UNIVERSITY OF WISCONSIN
DEPARTMENT OF BIOSTATISTICS

Technical Report #110

October 1996
Revised 1/97

ON CONSISTENCY OF THE NONPARAMETRIC MLE OF SURVIVAL FOR LEFT TRUNCATED AND INTERVAL CENSORED DATA

Wei Pan
Richard Chappell, Ph.D.
Michael Kosorok, Ph.D.

Department of Biostatistics

MADISON, WISCONSIN

ON CONSISTENCY OF THE NONPARAMETRIC MLE OF SURVIVAL FOR LEFT TRUNCATED AND INTERVAL CENSORED DATA[1]

BY WEI PAN, RICK CHAPPELL AND MICHAEL R. KOSOROK

University of Wisconsin–Madison

The nonparametric monotone MLE is used to overcome the severe under-estimation of survival functions by the NPMLE for left truncated and interval censored data when a monotone hazard assumption is appropriate (Pan and Chappell, 1996). In this paper, we establish the consistency of the monotone MLE for interval censored data with or without left truncation in two different ways corresponding to two different realistic conditions.

Supported in part by the National Eye Institute Grant 1R29-EY10769-01.

AMS 1991 subject classifications. Primary 62G20; secondary 62G07.

Key words and phrases. Cumulative hazard function, monotone hazard, nonparametric maximum likelihood, vague convergence.

[1]Revised on February 3, 1997

1. Introduction. It is well known that the nonparametric maximum likelihood estimator (NPMLE for short) of survival functions tends to *severely* under-estimate the survival function at early times for small to medium sized samples with left-truncation with or without right-censoring (Lynden-Bell, 1971; Woodroofe, 1985; Tsai, 1988). Tsai (1988) suggests that in many applications it may be assumed or known that the survival function has a nondecreasing hazard function, and has derived the maximum likelihood estimate under this nondecreasing hazard assumption (called monotone MLE in the sequel) for right-censored data. By using the Channing House data (Hyde, 1977), Tsai has clearly shown that the monotone MLE overcomes the severe under-bias from the NPMLE for right-censored data. On the other hand, there has also been substantial interest in investigating the monotone MLE for other reasons: see also Grenander (1956) for complete data and Padgett and Wei (1980) for right censored data. Pan and Chappell (1996) give an algorithm based on the gradient projection to compute the monotone MLE for *interval-censored* data with or without left-truncation. Their simulation shows that the under-estimation from the NPMLE still persists for small to medium sized samples while the monotone MLE overcomes this under-estimation well; for large samples, both the monotone MLE and NPMLE tend to give unbiased estimates, though the former gives smaller variance. This suggests that both estimators may be consistent. It should be mentioned here that to our best knowledge, there are no results on consistency for either estimators with truncated and interval-censored data, although the consistency of the NPMLE for interval-censored data without truncation has already been established (Groeneboom and Wellner, 1992).

We emphasize that this study was motivated by real datasets subject to interval censoring and truncation such as the Massachusetts Health Care Panel Study (Chappell, 1991) and the Wisconsin Epidemiologic Study of Diabetic Retinopathy (Pan and Chappell, 1996). In these panel studies, baseline determinations of disease were followed by one or more followup examinations, resulting in interval or right censoring. Left truncation is induced by cohorts of subjects entering the studies at varying ages and event-free. Since these examples measure degenerations associated with aging, monotone increasing hazard functions are realistic.

Clinical examples of the monotone hazard are too numerous to mention. Certainly the term "aging" has direct implications of increasing risk. In product reliability research, a monotone hazard represents a "wear-out type of behavior" (Nelson, 1982).

In this paper, the consistency of the monotone MLE for interval censored data with or without left truncation is established under each of two different realistic conditions, as in Groeneboom and Wellner (1992) and Wellner (1995), respectively, for interval-censored data (without truncation). Roughly speaking, one is a *closeness* assumption requiring the two examination times to be sufficiently close, while the other is a *separation* assumption requiring the second examination to happen after at least a fixed period of time following the first examination. It seems that both assumptions are realistic in practice. On the other hand, for only interval-censored data, Groeneboom and Wellner (1992) hypothesize that under the closeness assumption the NPMLE of the survival function has an $(n \log n)^{-1/3}$ convergence rate. In contrast, Wellner (1995) conjectures that under the separation-like assumption the convergence rate is only $n^{-1/3}$. Hence each assumption has its own theoretical significance.

Consistency results under the closeness assumption are established by a general method which has also been used to prove the consistency of the NPMLE for interval censored data in Groeneboom and Wellner (1992). They attribute this approach to Jewell (1982) in proving the consistency of the NPMLE for mixing distributions. We begin by proving in Section 2 the consistency of the monotone MLE—in contrast to the usual NPMLE—for interval censored data. This will facilitate our discussion with left truncated and interval censored data in Section 3. Finally, in Section 4, we establish consistency under the separation condition.

2. Consistency with interval-censored data. In this section we establish the consistency of the monotone MLE for interval censored data. For simplicity of notation, we only investigate the interval censoring case 1 in Groeneboom and Wellner (1992), which corresponds to current status data. The result might be readily extended to case 2 of Groeneboom and Wellner (1992), but this is not pursued here.

Let $(X_1, U_1), \ldots, (X_n, U_n)$ be a sample of random variables in R_+^2, where X_i and U_i are independent random variables with distribution functions F_0 and G respectively. The only observations available are U_i and $\delta_i = I(X_i \le U_i)$, where $I(A)$ is the indicator of an event A. Our assumptions are:

(A1) X_i and U_i are independent.

(A2) F_0 is continuous, and $P_{F_0} \ll P_G$ (the probability measure P_{F_0} introduced by F_0, is absolutely continuous with respect to the probability measure P_G, induced by G).

(A3) X_i has **nondecreasing** hazard function λ_0.

Notice that except (A3) (which is necessary in discussing the monotone MLE), the other conditions are the same as for case 1 interval censoring in Groeneboom and Wellner (1992, page 75).

From the given sample, the log-likelihood is

$$\log L = \sum_{i=1}^{n} \delta_i \log(F_0(U_i)) + (1-\delta_i)\log(1-F_0(U_i)) = \sum_{i=1}^{n}\left[\delta_i \log\left(e^{\int_0^{U_i}\lambda_0(t)dt} - 1\right) - \int_0^{U_i}\lambda_0(t)dt\right].$$

The cumulative hazard is defined as $V(u) = \int_0^u \lambda(t)dt$. It is not difficult to verify that "$\lambda(u)$ is nondecreasing (and nonnegative)" is equivalent to "$V(u)$ is convex (and nondecreasing)". The above log-likelihood in terms of the cumulative hazard V of X_i, divided by n, can be rewritten as

$$\psi(V) := \int_{R^2}\left\{I(x \le u)\log\{e^{V(u)} - 1\} - V(u)\right\} dP_n(x,u), \tag{2.1}$$

where P_n is the empirical probability measure of the pairs (X_i, U_i), $1 \le i \le n$. The *monotone maximum likelihood estimator* (monotone MLE) F_n of F is a (right-continuous) distribution function $1-e^{-V_n}$, where V_n maximizes (2.1) and is **convex**. Naturally, we denote the survival functions $S_n = 1 - F_n$ and $S = 1 - F$.

From now on, we denote the true cumulative hazard and true distribution function of X_i as V_0 and S_0 respectively. Since the monotone MLE V_n maximizes (2.1), we have, for each $\epsilon \in (0,1)$:

$$\psi\left((1-\epsilon)V_n + \epsilon V_0\right) - \psi(V_n) \le 0,$$

by using the fact that both V_n and V_0 are convex implies that $(1-\epsilon)V_n + \epsilon V_0$ is also convex. Hence we obtain

$$\lim_{\epsilon \downarrow 0} \frac{\psi\left((1-\epsilon)V_n + \epsilon V_0\right) - \psi(V_n)}{\epsilon} \le 0.$$

Now we evaluate this limit.

$$\begin{aligned}
\log \frac{e^{(1-\epsilon)V_n + \epsilon V_0} - 1}{e^{V_n} - 1} &= \log \frac{e^{\epsilon(V_0 - V_n)} - e^{-V_n}}{1 - e^{-V_n}} \sim \log \frac{1 + \epsilon(V_0 - V_n) - e^{-V_n}}{1 - e^{-V_n}} \\
&= \log\left\{1 + \frac{\epsilon(V_0 - V_n)}{1 - e^{-V_n}}\right\} \sim \frac{\epsilon(V_0 - V_n)}{1 - e^{-V_n}}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\lim_{\epsilon \downarrow 0} \frac{\psi\left((1-\epsilon)V_n + \epsilon V_0\right) - \psi(V_n)}{\epsilon} \\
&= \int \left\{ I(x \le u) \frac{V_0(u) - V_n(u)}{1 - e^{-V_n(u)}} - (V_0(u) - V_n(u)) \right\} dP_n(x, u) \\
&= \int \left\{ I(x \le u) \frac{e^{-V_n(u)}(V_0(u) - V_n(u))}{1 - e^{-V_n(u)}} - I(x > u)(V_0(u) - V_n(u)) \right\} dP_n(x, u),
\end{aligned}$$

and we therefore obtain

$$\int \left\{ I(x \le u) \frac{S_n(u)}{1 - S_n(u)} \log \frac{S_n(u)}{S_0(u)} - I(x > u) \log \frac{S_n(u)}{S_0(u)} \right\} dP_n(x, u) \le 0. \tag{2.2}$$

Now we take our sample space Ω as the space of all infinite sequences

$$(X_1, U_1), (X_2, U_2), \ldots,$$

endowed with the Borel σ-algebra generated by the product topology on $\prod_1^\infty R^2$, and the product measure $\mathbf{P}$. Denote a point of this sample space by ω. To indicate the dependence on ω, we will write $S_n(u; \omega)$ instead of $S_n(u)$, and likewise $P_n(x, u; \omega)$ instead of $P_n(x, u)$.

Fix $\epsilon \in (0, \frac{1}{2})$ and let a and b be chosen such that

$$S_0(a) = 1 - \epsilon \text{ and } S_0(b) = \epsilon. \tag{2.3}$$

By the strong law of large numbers, we have that $P_n(\cdot, \cdot; \omega)$ converges weakly to P for all ω in a set B such that $\mathbf{P}(B) = 1$. Fix an $\omega \in B$. By the Helly selection theorem, the sequence of functions $S_n(\cdot;\omega)$ has a subsequence $S_{n_k}(\cdot;\omega)$ that converges vaguely to a nonincreasing right continuous function S taking values in $[0,1]$. (At this point we have not shown that S is a survival function.)

Now we prove that both $Q_1^n := S_n(\log S_n - \log S_0)/(1-S_n)$ and $Q_2^n := \log S_0 - \log S_n$ are bounded for $t \in [a,b]$ and all n sufficiently large. Since

$$\lim_{y\to 0} y\log y = 0 \text{ and } \lim_{y\to 1}\frac{y\log y}{1-y} = -1,$$

the only possibilities for Q_1^n and Q_2^n to be unbounded are S_n tending to 1 and 0, respectively. In either cases, Q_1^n and Q_2^n will go to positive infinity. But this cannot happen due to (2.2) and the fact that $P_n(\cdot,\cdot;\omega)$ converges weakly to P.

For example, if Q_2^n is unbounded, then for any large M and $n<\infty$, $\exists\ t_0 \in [a,b]$ such that $\log S_0(t_0) - \log S_n(t_0) > M$. First consider $t_0 < b$. Then by the monotonicity of S_n, we have $\log S_0(u) - \log S_n(u) > M$ for all $u \in [t_0, b]$. By weak convergence of $P_n(\cdot,\cdot;\omega)$ to P, this will contradict (2.2). For $t_0 = b$, we may replace ϵ with $\epsilon/2$ and repeat the above argument. A similar argument applies to Q_1^n.

Hence $\exists\ M \in (0,\infty)$ such that for all n large enough,

$$|S_n(\log S_n - \log S_0)/(1-S_n)| + |\log S_0 - \log S_n| \le M \tag{2.4}$$

and

$$|S(\log S - \log S_0)/(1-S)| + |\log S_0 - \log S| \le M. \tag{2.5}$$

The last inequality directly follows from the vague convergence of $S_{n_k}(\cdot;\omega)$ to S.

As in Groeneboom and Wellner (1992), we have the following key lemma:

LEMMA 1 *Let a and b be defined as in (2.3). We then have*

$$\begin{aligned}\lim_{k\to\infty}\int_{R\times[a,b]}&\left\{I(x\le u)\frac{S_{n_k}(u;\omega)}{1-S_{n_k}(u;\omega)}\log\frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x>u)\log\frac{S_{n_k}(u;\omega)}{S_0(u)}\right\}dP_n(x,u;\omega)\\ &= \int_{R\times[a,b]}\left\{I(x\le u)\frac{S(u)}{1-S(u)}\log\frac{S(u)}{S_0(u)} - I(x>u)\log\frac{S(u)}{S_0(u)}\right\}dP(x,u)\end{aligned} \tag{2.6}$$

and

$$(2.7) \quad \int_{R\times[a,b]} \left\{ I(x \le u) \frac{S(u)}{1-S(u)} \log \frac{S(u)}{S_0(u)} - I(x > u) \log \frac{S(u)}{S_0(u)} \right\} dP(x,u) \le 0.$$

PROOF The proof follows the same arguments used in Groeneboom and Wellner (1992). Fix $0 < \delta < 1$ and take a grid of points $a = t_0 < t_1 < ... < t_m = b$ on $[a,b]$ such that $m = 1 + [1/\delta^2]$ and

$$G(t_i) - G(t_{i-1}) = \{G(b) - G(a)\}/m, i = 1, ..., m.$$

First we suppose, for simplicity, that the points t_i are points of continuity for the function S, which is the vague limit of the sequence of functions $S_{n_k}(\cdot;\omega)$.

Let K be the (possibly empty) set of indices i, $i = 1, ..., m$ such that

$$\max\left\{ \left| \frac{S(t_{i-1})}{1-S(t_{i-1})} \log \frac{S(t_{i-1})}{S_0(t_{i-1})} - \frac{S(t_i)}{1-S(t_i)} \log \frac{S(t_i)}{S_0(t_i)} \right|, \left| \log \frac{S(t_{i-1})}{S_0(t_{i-1})} - \log \frac{S(t_i)}{S_0(t_i)} \right| \right\} \ge \delta.$$

By (2.5), the number of indices of this type is not bigger than $1 + [M/\delta]$. Let L be the remaining set of indices i, $i = 1, ..., m$. Denoting the intervals $[t_0, t_1]$ by J_1 and the intervals $(t_{i-1}, t_i]$ by J_i, $i > 1$, we get

$$\begin{aligned} &\int_{R\times[a,b]} \left\{ I(x \le u) \frac{S_{n_k}(u;\omega)}{1-S_{n_k}(u;\omega)} \log \frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x > u) \log \frac{S_{n_k}(u;\omega)}{S_0(u)} \right\} dP_n(x,u;\omega) \\ = &\sum_{i=1}^{m} \int_{R\times J_i} \left\{ I(x \le u) \frac{S_{n_k}(u;\omega)}{1-S_{n_k}(t;\omega)} \log \frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x > u) \log \frac{S_{n_k}(u;\omega)}{S_0(u)} \right\} dP_n(x,u;\omega). \end{aligned}$$

Since $S_{n_k}(u_k;\omega)$ converges to $S(u_i)$ for each i, $0 \le i \le m$, we get, for sufficiently large k:

$$(2.8) \quad \begin{aligned} \left| \frac{S_{n_k}(t_{i-1};\omega)}{1-S_{n_k}(t_{i-1};\omega)} \log \frac{S_{n_k}(t_{i-1};\omega)}{S_0(t_{i-1})} - \frac{S_{n_k}(t_i;\omega)}{1-S_{n_k}(t_i;\omega)} \log \frac{S_{n_k}(t_i;\omega)}{S_0(t_i)} \right| &< 2\delta, i \in L \\ \left| \log \frac{S_{n_k}(t_{i-1};\omega)}{S_0(t_{i-1})} - \log \frac{S_{n_k}(t_i;\omega)}{S_0(t_i)} \right| &< 2\delta, i \in L. \end{aligned}$$

Hence

$$\int_{R\times[a,b]} \left\{ I(x \le u) \frac{S_{n_k}(u;\omega)}{1-S_{n_k}(u;\omega)} \log \frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x > u) \log \frac{S_{n_k}(u;\omega)}{S_0(u)} \right\} dP_n(x,u;\omega)$$

$$
\begin{aligned}
&= \sum_{i\in K}\int_{R\times J_i}\left\{I(x\le u)\frac{S_{n_k}(u;\omega)}{1-S_{n_k}(u;\omega)}\log\frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x>u)\log\frac{S_{n_k}(u;\omega)}{S_0(u)}\right\}dP_n(x,u;\omega)\\
&+ \sum_{i\in L}\int_{R\times J_i}\left\{I(x\le u)\frac{S_{n_k}(u;\omega)}{1-S_{n_k}(u;\omega)}\log\frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x>u)\log\frac{S_{n_k}(u;\omega)}{S_0(u)}\right\}dP_n(x,u;\omega)
\end{aligned}
$$

(2.9)

$$
\begin{aligned}
&= \int_{R\times[a,b]}\left\{I(x\le u)\frac{S_{n_k}(u;\omega)}{1-S_{n_k}(u;\omega)}\log\frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x>u)\log\frac{S_{n_k}(u;\omega)}{S_0(u)}\right\}dP(x,u)\\
&+r_k(\omega),
\end{aligned}
$$

where $|r_k(\omega)|\le c\delta$, for a constant $c>0$. This can be seen by replacing $S_{n_k}(u;\omega)$ on each interval J_i by its value $S_{n_k}(u_i;\omega)$ at the right end point of the interval, and by noting that for large k

$$
\left|\log\frac{S_{n_k}(u;\omega)}{S_0(u)} - \log\frac{S_{n_k}(t_i;\omega)}{S_0(t_i)}\right| < 2\delta,
$$

if $i\in L$ (with a similar inequality for the other term). On the intervals J_i, with $i\in K$, we use (2.4). Note that

$$
\sum_{i\in K} P(R\times J_i)\to 0,\ \text{ if } \delta\downarrow 0,
$$

since $P(R\times J_i)$ is of order $O(\delta^2)$, while the number of intervals J_i such that $i\in K$ is of order $O(1/\delta)$.

On the other hand we have by dominated convergence:

$$
\begin{aligned}
&\lim_{k\to\infty}\int_{R\times[a,b]}\left\{I(x\le u)\frac{S_{n_k}(u;\omega)}{1-S_{n_k}(u;\omega)}\log\frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x>u)\log\frac{S_{n_k}(u;\omega)}{S_0(u)}\right\}dP(x,u)\\
(2.10)\qquad &= \int_{R\times[a,b]}\left\{I(x\le u)\frac{S(u)}{1-S(u)}\log\frac{S(u)}{S_0(u)} - I(x>u)\log\frac{S(u)}{S_0(u)}\right\}dP(x,u).
\end{aligned}
$$

Combining (2.9)and (2.10) we obtain

$$
\begin{aligned}
&\int_{R\times[a,b]}\left\{I(x\le u)\frac{S_{n_k}(u;\omega)}{1-S_{n_k}(u;\omega)}\log\frac{S_{n_k}(u;\omega)}{S_0(u)} - I(x>u)\log\frac{S_{n_k}(u;\omega)}{S_0(u)}\right\}dP_{n_k}(x,u;\omega)\\
(2.11)\qquad &= \int_{R\times[a,b]}\left\{I(x\le u)\frac{S(u)}{1-S(u)}\log\frac{S(u)}{S_0(u)} - I(x>u)\log\frac{S(u)}{S_0(u)}\right\}dP(x,u)+r_k'(\omega),
\end{aligned}
$$

where $|r_k'(\omega)|\le c'\delta$, for a constant $c'>0$.

If one or more of the points t_i is not a point of continuity of S, we shift the point t_i a bit to the left or right, in order to get continuity points (using the fact that the continuity points of S are dense). So in all cases we get a relation of type (2.11). Since δ can be chosen arbitrarily small, (2.6) now follows. Relation (2.7) immediately follows from (2.6) and (2.2). □

By monotone convergence we now obtain from (2.7):

$$
\begin{aligned}
&\int_{R^2}\left\{I(x\le u)\frac{S(u)}{1-S(u)}\log\frac{S(u)}{S_0(u)}-I(x>u)\log\frac{S(u)}{S_0(u)}\right\}dP(x,u)\\
(2.12)\qquad &= \lim_{a\downarrow 0,b\to\infty}\int_{R\times[a,b]}\left\{I(x\le u)\frac{S(u)}{1-S(u)}\log\frac{S(u)}{S_0(u)}-I(x>u)\log\frac{S(u)}{S_0(u)}\right\}dP(x,u)\\
&\le 0.
\end{aligned}
$$

However, this can only happen if $S=S_0$. Since

$$
\begin{aligned}
&\int_{R^2}\left\{I(x\le u)\frac{S(u)}{1-S(u)}\log\frac{S(u)}{S_0(t)}-I(x>u)\log\frac{S(u)}{S_0(u)}\right\}dP(x,u)\\
&= \int_R\left\{\frac{S(u)}{1-S(u)}\log(\frac{S(u)}{S_0(u)})(1-S_0(u))-\log(\frac{S(u)}{S_0(u)})S_0(u)\right\}dG(u)\\
&= \int_R\left\{\frac{S(u)-S_0(u)}{1-S(u)}(\log S(u)-\log S_0(u))\right\}dG(u),
\end{aligned}
$$

we claim that the last expression is strictly greater than 0, unless $S=S_0$.

We now prove our claim. If $0<S_0(u)<1$ and $y\in(0,1)$, then

$$
(2.13)\qquad \frac{y-S_0(u)}{1-y}(\log y-\log S_0(u))\begin{cases} =0, & \text{if } y=S_0(u);\\ >0, & \text{if } y\ne S_0(u).\end{cases}
$$

By the monotonicity of S, the monotonicity and continuity of S_0, and the absolute continuity of P_{S_0} with respect to P_G, we have $S\ne S_0\Longrightarrow S(u)\ne S_0(u)$ on an interval of increase in G; hence by (2.13),

$$
\int_R\left\{\frac{S(u)-S_0(u)}{1-S(u)}(\log S(u)-\log S_0(u))\right\}dG(u)>0
$$

if $S\ne S_0$. This contradicts (2.12).

Thus we have proved that for each ω outside a set of probability zero, each subsequence of the sequence of functions $S_n(\cdot;\omega)$ has a vaguely convergent subsequence, and that all these convergent subsequences have the same limit S_0. This yields the following consistency result for the monotone MLE with interval censored data:

Theorem 1 *With interval censored data and under conditions (A1),(A2) and (A3), as the sample size $n \to \infty$, the monotone MLE S_n converges weakly to S_0, with probability one.*

In fact, as in Groeneboom and Wellner (1992), since S_0 is continuous, we have that S_n converges with probability one to S_0 in the supremum distance on the set of distribution functions, i.e.

$$\mathbf{P}\left\{\lim_{n\to\infty}\sup_{t\in R}|S_n(t)-S_0(t)|=0\right\}=1.$$

Though the consistency of the NPMLE under the conditions (A1) and (A2) has already been established in Groeneboom and Wellner (1992), the above argument can also establish it. The key point is that (2.2) still holds if S_n is the NPMLE without condition (A3). However, the difference between our argument and that in Groeneboom and Wellner (1992) is that our log-likelihood (2.1) is parameterized in terms of the cumulative hazard function instead of the distribution function.

3. Consistency with left-truncated and interval-censored data under the closeness assumption. We now establish the consistency of the monotone MLE for left truncated and interval censored data under the closeness assumption. Let (X_1, T_1, U_1), ...,(X_n, T_n, U_n) be a sample of random variables in R^3_+, where X_i is a non-negative random variable whose distribution function F_0 (or equivalently its survival function S_0) is what we are interested in estimating, and where T_i and U_i are truncation and censoring random variables respectively. The only observations available are $(T_i, U_i, I(T_i \le X_i \le U_i), I(X_i > U_i))$. Note that due to left-truncation, $X_i \ge T_i$ for all $i = 1, ..., n$. Our assumptions are:

(B1) T_i and U_i are independent of X_i.

(B2) F_0 is continuous; T_i and U_i have a joint continuous distribution function H such that $T_i < U_i$ with probability one.

(B3) H has a density h with respect to Lebesgue measure, satisfying

$$h(t,u) > 0, \text{if } 0 < F_0(t) < F_0(u) < 1. \tag{3.1}$$

(B4) X_i has **nondecreasing** hazard function λ_0.

The conditions (B1)-(B3) are the same as those for interval censoring case 2 in Groeneboom and Wellner (1992, page 82); in particular, (B3) is the so called *"closeness"* assumption which guarantees that the two examination times T and U can be close enough with positive probability. (B4) arises naturally in our context.

The log-likelihood divided by n, in terms of the cumulative hazard V of X_i, can be written as:

$$\psi(V) := \int_{R^3} \left\{ I(t \le x \le u) \log\{e^{V(u)-V(t)} - 1\} - I(x \ge t)\{V(u) - V(t)\} \right\} dP_n(x,t,u). \tag{3.2}$$

If V_n is convex and maximizes the likelihood (i.e., V_n is the monotone MLE of V_0) under the condition V_0 is convex (since λ_0 is nondecreasing), as in the last section, we have

$$\lim_{\epsilon \downarrow 0} \frac{\psi\left((1-\epsilon)V_n + \epsilon V_0\right) - \psi(V_n)}{\epsilon} \le 0.$$

Evaluating this limit,

$$\begin{aligned} &\lim_{\epsilon \downarrow 0} \frac{\psi\left((1-\epsilon)V_n + \epsilon V_0\right) - \psi(V_n)}{\epsilon} \\ &= \int \left\{ \left[\frac{I(t \le x \le u)}{1 - e^{V_n(t) - V_n(u)}} - I(x \ge t) \right] [V_0(u) - V_n(u) + V_n(t) - V_0(t)] \right\} dP_n(x,t,u). \end{aligned}$$

Writing this in terms of survival functions, we then have

$$\int \left\{ \left[\log \frac{S_n(u) S_0(t)}{S_0(u) S_n(t)} \right] \left[\frac{I(t \le x \le u)}{1 - \frac{S_n(u)}{S_n(t)}} - I(x \ge t) \right] \right\} dP_n(x,t,u) \le 0. \tag{3.3}$$

We take our sample space to be similar to that given in the last section. By the strong law of large numbers, we see that $P_n(\cdot,\cdot,\cdot;\omega)$ converges weakly to P, for all ω in a set B such that $\mathbf{P}(B) = 1$. Fix an $\omega \in B$. By the Helly selection theorem we have that the sequence of functions $S_n(\cdot;\omega)$ has a subsequence $S_{n_k}(\cdot;\omega)$ converging vaguely to a nonincreasing right continuous function S, taking values in $[0,1]$.

Fix $\epsilon \in (0, \frac{1}{2})$ and let a and b be chosen such that

$$S_0(a) = 1 - \epsilon \text{ and } S_0(b) = \epsilon. \tag{3.4}$$

Define the set A_ϵ by

$$A_\epsilon = \{(t,u) : S_0(t) \leq 1-\epsilon, S_0(u)/S_0(t) \leq 1-\epsilon, S_0(u) \geq \epsilon\}.$$

Noticing that

$$\lim_{y\to 0} \log y \left(\frac{1}{1-y} - 1\right) = 0 \text{ and } \lim_{y\to 1} \frac{\log y}{1-y} = -1,$$

we have that if

$$\left[\log \frac{S_n(u;\omega)S_0(t)}{S_0(u)S_n(t;\omega)}\right]\left[\frac{1}{1-\frac{S_n(u;\omega)}{S_n(t;\omega)}} - 1\right]$$

is unbounded, it must be tending to positive infinity. But this cannot happen. Otherwise it would contradict (3.3). The proof is similar to that given in the last section and in Groeneboom and Wellner (1992, page 82). Hence we may assume that there exists an $M \in (0,\infty)$ such that

$$\left|\left[\log \frac{S_n(u;\omega)S_0(t)}{S_0(u)S_n(t;\omega)}\right]\left[\frac{1}{1-\frac{S_n(u;\omega)}{S_n(t;\omega)}} - 1\right]\right| + \left|\log \frac{S_n(u;\omega)S_0(t)}{S_0(u)S_n(t;\omega)}\right| \leq M, \tag{3.5}$$

for $(t,u) \in A_\epsilon$ and for all sufficiently large n.

As before, we have the following key lemma:

LEMMA 2 *With the above notation, we have that*

$$\begin{aligned} &\lim_{k\to\infty} \int_{R\times A_\epsilon} \left\{\left[\log \frac{S_n(u;\omega)S_0(t)}{S_0(u)S_n(t;\omega)}\right]\left[\frac{I(t\leq x\leq u)}{1-\frac{S_n(u;\omega)}{S_n(t;\omega)}} - I(x\geq t)\right]\right\} dP_n(x,t,u;\omega) \\ &= \int_{R\times A_\epsilon} \left\{\left[\log \frac{S(u)S_0(t)}{S_0(u)S(t)}\right]\left[\frac{I(t\leq x\leq u)}{1-\frac{S(u)}{S(t)}} - I(x\geq t)\right]\right\} dP(x,t,u) \end{aligned} \tag{3.6}$$

and

$$\int_{R\times A_\epsilon} \left\{\left[\log \frac{S(u)S_0(t)}{S_0(u)S(t)}\right]\left[\frac{I(t\leq x\leq u)}{1-\frac{S(u)}{S(t)}} - I(x\geq t)\right]\right\} dP(x,t,u) \leq 0. \tag{3.7}$$

PROOF. The proof is similar to that given in Section 2 and in Groeneboom and Wellner (1992, page 83-83) and is therefore omitted. □

By monotone convergence we now obtain that

$$\begin{aligned}
&\int_{R^3}\left\{\left[\log\frac{S(u)S_0(t)}{S_0(u)S(t)}\right]\left[\frac{I(t\le x\le u)}{1-\frac{S(u)}{S(t)}}-I(x\ge t)\right]\right\}dP(x,t,u)\\
(3.8)\qquad &= \lim_{\epsilon\downarrow 0}\int_{R\times A_\epsilon}\left\{\left[\log\frac{S(u)S_0(t)}{S_0(u)S(t)}\right]\left[\frac{I(t\le x\le u)}{1-\frac{S(u)}{S(t)}}-I(x\ge t)\right]\right\}dP(x,t,u)\\
&\le 0.
\end{aligned}$$

However, this can only happen if $S = S_0$, provided we can show that

$$\begin{aligned}
&\int_{R^3}\left\{\left[\log\frac{S(u)S_0(t)}{S_0(u)S(t)}\right]\left[\frac{I(t\le x\le u)}{1-\frac{S(u)}{S(t)}}-I(x\ge t)\right]\right\}dP(x,t,u)\\
&= \int_{R^2}\left\{\left[\log\frac{S(u)S_0(t)}{S_0(u)S(t)}\right]\frac{\frac{S(u)S_0(t)}{S_0(u)S(t)}-1}{1-\frac{S(u)}{S(t)}}S_0(u)\right\}dH(t,u)
\end{aligned}$$

implies that (3.8) is strictly greater than 0. We now prove this.

If $0 < S_0(u) < S_0(t) < 1$ and $0 < S(u) < S(t) < 1$, then

$$(3.9)\qquad \left[\log\frac{S(u)S_0(t)}{S_0(u)S(t)}\right]\frac{\frac{S(u)S_0(t)}{S_0(u)S(t)}-1}{1-\frac{S(u)}{S(t)}}S_0(u)\begin{cases}=0, & \text{if } \frac{S(u)S_0(t)}{S_0(u)S(t)}=1;\\ >0, & \text{otherwise.}\end{cases}$$

If there exists (t_0, u_0) such that $\frac{S(u_0)S_0(t_0)}{S_0(u_0)S(t_0)} \neq 1$ (and also both $0 < F_0(t_0) < F_0(u_0) < 1$ and $0 < F(t_0) < F(u_0) < 1$), then $S(u_0)/S_0(u_0) \neq S(t_0)/S_0(t_0)$. But both S and S_0 are monotone and right continuous, so there must exist $\delta > 0$ such that $S(u)/S_0(u) \neq S(t)/S_0(t)$ on $[t_0, t_0+\delta] \times [u_0, u_0+\delta]$, on which $h(t,u) > 0$. Hence we will have that the left hand side of (3.8) is strictly greater than 0, which is a contradiction. Therefore, we have that $S(u) = S_0(u)S(t)/S_0(t)$ for any (t,u) satisfying both $0 < F_0(t) < F_0(u) < 1$ and $0 < F(t) < F(u) < 1$. Again, since both S and S_0 are right continuous and $S(0) = S_0(0) = 1$, we have for any $t < u$,

$$S(u) = \lim_{t\downarrow 0}\frac{S(t)}{S_0(t)}S_0(u) = \frac{S(0)}{S_0(0)}S_0(u) = S_0(u).$$

We have thus proved that $S = S_0$. As in Section 2, this provides the following consistency result for the monotone MLE with left truncated and interval censored data:

Theorem 2 *With left truncated and interval censored data, under conditions (B1), (B2), (B3) and (B4), as the sample size $n \to \infty$, the monotone MLE S_n converges weakly to S_0 with probability one.*

One evident implication is that, as in the previous section, since S_0 is continuous, S_n (the monotone MLE in Theorem 2) also converges with probability one to S_0 in the supremum distance on the set of distribution functions.

4. Consistency with left-truncated and interval-censored data under the separation assumption. Now we establish the consistency of the monotone MLE for left truncated and interval censored data under the separation assumption. Our approach is similar to that of Huang and Wellner (1995) in proving the consistency of the NPMLE for the Cox proportional hazards model with interval censored data. The assumptions needed are the same as in Section 3 except for the following *separation* assumption:

(C3) There exist positive constants $0 < \delta < a$ such that $0 < S_0(a) < 1$ and

$$g(t,u) \begin{cases} > 0, & \text{if } u - t \geq \delta \text{ and } (t,u) \in [0,a]^2; \\ = 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

Theorem 3 *With left truncated and interval censored data, under conditions (B1), (B2), (C3) and (B4), the monotone MLE $S_n(t)$ converges to $S_0(t)$ uniformly on $[0,a]$ with probability one.*

PROOF. The likelihood $L_i(S;Y_i)$ for the ith sample $Y_i = (X_i, T_i, U_i)$ is

$$L_i(S;Y_i) = \left(\frac{S(T_i) - S(U_i)}{S(T_i)}\right)^{I(T_i \leq X_i \leq U_i)} \left(\frac{S(U_i)}{S(T_i)}\right)^{I(X_i > U_i)}$$

Since the monotone MLE S_n maximizes the log-likelihood under the nondecreasing hazard assumption and S_0 has nondecreasing hazard by (B4), we have

$$\sum_{i=1}^{n} \log L(S_n;Y_i) - \log L(S_0;Y_i) = \sum_{i=1}^{n} \log \frac{L(S_n;Y_i)}{L(S_0;Y_i)} \geq 0. \tag{4.2}$$

But, by the concavity of the function $\log(x)$ and by Jensen's inequality, for any $0 < \alpha < 1$ and $x > 0$ we have

$$\log(1 - \alpha + \alpha x) \geq (1 - \alpha)\log 1 + \alpha \log x = \alpha \log x,$$

so we have

$$\frac{1}{n}\sum_{i=1}^{n} \log\left(1 - \alpha + \alpha\frac{L(S_n; Y_i)}{L(S_0; Y_i)}\right) = \int \log\left(1 - \alpha + \alpha\frac{L(S_n; y)}{L(S_0; y)}\right) dP_n(y) \geq 0. \tag{4.3}$$

As before, we take our sample space Ω to be the space of all infinite sequences

$$(X_1, T_1, U_1), (X_2, T_2, U_2), \ldots,$$

endowed with the usual Borel σ-algebra generated by the product topology on $\prod_1^\infty R^3$ and the product measure $\mathbf{P}$. By the strong law of large numbers, we have that $P_n(\cdot, \cdot; \omega)$ converges weakly to P for all ω in a set B such that $\mathbf{P}(B) = 1$. Fix an $\omega \in B$. By the Helly selection theorem, for the sequence S_n, there exists a subsequence (indexed as n_k later on) converging vaguely to some nonincreasing function S. Under condition (C3) it is easy to verify that $L(S_{n_k}; Y_i)/L(S_0; Y_i)$ is bounded. By the bounded convergence theorem and some standard arguments,

$$\lim_{k\to\infty} \int \log\left(1 - \alpha + \alpha\frac{L(S_{n_k}; y)}{L(S_0; y)}\right) dP_{n_k}(y) = \int \log\left(1 - \alpha + \alpha\frac{L(S; y)}{L(S_0; y)}\right) dP(y).$$

By (4.3) the above expression must be nonnegative. However, by Jensen's inequality it must be nonpositive. Therefore it must be zero, which leads to

$$S(t)/S(u) = S_0(t)/S_0(u) \qquad G \text{ a.s.}$$

Then by the facts that $S(0) = S_0(0) = 1$, that S and S_0 are nonincreasing, and that S_0 is continuous, we have $S = S_0$ on $[0, a]$. The uniform convergence is then implied by the continuity of S_0. This completes the proof. □

Acknowledgement The first author would like to thank Jerome Klotz and Michael Newton for their encouragement and suggestions; thanks also go to Yunlei Zhang, Jianjian Gong and Yinzhong Chen for their stimulating discussions.

REFERENCES

CHAPPELL, R. (1991). Sampling design of multiwave studies with an application to the Massachusetts Health Care Panel Study. *Statist. in Medicine* **10**, 1945-1958.

GRENANDER, U. (1956). On the theory of mortality measurement. Part II. *Skand. Aktuar.* **39**, 125-153.

GROENEBOOM, P. and WELLNER, J.A. (1992). Information bounds and nonparametric maximum likelihood estimation. *DMV Seminar Band* **19**, Birhauser Verlag, Basel.

HUANG, J. and WELLNER, J.A. (1995). Efficient estimation for the proportional hazards model with case 2 interval censoring. Tech. Rept., Statistics Dept., Univ. of Washington.

JEWELL, N.P. (1982). Mixtures of exponential distributions. *Ann. Stat* **10**, 1184-1186.

LYNDEN-BELL, D. (1971). A method for allowing for known observational selection in small samples applied to 3CR quasars. *Mon. Nat. Royal Astr. Soc.* **155**, 95-118.

NELSON, W.B. (1982). *Applied life data analysis.* New York: John Wiley & Sons.

PADGETT, W.J. and WEI, L.J. (1980). Maximum Likelihood Estimation of a Distribution Function with Increasing Failure Rate Based on Censored Observations. *Biometrika* **67**, 2, 470-474.

PAN, W. and CHAPPELL, R. (1996). Estimating Survival Curves with Left Truncated and Interval Censored Data under Monotone Hazards. Technical Report 109, Department of Biostatistics, Univ. of Wisconsin-Madison.

TSAI, W.Y. (1988). Estimation of the Survival Function with Increasing Failure Rate Based on Left Truncated and Right Censored Data. *Biometrika* **75**, 2, 319-324.

WELLNER, J.A. (1995). Interval Censoring, Case 2: Alternative Hypotheses. Tech. Rept., Statistics Dept., Univ. of Washington.