, dQ(\underline{t}_k|K = k) \lesssim 4\epsilon^2\,(EK + 1) < \infty$, if $EK < \infty$. Here, symbol $\lesssim$ denotes that the left term is less than the right term up to a constant. This implies that $\log N_{[\,]}(\epsilon, \mathcal{M}, L_2(P)) = O(1/\epsilon)$.

Proof of Theorem 3.2. In order to apply Corollary 3.2.6 in Van der Vaart and Wellner (1996), we need to use Lemma 3.4.2 in Van der Vaart and Wellner (1996) to find an upper bound for $E^*\|\mathbb{G}_n\|_{\mathcal{M}_\delta}$, where $\mathcal{M}_\delta \equiv \{m_F : h(p_F, p_{F_0}) < \delta\}$. For $m_F \in \mathcal{M}_\delta$, $P(m_F^2) \leq \sqrt{2}h(p_F, p_{F_0}) < \sqrt{2}\,\delta$. Also, for all $m_F \in \mathcal{M}_\delta$, $\|m_F^2\|_\infty \leq 1$. Thus, by Lemma 3.4.2 in Van der Vaart and Wellner (1996), $E^*\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \leq \tilde{J}_{[\,]}(\delta,\, \mathcal{M}_\delta,\, L_2(P))\,(1 + \tilde{J}_{[\,]}(\delta,\, \mathcal{M}_\delta,\, L_2(P))/(\delta^2\sqrt{n}))$, where $\tilde{J}_{[\,]}(\delta, \mathcal{M}_\delta, L_2(P)) \equiv \int_0^\delta \{1 + \log N_{[\,]}(\epsilon, \mathcal{M}_\delta, L_2(P))\}^{1/2} d\epsilon \leq \int_o^\delta \{1 + \log N_{[\,]}(\epsilon,\, \mathcal{M}_{\mathcal{F}},\; L_2(P))\}^{1/2} d\epsilon = \int_0^\delta \{1 + O(1/\epsilon)\}^{1/2} d\epsilon = O(\delta^{1/2})$. Then it follows from Corollary 3.2.6 in Van der Vaart and Wellner (1996) that $E^*\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim \delta^{1/2}\;(1 + \delta^{1/2}/(\delta^2\sqrt{n})) = \delta^{1/2} + 1/(\delta\sqrt{n}) \equiv \phi(\delta)$. From $r_n^2 \cdot \phi(1/r_n) = r_n^2(r_n^{-1/2} + r_n/\sqrt{n}) \leq \sqrt{n}$, we have $r_n \leq n^{1/3}$. Thus the rate is $n^{1/3}$.

Proof of Theorem 3.3. We use Lemma 3.1 to control the bracketing entropy of the class $\mathcal{M}$. In this case, $r = 2$, so $dQ_\sigma = (1_{[p_{F_0} > \sigma]}/p_{F_0})\, d\nu$. Let $[p_i^l, p_i^r]$ be a pair of bracket for the class $\mathcal{P}$ as defined as $p_{F,i}^l \equiv \Pi_{j=1}^{K+1}(F_i^l(t_{K,j}) - F_i^r(t_{K,j-1}))^{\delta_{K,j}}$, $p_{F,i}^r \equiv \Pi_{j=1}^{K+1}(F_i^r(t_{K,j}) - F_i^l(t_{K,j-1}))^{\delta_{K,j}}$. Note that $\|p_i^r - p_i^l\|_{Q_\sigma,2}^2 = \int(F_i^r(t_{k,j}) - F_i^l(t_{k,j}))^2\, 1_{[p_{F_0} > \sigma]} p_{F_0}^{-1}\, d\nu + \int(F_i^r(t_{k,j-1}) - F_i^l(t_{k,j-1}))^2\, 1_{[p_{F_0} > \sigma]} p_{F_0}^{-1} d\nu + \int 2(F_i^r(t_{k,j}) - F_i^l(t_{k,j}))\,(F_i^r(t_{k,j-1}) - F_i^l(t_{k,j-1}))\, 1_{[p_{F_0} > \sigma]} p_{F_0}^{-1}\, d\nu$. Define $d\bar{Q}(\underline{t}_k|K = k) = (1_{[p_{F_0} > \sigma]}/\,(p_{F_0}\, \int(1_{[p_{F_0} > \sigma]}/p_{F_0})\, d\nu))d\nu$. Then $\int(F_i^r(t_{k,j}) - F_i^l(t_{k,j}))^2 d\bar{Q}(\underline{t}_k|K = k) \leq \epsilon^2$, and $\int(F_i^r(t_{k,j-1}) - F_i^l(t_{k,j-1}))^2\, d\bar{Q}(\underline{t}_k|K = k) \leq \epsilon^2$. By the Cauchy-Schwarz inequality, $\int(F_i^r(t_{k,j}) - F_i^l(t_{k,j}))\,(F_i^r(t_{k,j-1}) - F_i^l(t_{k,j-1}))\, d\bar{Q}(\underline{t}_k|K = k) \leq \{\int(F_i^r(t_{k,j}) - F_i^l(t_{k,j}))^2\, d\bar{Q}(\underline{t}_k|K = k) \cdot \int(F_i^r(t_{k,j-1}) - F_i^l(t_{k,j-1}))^2\, d\bar{Q}(\underline{t}_k|\; K = k)\}^{1/2} \leq \epsilon \cdot \epsilon = \epsilon^2$. Thus, we have $\|p_i^r - p_i^l\|_{Q_\sigma,2} \leq 2\,\epsilon\, \{\int(1_{[p_{F_0} > \sigma]}/p_{F_0})\, d\nu\}^{1/2} =$

$2\,\epsilon\,\{\int dQ_\sigma\}^{1/2}$. This shows $\log N_{[\,]}(\epsilon, \mathcal{P}, L_2(Q_\sigma)) = O(\{\int dQ_\sigma\}^{1/2}/\epsilon)$. Now for fixed $k \ge 1$ and $1 \le j \le k$, $\int 1_{[F_0(t_{k,j})-F_0(t_{k,j-1})>\sigma]}/[F_0(t_{k,j})-F_0(t_{k,j-1})]dQ(\underline{t}_k|K = k) = \int 1_{[F_0(t_{k,j})-F_0(t_{k,j-1})>\sigma]} q_{k,j,j-1} f_0(t_{k,j}) f_0(t_{k,j-1})/[(F_0(t_{k,j})-F_0(t_{k,j-1}))f_0(t_{k,j}) f_0(t_{k,j-1})]dt_{k,j}dt_{k,j-1} \le c_0 c_1 \log(c_1/\sigma)$, where c_0, c do not depend on j and k by assumption D. Hence we have $\int dQ_\sigma \le \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} c_0 c_1 \log(c_1/\sigma) = c_0 c_1 (EK)\, \log(c_1/\sigma)$.

Recall that we require $\int_{p_{F_0} \le \sigma} p_{F_0}\, d\nu \le \epsilon^2$, and $\int_{p_{F_0} \le \sigma} p_{F_0}\, d\nu \sigma(EK)$. Take $\sigma = \epsilon^2/EK$. Then, if $EK < \infty$, $\log N_{[\,]}(\epsilon, \mathcal{P}, L_2(Q_\sigma)) \lesssim (1/\epsilon)\, \log^{1/2}(1/\epsilon)$. By Lemma 3.1, $\log N_{[\,]}(\epsilon, \mathcal{M}_\delta, L_2(P)) \lesssim (1/\epsilon)\, \log^{1/2}(1/\epsilon)$. We have shown that $\int m_F^2 1_{[h(p_F, p_{F_0}) < \delta]} dP < \sqrt{2}\,\delta$. It follows that

$$\begin{aligned}\tilde{J}_{[\,]}(\delta, \mathcal{M}_\delta, L_2(P)) &= \int_0^\delta \sqrt{1 + \log N_{[\,]}(\epsilon, \mathcal{M}_\delta, L_2(P))}\, d\epsilon \\ &\lesssim \int_o^\delta \sqrt{\frac{1}{\epsilon}\, \log^{1/2}(\frac{1}{\epsilon})}\, d\epsilon \\ &\lesssim \delta^{1/2}\, \log^{1/4} \frac{1}{\delta}.\end{aligned}$$

By Lemma 3.4.2 in Van der Vaart and Wellner (1996), $E^*\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim \delta^{1/2} \log^{1/4} (1/\delta)\,(1 + (\delta^{1/2} \log^{1/4}(1/\delta))/(\sqrt{n}\,\delta^2)) = \delta^{1/2} \log^{1/4}(1/\delta) + \log^{1/2}(1/\delta)/(\sqrt{n}\,\delta) \equiv \phi(\delta)$. By Corollary 3.2.6 in Van der Vaart and Wellner (1996), we have $r_n^2 \phi(1/r_n) = r_n^2(r_n^{-1/2} \log^{1/4} r_n + (r_n\sqrt{\log r_n})/\sqrt{n}) \le \sqrt{n}$. This yields $r_n \le n^{1/3}\, \log^{-1/6} n$. Thus the rate is $n^{1/3}\, \log^{-1/6} n$.

Acknowledgements

This paper is part of author's Ph.D. dissertation under supervision of Professor Jon A. Wellner. Some writing and computation was carried out while the author was employed by the Boeing Company. The author thanks an editor, an associate editor and an anonymous referee for their valuable suggestions. This research was partially supported by National Science Foundation grant DMS-9971951.

References

Aragón, J. and Eberly, D. (1992). On convergence of convex minorant algorithm for distribution estimation with interval-censored data. *J. Comput. Graph. Statist.* **1**, 129-140.

Becker, N. G. and Melbye, M. (1991). Use of log-linear model to compute the empirical survival curve from interval-censored failure time data, with application to data on tests for HIV positivity. *Austral. J. Statist.* **33**, 125-133.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845-854.

Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41**, 993-945.

Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. Technical Report 378, Department of Statistics, Stanford University.

Groeneboom, P. (1996). Lectures on inverse problems. In *Lectures on Probability Theory and Statistics* **1648**, Lectures Notes in Mathematics (Edited by P. Bernard), 67-136. Springer-Verlag, Berlin.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, DMV Seminar Band 19, Birkhäuser, Basel.

Groeneboom, P. and Wellner, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10**, 388-400.

Huang, J. and Wellner, J. A. (1997) Interval censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* **123**, Lecture Notes in Statistics (Edited by D. Y. Lin and Thomas. R. Fleming), 123-169. Springer Verlag, Berlin.

Jongbloed, G. (1995). Three statistical inverse problems. Ph.D. thesis, Delft Technological University, The Netherlands.

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

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

Self, S. G. and Grossman, E. A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* **42**, 521-530.

Song, S. (2001). Estimation with bivariate interval censored data. Ph. D. thesis, University of Washington.

Song, S. (2002). Estimation with univariate "mixed case" interval censored data. Technical Report 414, Department of Statistics, University of Washington.

Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38**, 290-295.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical processes with Application to Statistics*. Springer, New York.

Van der Vaart, A. W. and Wellner, J. A. (2000). Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In *High Dimensional Probability* II (Edited by E. Giné, D. M. Mason and J. A. Wellner), 115-133. Birkhäuser, Basel.

Wellner, J. A. (1995). Interval censoring case 2: alternative hypotheses. In *Analysis of Censored Data* **27**, IMS Lecture Notes - Monograph Series (Edited by H. L. Koul and J. V. Deshpande), 271-291. IMS, Hayward.

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**, 272–291.

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting processes with panel count data. *Ann. Statist.* **28**, 779-814.

Applied Statistics, Phantom Works, The Boeing Company, P.O. Box 3707, MS 7L-22, Seattle, Washington 98124-2207, U.S.A.

E-mail: shuguang.song@boeing.com

(Received October 2002; accepted March 2003)