Biometrika (2007), **94**, 1, *pp*. 71–86
doi:10.1093/biomet/asm011

A pseudolikelihood method for analyzing interval censored data

BY BODHISATTVA SEN AND MOULINATH BANERJEE
Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
bodhi@umich.edu moulib@umich.edu

SUMMARY

We introduce a method based on a pseudolikelihood ratio for estimating the distribution function of the survival time in a mixed-case interval censoring model. In a mixed-case model, an individual is observed a random number of times, and at each time it is recorded whether an event has happened or not. One seeks to estimate the distribution of time to event. We use a Poisson process as the basis of a likelihood function to construct a pseudolikelihood ratio statistic for testing the value of the distribution function at a fixed point, and show that this converges under the null hypothesis to a known limit distribution, that can be expressed as a functional of different convex minorants of a two-sided Brownian motion process with parabolic drift. Construction of confidence sets then proceeds by standard inversion. The computation of the confidence sets is simple, requiring the use of the pool-adjacent-violators algorithm or a standard isotonic regression algorithm. We also illustrate the superiority of the proposed method over competitors based on resampling techniques or on the limit distribution of the maximum pseudolikelihood estimator, through simulation studies, and illustrate the different methods on a dataset involving time to HIV seroconversion in a group of haemophiliacs.

Some key words: Greatest convex minorant; Haemophilia data; Mixed-case interval censoring; Panel count data; Pseudolikelihood.

1. INTRODUCTION

In the mixed-case interval censoring model each individual is followed up at the clinic for a number of times, where this number and the times of inspection themselves can vary from individual to individual. It is determined between which two successive observation times the individual succumbed to infection/illness. It is of course possible that infection/illness may not occur by the last follow-up time. The term 'mixed-case' is used to indicate that the number of inspection times is patient-specific, and was first used by Schick & Yu (2000). Our interest lies in constructing confidence sets for F, the distribution of the random time to infection/illness.

The simplest form of mixed-case censoring is current status data, where the number of observation times on each patient is exactly one; see for example Groeneboom & Wellner (1992), Jewell & van der Laan (1995), Shiboski (1998), Banerjee & Wellner (2001, 2005) and Jewell et al. (2003). In this model, the distribution of the indicator of time to infection/illness, conditional on the single inspection time, is a Bernoulli random variable, and this makes likelihood inference easy. Banerjee & Wellner (2001) showed that the likelihood-ratio statistic for testing a pointwise hypothesis of the type $H_0 : F(t_0) = \theta_0$, for some prespecified point t_0, is asymptotically pivotal under H_0. This immediately provides a way of constructing

pointwise confidence bands for F by standard inversion of the likelihood-ratio statistic, with the critical values determined by the quantiles of the limiting pivotal distribution. While this result is, in principle, generalizable to mixed-case interval censoring, dealing with the likelihood function in the mixed-case model is considerably more difficult, at both a theoretical and a computational level. Only partial results, in fairly restrictive settings, exist thus far, about the limiting behaviour of the nonparametric maximum likelihood estimator; consequently, the limiting behaviour of the likelihood-ratio statistic for testing a pointwise null hypothesis is not tractable either. See for example, Groeneboom (1996), where the asymptotics of the behaviour of the nonparametric maximum likelihood estimator of F in a particular version of the Case 2 censoring model is established, and Song (2004), where estimation procedures for mixed-case censoring models and associated issues are presented.

The nonparametric maximum likelihood estimator and likelihood-ratio statistic for current status data are readily computable using appropriately modified versions of the pool-adjacent-violators algorithm (Robertson et al., 1988). The computations are based on explicit representations of the maximum likelihood estimates in terms of the given data and do not involve iterative schemes. However, maximization of the likelihood function in the mixed-case setting is much more complex and requires sophisticated optimization techniques. The EM algorithm can be employed but is extremely slow (Jongbloed, 1998); a faster algorithm is the modified iterative convex minorant algorithm of Jongbloed (1998), based on the Kuhn–Tucker conditions associated with the maximization problem. However, both methods involve iterating till convergence, and can therefore be quite slow. Alternative methods for computing nonparametric maximum likelihood estimators and likelihood-ratio-based intervals for interval censored data have been developed by Vandal et al. (2005) using graph-theoretic representations of the unconstrained and constrained estimators. These involve reduction techniques as well as versions of the EM algorithm and the vertex exchange method. It is not known how these methods compare to the modified iterative convex minorant algorithm in terms of speed. Finally, even if one obtains likelihood-ratio-based intervals using methods of the type discussed above, it is not clear how to calibrate them, since concrete inferential results for the likelihood ratio do not exist, though conjectures do.

Our approach is to think of mixed-case interval censored data as data on a one-jump counting process with counts available only at the inspection times and to use a pseudolikelihood function based on the marginal likelihood of a Poisson process to construct a pseudolikelihood-ratio statistic for testing null hypotheses of the form $H_0 : F(t_0) = \theta_0$. We show that under such a null hypothesis the statistic converges to a pivotal quantity. This result can now be used to construct confidence intervals for $F(t_0)$. The pseudolikelihood method that we adopt is based on an estimator originally proposed by Sun & Kalbfleisch (1995) whose asymptotic properties, under appropriate regularity conditions, were studied in Wellner & Zhang (2000). Indeed, our key result in § 2 draws freely on the work of Wellner & Zhang (2000) and our point of view here, the fact that the interval censoring situation can be thought of as a one-jump counting process to which, consequently, the results on the pseudolikelihood-based estimators can be applied, is motivated by their work. That said, our likelihood-ratio approach for computing confidence intervals has major advantages over the Wald-type intervals that can be derived from their work.

We now introduce the stochastic processes and derived functionals that are needed to describe the asymptotic distributions. For a real-valued function f defined on $\mathbb{R}$, let

slogcm(f, I) denote the left-hand slope of the greatest convex minorant of the restriction of f to the interval I. We abbreviate slogcm$(f, \mathbb{R})$ to slogcm(f). Also define

$$\text{slogcm}^0(f) = \{\text{slogcm}\,(f, (-\infty, 0]) \wedge 0\}\, 1_{(-\infty,0]} + [\text{slogcm}\,\{f, (0, \infty)\} \vee 0] 1_{(0,\infty)}\,.$$

For positive constants c and d define the process $X_{c,d}(z) = c\,W(z) + dz^2$, where $W(z)$ is standard two-sided Brownian motion starting from 0. Set $g_{c,d} = \text{slogcm}(X_{c,d})$ and $g^0_{c,d} = \text{slogcm}^0\,(X_{c,d})$. It is known that $g_{c,d}$ is a piecewise-constant increasing function, with finitely many jumps in any compact interval. The function $g^0_{c,d}$ has the same characteristics and differs, almost surely, from $g_{c,d}$ on a finite interval containing 0. In fact, with probability 1, $g^0_{c,d}$ is identically 0 in some random neighbourhood of 0, whereas $g_{c,d}$ is almost surely nonzero in some random neighbourhood of 0. Also, the length of the interval $D_{c,d}$ on which $g_{c,d}$ and $g^0_{c,d}$ differ is $O_p(1)$. For more detailed descriptions of the processes $g_{c,d}$ and $g^0_{c,d}$, see Banerjee and Wellner (2001), Wellner (2003) and M. Banerjee's unpublished 2000 Ph.D. thesis from the University of Washington. Thus, $g_{1,1}$ and $g^0_{1,1}$ are the unconstrained and constrained versions of the slope processes associated with the canonical process $X_{1,1}(z)$. By Brownian scaling, the slope processes $g_{c,d}$ and $g^0_{c,d}$ can be related in distribution to the canonical slope processes $g_{1,1}$ and $g^0_{1,1}$. This leads to the following lemma.

LEMMA 1. *For positive a and b, set*

$$\mathbb{D}_{a,b} = \int \left[\{g_{a,b}(u)\}^2 - \{g^0_{a,b}(u)\}^2\right] du$$

and abbreviate $\mathbb{D}_{1,1}$ *to* $\mathbb{D}$. *Then* $\mathbb{D}_{a,b}$ *has the same distribution as* $a^2\,\mathbb{D}$.

This is proved in Chapter 3 of M. Banerjee's thesis; alternatively, see Banerjee and Wellner (2001).

2. A PSEUDOLIKELIHOOD METHOD FOR ANALYZING MIXED-CASE INTERVAL CENSORED DATA

We describe our method more broadly in the context of a counting process and then specialize to the interval censoring situation. Suppose that $N = \{N(t) : t \geqslant 0\}$ is a counting process with mean function $E\,N(t) = \Lambda(t)$, K is an integer-valued random variable and $T = \left\{T_{k,j},\, j = 1, \ldots, k,\ k = 1, 2, \ldots\right\}$ is a triangular array of potential observation times. It is assumed that N and (K, T) are independent, that K and T are independent and that $T_{k,j-1} \leqslant T_{k,j}$, for $j = 1, \ldots, k$, for every k; we interpret $T_{k,0}$ as 0. Let $X = (N_K, T_K, K)$ be the observed random vector for an individual. Here K is the number of times that the individual was observed during a study, $T_{K,1} \leqslant T_{K,2} \leqslant \cdots \leqslant T_{K,K}$ are the times when they were observed and $N_K = \{N_{K,j} \equiv N(T_{K,j})\}_{j=1}^K$ are the observed counts at those times. The above scenario specializes easily to the mixed-case interval censoring model, when the counting process is $N(t) = 1(S \leqslant t)$, S being a positive random variable with distribution function F and independent of (T, K).

Suppose that we have data on n individuals; thus, we observe n independent and identically distributed copies of X, say $\mathcal{X} = (X_1, X_2, \ldots, X_n)$ where $X_i = (N^{(i)}_{K_i}, T^{(i)}_{K_i}, K_i)$, for $i = 1, \ldots, n$. We are interested in estimating the mean function $\Lambda(t)$ at a prespecified point of interest t_0.

Based on our data, we can construct a pseudolikelihood estimator, in the following manner. Pretend that the process $N(t)$ is a nonhomogeneous Poisson process. Then the

marginal distribution of $N(t)$ is

$$\mathrm{pr}\{N(t) = k\} = \exp\{-\Lambda(t)\}\,\frac{\Lambda^k(t)}{k!}, \quad (k = 0, 1, 2, \ldots).$$

Note that, under the Poisson process assumption, the successive counts on an individual $(N_{K,1}, N_{K,2}, \ldots)$, conditional on the $T_{K,j}$'s, are actually dependent. However, we choose to ignore the dependence in writing down a likelihood function for the data, conditional on the $T^{(i)}$'s and the K_i's. These do not involve Λ and hence will not contribute to the estimation procedure. Our likelihood function is

$$L_n^{\mathrm{ps}}(\Lambda \mid \mathcal{X}) = \prod_{i=1}^{n}\prod_{j=1}^{K_i} \exp\{-\Lambda(T_{K_i,j}^{(i)})\}\,\frac{\Lambda(T_{K_i,j}^{(i)})^{N_{K_i,j}^{(i)}}}{N_{K_i,j}^{(i)}!}.$$

Thus, the loglikelihood function, up to an additive constant not depending upon the parameter, is given by

$$l_n^{\mathrm{ps}}(\Lambda \mid \mathcal{X}) = \sum_{i=1}^{n}\sum_{j=1}^{K_i} \{N_{K_i,j}^{(i)} \log \Lambda(T_{K_i,j}^{(i)}) - \Lambda(T_{K_i,j}^{(i)})\}.$$

The above loglikelihood can be written in the following slightly neater way. Let $T_{(1)} < T_{(2)} < \cdots < T_{(M)}$ denote the ordered distinct observation times in the set of all observation time-points $\{T_{K_i,j}^{(i)}, j = 1, \ldots, K_i, \;\; i = 1, \ldots, n\}$. For $1 \leqslant l \leqslant M$, define

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

Thus w_l is the frequency of the lth largest observation time in the sample and $w_l\,\bar{N}_l$ is the total number of events that happened by the lth largest time. Writing $\Lambda(T_{(l)})$ as Λ_l, for convenience, we can represent the loglikelihood as

$$l_n^{\mathrm{ps}}(\Lambda \mid \mathcal{X}) = \sum_{l=1}^{M}\Big(w_l\,\bar{N}_l \log \Lambda_l - w_l\,\Lambda_l\Big). \tag{1}$$

We define the nonparametric estimator $\hat{\Lambda}_n$ of Λ to be the unique nondecreasing right-continuous step-function with possible jumps only occurring at the $T_{(i)}$'s, such that the above expression is maximized. Of course, only $\Lambda_1, \ldots, \Lambda_M$ are identifiable; the choice of $\hat{\Lambda}_n$ made above is arbitrary. Other conventions are possible, but will make no difference to the asymptotics. Thus, $\hat{\Lambda}_n$, which we will subsequently refer to as $\hat{\Lambda}$ for convenience, is the unconstrained maximum pseudolikelihood estimator. The constrained estimator $\hat{\Lambda}_n^{(0)}$, to be referred to subsequently as $\hat{\Lambda}^{(0)}$, is defined to be the unique nondecreasing step-function with possible jumps only at the $T_{(i)}$'s and at t_0, that maximizes (1) subject to the additional constraint that $\Lambda(t_0) = \theta_0$. Using the theory of generalized isotonic regression (Robertson et al., 1988, § 1·5), or by appealing to the Kuhn–Tucker theorem (Robertson et al., 1988, § 6·4), we can show that $\hat{\Lambda}(T_{(i)})$ is $\hat{f}_i$, where $(\hat{f}_1, \ldots, \hat{f}_M)$ minimizes $\sum_{i=1}^{M} w_i\,(g_i - f_i)^2$ over all $f_1 \leqslant \cdots \leqslant f_M$, with $g_i \equiv \bar{N}_i$ and w_i as defined above. Also, $\hat{\Lambda}^{(0)}(T_{(i)})$ is $\hat{f}_i^{(0)}$, where $(\hat{f}_1^{(0)}, \hat{f}_2^{(0)}, \ldots, \hat{f}_M^{(0)})$ solves the constrained isotonic least squares problem of minimizing $\sum_{i=1}^{M} w_i\,(g_i - f_i)^2$ over all $f_1 \leqslant \cdots \leqslant f_m \leqslant \theta_0 \leqslant f_{m+1} \leqslant \cdots \leqslant f_M$, with $T_{(m)} < t_0 < T_{(m+1)}$. The fact that none of the $T_{(i)}$'s can actually be equal to t_0, with probability 1, is guaranteed

by the regularity conditions under which the asymptotic results for this model will be established; in particular, see Assumptions A6 and A7 in the Appendix.

For points $\{(x_0, y_0), (x_1, y_1), \ldots, (x_k, y_k)\}$ where $x_0 = y_0 = 0$ and $x_0 < x_1 < \cdots < x_k$, consider the left-continuous function $P(x)$ such that $P(x_i) = y_i$ and such that $P(x)$ is constant on (x_{i-1}, x_i). We will denote the vector of slopes, i.e., left-derivatives, of the greatest convex minorant of $P(x)$ computed at the points $(x_1, x_2, \ldots, x_k)$ by slogcm $\{(x_i, y_i)\}_{i=0}^k$.

It is not difficult to see that

$$\{\hat{\Lambda}_i\}_{i=1}^M = \text{slogcm}\left\{\sum_{j=1}^i w_j, \sum_{j=1}^i w_j \bar{N}_j\right\}_{i=0}^M, \tag{2}$$

where summation over an empty set is interpreted as 0. Also,

$$\{\hat{\Lambda}_i^{(0)}\}_{i=1}^m = \theta_0 \wedge \text{slogcm}\left\{\sum_{j=1}^i w_j, \sum_{j=1}^i w_j \bar{N}_j\right\}_{i=0}^m, \tag{3}$$

where the minimum is interpreted as being taken componentwise, while

$$\{\hat{\Lambda}_i^{(0)}\}_{i=m+1}^M = \theta_0 \vee \text{slogcm}\left\{\sum_{j=m+1}^i w_j, \sum_{j=m+1}^i w_j \bar{N}_j\right\}_{i=m}^M, \tag{4}$$

where the maximum is once again interpreted as being taken componentwise.

Define the pseudolikelihood-ratio statistic as

$$2 \log \lambda_n = 2\{l_n^{\text{ps}}(\hat{\Lambda} \mid \mathcal{X}) - l_n^{\text{ps}}(\hat{\Lambda}^{(0)} \mid \mathcal{X})\}.$$

The limit distribution of $2 \log \lambda_n$ will be established under a number of assumptions. These are minor modifications of conditions given in Wellner & Zhang (2000), but, for the sake of completeness, we state them in the Appendix and there discuss the implications of these conditions in the interval censoring framework.

Under Assumptions A1–A4, there exist $a_0 < t_0 < b_0$ such that

$$\sup_{x \in [a_0, b_0]} \mid \hat{\Lambda}_n(x) - \Lambda(x) \mid \to 0,$$

almost surely. Also, if the null hypothesis holds,

$$\sup_{x \in [a, b]} \mid \hat{\Lambda}_n^{(0)}(x) - \Lambda(x) \mid \to 0,$$

almost surely. This consistency result will not be established here.

We now state the main result of this paper, which concerns the limiting behaviour of $2 \log \lambda_n$.

THEOREM 1. *Under Assumptions* A1–A9*, the pseudolikelihood ratio statistic,*

$$2 \log \lambda_n \equiv 2\{l_n^{\text{ps}}(\hat{\Lambda}_n \mid \mathcal{X}) - l_n^{\text{ps}}(\hat{\Lambda}_n^{(0)} \mid \mathcal{X})\} \to \frac{\sigma^2(t_0)}{\Lambda(t_0)} \mathbb{D},$$

in distribution, when $H_0 : \Lambda(t_0) = \theta_0$ *holds.*

A sketch proof of this theorem is given in the Appendix and uses the following theorem on the limit distribution of the nonparametric maximum likelihood estimators of Λ. Define

$$X_n(z) = n^{1/3}\left\{\hat{\Lambda}_n(t_0 + z\, n^{-1/3}) - \theta_0\right\}, \quad Y_n(z) = n^{1/3}\left\{\hat{\Lambda}_n^{(0)}(t_0 + z\, n^{-1/3}) - \theta_0\right\}.$$

THEOREM 2. *Suppose that Assumptions* A1–A9 *hold and set*

$$a = \left\{ \frac{\sigma^2(t_0)}{G'(t_0)} \right\}^{1/2}, \qquad b = \tfrac{1}{2}\Lambda'(t_0).$$

Then, under H_0,

$$(X_n(z), Y_n(z)) \to \left(g_{a,b}(z), g^0_{a,b}(z)\right),$$

in distribution, finite-dimensionally and also in the space $\mathcal{L} \times \mathcal{L}$, *where* $\mathcal{L}$ *is the space of functions from* $\mathbb{R} \to \mathbb{R}$ *that are bounded on every compact set, equipped with the topology of* L_2*-convergence with respect to Lebesgue measure on compact sets.*

For a sketch proof of this theorem, which can be established by extending the arguments in Theorem 4·3 of Wellner & Zhang (2000), or by continuous mapping arguments, see a longer version of this paper available from the authors.

Theorem 1 gives an easy way of constructing a likelihood-ratio-based confidence set for $F(t_0)$ in the mixed-case interval censoring model. This is based on the observation that, under the mixed-case interval censoring framework, where the counting process $N(t)$ is $1(S \leqslant t)$ with S following distribution F independently of (K, T), the pseudolikelihood-ratio statistic in Theorem 1 converges to $(1-\theta_0)\,\mathbb{D}$ under the null hypothesis $F(t_0) = \theta_0$. Thus, $(1-\theta_0)^{-1}\, 2 \log \lambda_n$ converges in distribution to $\mathbb{D}$, so that an asymptotic level-$(1-\alpha)$ confidence set for $F(t_0)$ is given by $\{\theta : (1-\theta)^{-1}\, 2 \log \lambda_n(\theta) \leqslant q(\mathbb{D}, 1-\alpha)\}$, where $q(\mathbb{D}, 1-\alpha)$ is the $(1-\alpha)$th quantile of $\mathbb{D}$ and $2 \log \lambda_n(\theta)$ is the pseudolikelihood-ratio statistic computed under the null hypothesis $H_{0,\theta} : F(t_0) = \theta$. Thus, finding the confidence set amounts to computing the likelihood ratio under a family of null hypotheses. The computation is a simple affair and can be done through using the elementary pool-adjacent-violators algorithm. Quantiles of $\mathbb{D}$ are tabulated in Banerjee & Wellner (2001).

Theorem 4·3 of Wellner & Zhang (2000) can also be derived as a special case of Theorem 2 by setting $z = 0$. Specialized to the mixed-case censoring scenario, it provides an alternative route to constructing confidence sets for $F(t_0)$. Denoting by $\hat{F}_n$ the pseudolikelihood estimate of F, from Theorem 4·3 of Wellner & Zhang (2000), we obtain

$$n^{1/3}\{\hat{F}_n(t_0) - F(t_0)\} \to \left\{ \frac{\theta_0\,(1-\theta_0)\, f(t_0)}{2\, G'(t_0)} \right\}^{1/3} 2\,\mathbb{Z}, \tag{5}$$

in distribution, where $\mathbb{Z} = \operatorname{argmin}_h \{W(h) + h^2\}$ and $f(t)$ is the derivative of $F(t)$. An approximate level-$(1-\alpha)$ confidence interval for $F(t_0)$ is

$$[\hat{F}_n(t_0) - 2\, C_n\, q(\mathbb{Z}, 1-\alpha/2)\ ,\ \hat{F}_n(t_0) + 2\, C_n\, q(\mathbb{Z}, 1-\alpha/2)],$$

where $q(\mathbb{Z}, 1-\alpha/2)$ is the $(1-\alpha/2)$th quantile of $\mathbb{Z}$ and

$$C_n = n^{-1/3} \left[\frac{\hat{F}_n(t_0)\,\{1 - \hat{F}_n(t_0)\}\, \hat{f}(t_0)}{2\, \hat{G}'(t_0)} \right]^{1/3},$$

with $\hat{f}$ and $\hat{G}'$ denoting estimators of f and G' respectively. Quantiles of $\mathbb{Z}$ are tabulated in Groeneboom & Wellner (2001). Estimating G' involves estimating first the probability density of K and then the marginal densities of the $T_{k,j}$'s; this can be done using kernel density methods with some optimal bandwidth selection procedure like least-squares crossvalidation (Loader, 1999). However, it is not difficult to see that, if K assumes a large number of values and the sample size n is moderate, there may not be sufficiently

many observations to estimate the density of each $T_{k,j}$ reliably. Finally there is also the problem of estimating f, which is a trickier affair, since observations from the distribution F are not available. A discussion of the issues involved in a similar situation can be found in Banerjee & Wellner (2005). In § 3, we estimate f by kernel smoothing the maximum pseudolikelihood estimator $\hat{F}_n$, as in Banerjee & Wellner (2005), using a likelihood-based crossvalidation criterion. The procedure followed is analogous to the one described in § 3·1 of that paper, the only difference being that the likelihood used for crossvalidation here is the pseudolikelihood, as opposed to the current status likelihood used in that paper.

Thus, the estimation of nuisance parameters turns out to be the major concern in the Wald-based approach: the variability introduced through nuisance parameter estimation will tend to make the confidence intervals much more unreliable, especially at smaller sample sizes. The likelihood-ratio-based method, on the other hand, does not involve nuisance parameter estimation and provides an extremely clear-cut way of constructing confidence intervals for $F(t_0)$. This makes it a much more attractive option. Yet another method of obtaining confidence sets is via the use of subsampling techniques. In view of the nonstandard asymptotics involved, as manifested in the cube-root convergence of the pseudolikelihood estimator to a nongaussian limit, the usual bootstrap is suspect, but subsampling without replacement works. Subsampling was implemented by drawing a large number of subsamples of size b from the original sample, without replacement, and estimating the limiting quantiles of $| n^{1/3}\{\hat{F}_n(t_0) - F(t_0)\} |$, using the empirical distribution of $| b^{1/3}\{\hat{F}_n^{\star}(t) - \hat{F}_n(t)\} |$; here $\hat{F}_n^{\star}(t)$ denotes the value of the maximum pseudolikelihood estimator, based on the subsample. For consistent estimation of the quantiles, b/n should converge to 0 as n increases. In the literature, b is referred to as the block-size. For details, see Politis et al. (1999, Ch. 2). The choice of b can affect the precision of the confidence intervals in finite samples. A data-driven choice of b is often resorted to but can be computationally very intensive. For a discussion of subsampling in the context of an interval censored model, see §§ 2 and 3 of Banerjee & Wellner (2005). Since the issues in the present case are similar, we do not go into an exhaustive discussion here.

The pseudolikelihood-based method for constructing confidence sets at a single point can be extended to finitely many points of interest; here the relevant limit distribution is the maximum of k independent copies of $\mathbb{D}$, where k is the number of points. However, the construction of likelihood-based simultaneous confidence bands for F is still an open problem.

3. Simulation studies and data analysis

3·1. *Simulation studies*

We present simulations from a mixed-case censoring model, in which the survival time distribution X was taken to follow the Ex(1) distribution. The random number K of observation times for an individual was generated from the uniform distribution on the integers $\{1, 2, 3, 4\}$ and, given $K = k$, the observation times $\{T_{k,i}\}_{i=1}^{k}$ were chosen as k order statistics from the uniform distribution on (0,3). We generated 1000 replicates for each sample size displayed in Table 1, and 95% confidence intervals for $F(\log 2) = 0{\cdot}5$ were computed by three different methods: pseudolikelihood ratio, limit distribution of the maximum pseudolikelihood estimator with kernel-based estimation of nuisance parameters, and subsampling with appropriate block-size. Kernel-based estimation was done in the way described in connection with the construction of confidence sets for

Table 1. *Simulation study for mixed-case interval censoring model: Average length* (AL) *and empirical coverage* (C) *of asymptotic* 95% *confidence intervals using pseudolikelihood-ratio* (PL), *maximum pseudolikelihood* (PMLE) *and subsampling-based* (SB) *methods.*

	PL		PMLE		SB	
n	AL	C	AL	C	AL	C
50	0·410	0·904	0·441	0·867	0·538	0·971
100	0·327	0·920	0·353	0·896	0·469	0·972
200	0·261	0·924	0·282	0·899	0·308	0·958
500	0·198	0·949	0·210	0·923	0·242	0·958
1000	0·157	0·938	0·167	0·914	0·174	0·945
1500	0·136	0·936	0·144	0·933	0·158	0·962
2000	0·124	0·943	0·131	0·921	0·144	0·965

$F(t_0)$ in § 2. For the subsampling-based intervals, we did not resort to a data-driven block-size selection algorithm, since this would have increased computational complexity by orders of magnitude. Since the data-generating process here is known, we generated separate datasets, 1000 replicates, from the mixed-case model for each sample size, and computed subsampling-based intervals for $F(t_0) = 0{\cdot}5$ using a selection of block-sizes. We then computed the empirical coverage of the 1000 confidence intervals produced for each block-size, and chose, as the optimal block-size for the simulations presented here, as the one for which the empirical coverage was closest to 0·95. Thus, block-size selection was done via pilot simulations. Of course, this is not feasible in a real-life setting, since the data generating process is unknown. A natural way to circumvent this problem for real datasets is to use the bootstrap to generate 'pilot data' from the empirical measure of the observed data and choose the block size based on the bootstrapped samples. This idea from Delgado et al. (2001) is used in § 3·2. The results are reported in Table 1.

From Table 1 we see that the pseudolikelihood method produces the narrowest confidence intervals on average. While they tend to be anticonservative, the coverage nevertheless is quite satisfactory, being greater than or close to 94%, provided the sample size is moderately large. The subsampling-based intervals are the widest, and not surprisingly conservative in general. The kernel-based intervals perform quite poorly at lower sample sizes, being extremely anticonservative but also giving wider confidence intervals than the likelihood ratio, and they remain anticonservative at higher sample sizes as well. The overall picture indicates the superiority of our pseudolikelihood-ratio method. This, added to the relative computational simplicity of our method in comparison to its competitors, where one needs to contend with the choice of a smoothing parameter or block-size, makes it an attractive choice.

In the longer version of the paper we present simulation results for examples of current status data and Case 2 interval censoring. Case 2 censoring is a special case of mixed-case censoring where the number of observation times for each individual is identically 2. The simulation results showed features similar to what was seen above.

3·2. *Illustration on a real dataset*

De Gruttola & Lagakos (1989) present an interval censored dataset of the time to HIV infection in a group of haemophiliacs. Since 1978, 262 people with Type A or B haemophilia had been treated at Hôpital Kremlin Bicêtre and Hôpital Coeur des Yvelines in France. Twenty-five of them were found to be infected on their first test for infection. By August 1988, 197 had become infected and 43 of these had developed some clinical symptoms relating to their HIV infection. All the infected persons are believed to have become infected by contaminated blood factor that they received for their haemophilia.

For each patient, the only information available is that $X \in [X_L, X_R]$, where X denotes the time to infection. Here time is measured in six-month intervals, with $X = 1$ denoting 1 July 1978. An individual was assigned $X_L = 1$ if they were found to be infected with HIV on their first test for infection. As mentioned above, there were 25 such individuals. For details see § 6 of De Gruttola & Lagakos (1989), and their Table 1, where the (X_L, X_R) values for each patient are provided. We are interested in estimating the distribution of X, the time to infection, based on the (X_L, X_R) pairs. We do the analysis separately for the two different groups into which the patients fell: the heavily-treated group of 105 patients received at least 1000 μg/kg of blood factor for at least one year between 1982 and 1985, and the lightly-treated group of 157 patients received less than 1000 μg/kg of blood factor per year.

We model the data as Case 2 censored data. The two censoring times U and V, with $U < V$, are defined as follows. If $1 = X_L < X_R < \infty$, we set $U = X_R$ and V to be the time till the end of the study. If $1 < X_L < X_R < \infty$, we set $U = X_L$ and $V = X_R$. If $1 < X_L < X_R = \infty$, we set $U = 1$ and $V = X_L$. If $(\Delta_1, \Delta_2, \Delta_3)$ denotes the vector of indicators, with $\Delta_1 = 1(X \leqslant U)$, $\Delta_2 = 1(U < X \leqslant V)$ and $\Delta_3 = 1(V \leqslant X)$, then for the first case this vector is $(1, 0, 0)$, for the second case it is $(0, 1, 0)$ and in the third case it is $(0, 0, 1)$. The given dataset is really an example of mixed-case censoring in which only the relevant inspection times have been noted. The formulation of the problem as a Case 2 model is a simplification that we adopt for the purpose of illustrating our method; because of lack of information about the other inspection times, the full mixed-case model cannot be fitted to the data.

The pseudolikelihood estimate of F, the distribution function of X, was computed for each of the two groups, and confidence intervals for the values of F at several different points were obtained using the three different methods illustrated in the simulation studies. The subsampling-based confidence interval at any given point was computed by first determining the block-size b using the bootstrap-based block selection algorithm referred to in § 3·1; see Banerjee & Wellner (2005) for a brief description and an application of this algorithm to current status data. Five hundred bootstrap samples were used for block-size selection, and once the optimal block size had been ascertained 1000 subsamples of that size were used to determine the confidence interval. As far as the estimation of nuisance parameters for the construction of the Wald-type confidence interval was concerned, $f(t_0)$ at a point of interest t_0 was computed by smoothing the maximum pseudolikelihood estimator using bandwidth determined by likelihood-based crossvalidation, as for the simulation experiments. However, least-squares crossvalidation, for choosing the optimal bandwidths to estimate $G'(t_0)$, did not perform well, and therefore G' was estimated by differentiating the piecewise-linear modification of the empirical distribution functions of U and V.

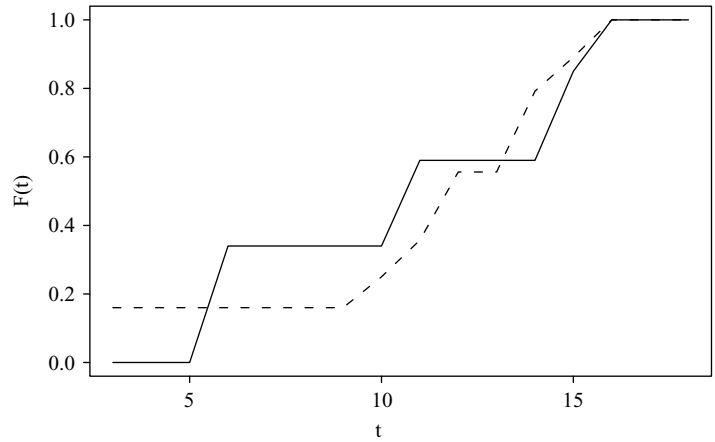

Fig. 1. HIV infection data. The estimated distribution functions of time to HIV infection in the two different groups: heavily treated, solid line; lightly treated, dashed line.

The estimated distribution functions of the time to infection are plotted for the two different groups in Fig. 1. The distribution function for the heavily-treated group dominates that for the lightly-treated group in the interval [6, 14); between 14 and 16, the distribution function for the lightly-treated group is higher; at 16, the two distributions coincide at the value 1. Individuals in the heavily-treated group received higher amounts of blood factor for at least a year between 1982 and 1985; the higher the amount of blood transfusion, the greater is the chance of infection through contaminated blood factor. The date of 1 July 1982 corresponds to $t = 9$, and $t = 16$, where the two distribution functions coincide, corresponds to 1 January 1986. In the range $9 - 16$, the distribution function for the heavily treated group is either equal to or almost equal to that for the lightly treated group or dominates it, except in the range [14, 15); this corresponds to the year 1985.

Table 2 gives confidence intervals at different time points obtained by the three different methods. The second column gives the value of the maximum pseudolikelihood estimator, the third gives the confidence intervals using the pseudolikelihood ratio, the fourth gives the Wald-type intervals and the fifth gives the subsampling-based intervals. Note that the left extremities of the confidence intervals for the distribution function in the heavily-treated group are generally shifted to the right of those for the corresponding time points in the lightly-treated group, with violations for large t. The general shift of the left extremities to the right is predictable. The violation of this property for large t is not surprising, since there we are dealing with the time range in which the two distribution functions are essentially 'catching up' with each other, as is evident from Fig. 1. Also note that the likelihood-ratio-based confidence intervals are somewhat less erratic than the two other intervals; they exhibit monotonicity of left as well as right endpoints with increasing t. Since F is monotone in t, this is a rather nice property. On the other hand, the Wald-type or the subsampling-based intervals tend to exhibit violations of this property, though there is an overall monotonic trend.

A beta version of the software employed for the simulations and data analysis is available from the authors, and will be developed into an R-package in the near future. The data are available in De Gruttola & Lagakos (1989) and also from the authors.

Table 2. HIV *infection data. Confidence intervals* (CI) *of three kinds for the distribution of time to* HIV *infection in the lightly and heavily treated groups at different times: likelihood-ratio-based* (*lrt*)*, Wald-type* (*Wald*) *and subsampling-based* (*subsampling*).

t	$\hat{F}(t)$	CI (lrt)	CI (Wald)	CI (subsampling)
		Lightly treated group		
6·0	0·160	0·068–0·285	0·000–0·354	0·068–0·252
7·0	0·160	0·068–0·285	0·000–0·409	0·046–0·274
8·0	0·160	0·068–0·298	0·000–0·410	0·042–0·278
9·0	0·160	0·069–0·321	0·000–0·379	0·026–0·294
10·0	0·250	0·069–0·458	0·000–0·463	0·048–0·451
11·0	0·357	0·099–0·546	0·000–0·623	0·174–0·540
12·0	0·556	0·187–0·660	0·381–0·730	0·396–0·716
13·0	0·556	0·402–0·700	0·277–0·834	0·361–0·750
14·0	0·792	0·439–0·888	0·553–1·000	0·660–0·923
15·0	0·891	0·637–0·943	0·712–1·000	0·786–0·996
		Heavily treated group		
6·0	0·340	0·000–0·442	0·067–0·613	0·087–0·593
7·0	0·340	0·092–0·442	0·171–0·509	0·220–0·459
8·0	0·340	0·240–0·442	0·113–0·567	0·184–0·496
9·0	0·340	0·240–0·442	0·179–0·501	0·213–0·467
10·0	0·340	0·240–0·451	0·206–0·474	0·213–0·467
11·0	0·588	0·242–0·665	0·437–0·739	0·459–0·717
12·0	0·588	0·472–0·665	0·451–0·725	0·490–0·686
13·0	0·588	0·484–0·673	0·462–0·715	0·496–0·680
14·0	0·588	0·504–0·676	0·450–0·727	0·478–0·699
15·0	0·852	0·504–0·927	0·751–0·953	0·740–0·964

Acknowledgement

M. Banerjee's research was partially supported by a grant from the U.S. National Science Foundation.

Appendix

Technical details

We first formally state the required assumptions.

Assumption A1. The observation times $T_{k,j}$, for $j = 1, \ldots, k$ and $k = 1, 2, \ldots$, are random variables taking values in the bounded set $[0, \tau]$, where $0 < \tau < \infty$ and $E(K) < \infty$.

Assumption A2. The mean function Λ satisfies $\Lambda(\tau) \leqslant M$ for some $0 < M < \infty$.

Assumption A3. The random variable M_0, defined as $M_0 = \sum_{j=1}^{K} N_{K,j} \log N_{K,j}$, satisfies $E(M_0) < \infty$. Here, interpret 0 log 0 as 0.

For Borel subsets B of $[0, \tau]$, define the measure μ as

$$\mu(B) = E\left\{\sum_{j=1}^{K} 1\,(T_{K,j} \in B)\right\}.$$

Let $G(t) \equiv \mu((0,t])$ be the distribution function corresponding to the measure μ. For each k, j, denote the distribution function of the random variable $T_{k,j}$ by $G_{k,j}$. Then

$$G(t) = E\left(\sum_{j=1}^{K} 1\left\{T_{K,j} \leqslant t\right\}\right) = \sum_{k=1}^{\infty} \operatorname{pr}(K=k) \sum_{j=1}^{k} \operatorname{pr}\left(T_{k,j} \leqslant t | K=k\right)$$
$$= \sum_{k=1}^{\infty} \operatorname{pr}(K=k) \sum_{j=1}^{k} G_{k,j}(t). \tag{A1}$$

Call x a support point of μ if, for every $\epsilon > 0$, it is the case that $\mu(x-\epsilon, x+\epsilon) > 0$. Let S_μ denote the set of all support points of μ.

Assumption A4. The point t_0 lies in the interior of S_μ.

Assumption A5 (a). The variable K has a finite moment of order greater than 2.

Assumption A5 (b). There exist $\alpha > 0$ and $M_1 > 0$ such that $E\{N^{2+\alpha}(t)\} \leqslant M_1$ for all $t \in S_\mu$.

Assumption A6. There is a neighbourhood $\mathcal{U}$ of $t_0 \in S_\mu$ such that the distribution functions $G_{k,j}$ have positive continuous derivatives on $\mathcal{U}$, which are bounded by a common constant B for all k, j.

Assumption A7. There is a neighbourhood $\mathcal{V}$ of $(t_0, t_0) \in \mathbb{R}^2$ such that, for all $k = 1, 2, \ldots$ and $1 \leqslant i \leqslant j \leqslant k$, $G_{k,i,j}(s,t) = \operatorname{pr}(T_{k,i} \leqslant s, T_{k,j} \leqslant t)$ is differentiable with respect to (s,t) and $g_{k,i,j}(s,t) = \partial^2 G_{k,i,j}(s,t)/\partial s\, \partial t$ exists. Furthermore, the functions $g_{k,i,j}$ are bounded on $\mathcal{V}$, by a common constant C, for all (k, i, j).

Assumption A8. The mean function Λ has a continuous bounded derivative on $\mathcal{U}$ and $\Lambda'(t_0) \neq 0$.

Assumption A9. The function $\sigma^2(t) \equiv \operatorname{var}\{N(t)\}$ is continuous in a neighbourhood of t_0.

We discuss the implications of our assumptions in the interval censoring framework. Assumption A2 is trivially satisfied in the interval censoring situation, since $0 \leqslant F(t) = \Lambda(t) \leqslant 1$. Assumption A3 is also easy to check; in the interval censored situation, $N_{k,j}$ is either 1 or 0, so that $M_0 = 0$. In so far as estimation at the point t_0 is concerned, it suffices to have a positive Lebesgue density for one of the $T_{k,j}$'s in a neighbourhood of the point t_0, along with $\operatorname{pr}(K=k) > 0$, for Assumption A4 to be satisfied. Assumption A5 is guaranteed for a K that is finitely supported, which is typically the case in applications, and for the interval censoring situation, since $N(t) \leqslant 1$. Assumption A8, in the interval censoring scenario, translates to $F(t)$ being continuously differentiable in a neighbourhood of t_0 with $f(t_0) \neq 0$. Finally, Assumption A9 is easily satisfied, since $\sigma^2(t) = F(t)\{1 - F(t)\}$.

We first define the following processes:

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

Thus, both V_n and G_n are piecewise-constant right-continuous processes, with possible jumps only at the distinct observation times; the jump of G_n at the point $T_{(l)}$ is simply w_l/n, whereas the jump

of V_n at the same point is $w_l\,\bar{N}_l/n$. Also, set

$$\xi_1(X,t) = \sum_{j=1}^{K} N_{K,j}\,1\left\{T_{K,j} \leqslant t\right\}, \qquad \xi_0(X,t) = \sum_{j=1}^{K} 1\left\{T_{K,j} \leqslant t\right\}.$$

Note that $G(t) = E\{\xi_0(X,t)\}$. Also, define $V(t) = E\{\xi_1(X,t)\}$. From (A1) we obtain that $G'(t) = \sum_{k=1}^{\infty} \mathrm{pr}(K=k)\sum_{j=1}^{k} G'_{k,j}(t)$. Also,

$$\begin{aligned} V(t) &= E\left(\sum_{j=1}^{K} N_{K,j}\,1\left\{T_{K,j} \leqslant t\right\}\right) = \sum_{k=1}^{\infty} \mathrm{pr}\,(K=k) \sum_{j=1}^{k} E\left(N_{k,j}\,1\{T_{k,j} \leqslant t\}\right) \\ &= \sum_{k=1}^{\infty} \mathrm{pr}\,(K=k) \sum_{j=1}^{k} \int_0^t \Lambda(x)\,dG_{k,j}(x) \\ &= \int_0^t \Lambda(x) \left\{\sum_{k=1}^{\infty} \mathrm{pr}\,(K=k) \sum_{j=1}^{k} G'_{k,j}(x)\right\} dx = \int_0^t \Lambda(x)\,G'(x)\,dx, \end{aligned}$$

whence

$$V'(t) = \Lambda(t)\,G'(t). \tag{A2}$$

Proof of Theorem 1. In the following derivation, we denote by $\hat{\Lambda}_l$ the value of the unconstrained estimator $\hat{\Lambda}_n$ at the point $T_{(l)}$, and by $\hat{\Lambda}_l^{(0)}$ the value of $\hat{\Lambda}_n^{(0)}$ at the point $T_{(l)}$. The likelihood ratio statistic is then given by

$$\begin{aligned} 2\log\lambda_n &= 2\sum_{l=1}^{M} w_l\,(\bar{N}_l \log\hat{\Lambda}_l - \hat{\Lambda}_l) - 2\sum_{l=1}^{M} w_l\,(\bar{N}_l \log\hat{\Lambda}_l^{(0)} - \hat{\Lambda}_l^{(0)}) \\ &= 2\sum_{l=1}^{M} w_l\,\bar{N}_l\,(\log\hat{\Lambda}_l - \log\hat{\Lambda}_l^{(0)}) - 2\sum_{l=1}^{M} w_l\,(\hat{\Lambda}_l - \hat{\Lambda}_l^{(0)}). \end{aligned}$$

In what follows, we assume that the null hypothesis holds, so that $\Lambda(t_0) \equiv \theta_0$. We will also denote the set of indices for which $\hat{\Lambda}_l$ differs from $\hat{\Lambda}_l^{(0)}$ by D. On Taylor expansion of $\log\hat{\Lambda}_l$ and $\log\hat{\Lambda}_l^{(0)}$ around θ_0, we obtain

$$\begin{aligned} 2\log\lambda_n = 2\sum_{l\in D} w_l\,\bar{N}_l \Bigg\{&\log\theta_0 + \frac{1}{\theta_0}(\hat{\Lambda}_l - \theta_0) - \frac{1}{2\theta_0^2}(\hat{\Lambda}_l - \theta_0)^2 + \frac{1}{3\,\Lambda_{l,\star}^3}(\hat{\Lambda}_l - \theta_0)^3 \\ &- \log\theta_0 - \frac{1}{\theta_0}(\hat{\Lambda}_l^{(0)} - \theta_0) + \frac{1}{2\theta_0^2}(\hat{\Lambda}_l^{(0)} - \theta_0)^2 - \frac{1}{3\,\Lambda_{l,\star\star}^3}(\hat{\Lambda}_l - \theta_0)^3\Bigg\} \\ &- 2\sum_{l\in D} w_l\,(\hat{\Lambda}_l - \hat{\Lambda}_l^{(0)}). \end{aligned}$$

Here $\hat{\Lambda}_{l,\star}$ is a point intermediate between $\hat{\Lambda}_l$ and θ_0, and $\hat{\Lambda}_{l,\star\star}$ is a point intermediate between $\hat{\Lambda}_l^{(0)}$ and θ_0. The above expression simplifies to

$$\begin{aligned} 2\log\lambda_n = 2\sum_{l\in D} w_l\,\bar{N}_l\,\frac{1}{\theta_0}\left\{(\hat{\Lambda}_l - \theta_0) - (\hat{\Lambda}_l^{(0)} - \theta_0)\right\} - 2\sum_{l\in D} w_l\left\{(\hat{\Lambda}_l - \theta_0) - (\hat{\Lambda}_l^{(0)} - \theta_0)\right\} \\ -\frac{1}{\theta_0^2}\sum_{l\in D}\left\{(\hat{\Lambda}_l - \theta_0)^2 - (\hat{\Lambda}_l^{(0)} - \theta_0)^2\right\} w_l\,\bar{N}_l + o_p(1), \end{aligned}$$

whence $2\log\lambda_n = T_1 - T_2 + o_p(1)$ with

$$T_1 = 2\sum_{l\in D} w_l\,\bar{N}_l\,\frac{1}{\theta_0}\left\{(\hat{\Lambda}_l - \theta_0) - (\hat{\Lambda}_l^{(0)} - \theta_0)\right\} - 2\sum_{l\in D} w_l\left\{(\hat{\Lambda}_l - \theta_0) - (\hat{\Lambda}_l^{(0)} - \theta_0)\right\},$$

and $T_2 = \frac{1}{\theta_0^2} \sum_{l \in D} \left\{ (\hat{\Lambda}_l - \theta_0)^2 - (\hat{\Lambda}_l^{(0)} - \theta_0)^2 \right\} w_l \bar{N}_l$. Consider T_2. Noting that $n^{-1} w_l \bar{N}_l$ is the jump of the right-continuous process V_n at the point $T_{(l)}$, letting D_n denote the set on which $\hat{\Lambda}_n$ and $\hat{\Lambda}_n^{(0)}$ differ and setting $\tilde{D}_n$ to be the set $n^{1/3}(D_n - t_0)$, which is an interval and can be easily shown to be $O_p(1)$, we can write

$$\begin{aligned} T_2 &= \frac{1}{\theta_0^2} n \int_{D_n} \left[\{\hat{\Lambda}_n(t) - \theta_0\}^2 - \{\hat{\Lambda}_n^{(0)}(t) - \theta_0\}^2 \right] d\, V_n(t) \\ &= \frac{1}{\theta_0^2} n \int_{D_n} \left[\{\hat{\Lambda}_n(t) - \theta_0\}^2 - \{\hat{\Lambda}_n^{(0)}(t) - \theta_0\}^2 \right] d\, V(t) + o_p(1) \qquad \text{(A3)} \\ &= \frac{1}{\theta_0^2} \int_{\tilde{D}_n} \Big[n^{2/3} \{\hat{\Lambda}_n(t_0 + n^{-1/3} z) - \theta_0\}^2 \\ &\quad - n^{2/3} \{\hat{\Lambda}_n^{(0)}(t_0 + n^{-1/3} z) - \theta_0\}^2 \Big] V'(t_0 + n^{-1/3} z)\, dz + o_p(1) \\ &= \frac{V'(t_0)}{\theta_0^2} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz + o_p(1)\,, \end{aligned}$$

where (A3) follows from the step above it, with V_n replaced by V, by a standard empirical process argument. Now, consider T_1. If we use the definitions of the processes V_n and G_n, it is straightforward to see that

$$\begin{aligned} T_1 &= \frac{2}{\theta_0} n \int_{D_n} \left[\{\hat{\Lambda}_n(t) - \theta_0\} - \{\hat{\Lambda}_n^{(0)}(t) - \theta_0\} \right] d\, \{V_n(t) - \theta_0\, G_n(t)\} \qquad \text{(A4)} \\ &= \frac{2}{\theta_0} n \int_{D_n} \left[\{\hat{\Lambda}_n(t) - \theta_0\}^2 - \{\hat{\Lambda}_n^{(0)}(t) - \theta_0\}^2 \right] d\, G_n(t) \qquad \text{(A5)} \\ &= \frac{2}{\theta_0} n \int_{D_n} \left[\{\hat{\Lambda}_n(t) - \theta_0\}^2 - \{\hat{\Lambda}_n^{(0)}(t) - \theta_0\}^2 \right] G'(t)\, dt + o_p(1) \qquad \text{(A6)} \\ &= \frac{2}{\theta_0} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} G'(t_0 + n^{-1/3} z)\, dz + o_p(1) \qquad \text{(A7)} \\ &= \frac{2\, G'(t_0)}{\theta_0} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz + o_p(1)\,, \end{aligned}$$

where (A5) follows from the characterization of the nonparametric maximum likelihood estimators in terms of the processes G_n and V_n and will be justified at the end, (A6) follows from (A5) with $dG_n(t)$ replaced by $d\, G(t) \equiv G'(t)\, dt$ using standard empirical process arguments, and (A7) follows if we transform to the local variable z and use the definitions of the processes X_n and Y_n. Thus,

$$2 \log \lambda_n = \frac{2\, G'(t_0)}{\theta_0} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz - \frac{V'(t_0)}{\theta_0^2} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz + o_p(1)\,.$$

Recalling that $a^2 = \sigma^2(t_0)/G'(t_0)$ from the statement of Theorem 2 and that $V'(t_0) = \Lambda(t_0)\, G'(t_0)$ from equation (A2), so that $\theta_0^2 V'(t_0) = \theta_0^{-1} G'(t_0)$, we have

$$\begin{aligned} 2 \log \lambda_n &= \frac{G'(t_0)}{\theta_0} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz = \frac{\sigma^2(t_0)}{\Lambda(t_0)} a^{-2} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz \\ &\to \frac{\sigma^2(t_0)}{\Lambda(t_0)} a^{-2} \int \left[\left\{ g_{a,b}(z) \right\}^2 - \left\{ g_{a,b}^0(z) \right\}^2 \right] dz\,, \qquad \text{(A8)} \end{aligned}$$

in distribution. Here (A8) follows from the previous step by application of Theorem 2 in conjunction with the continous mapping theorem for distributional convergence and the fact that $(f, g) \mapsto \int (f^2 - g^2)\, d\lambda$, with λ denoting Lebesgue measure, is a continuous function from

$\mathcal{L} \times \mathcal{L}$ to $\mathbb{R}$. However, $\{\sigma^2(t_0)/\Lambda(t_0)\}\, a^{-2} \int [\{g_{a,b}(z)\}^2 - \{g^0_{a,b}(z)\}^2]\, dz$ has the same distribution as $\{\sigma^2(t_0)/\Lambda(t_0)\}\, \mathbb{D}$, by Lemma 1. If in particular $N(t)$ is indeed a Poisson process, nonhomogeneous or otherwise, $\sigma^2(t_0) = \Lambda(t_0)$ and the limiting distribution is exactly $\mathbb{D}$.

It only remains to justify going from (A4) to (A5). It suffices to show that

$$d_1 \equiv \int_{D_n} \{\hat{\Lambda}_n(t) - \theta_0\}\, d\{V_n(t) - \theta_0\, G_n(t)\} = \int_{D_n} \{\hat{\Lambda}_n(t) - \theta_0\}^2\, d\, G_n(t),$$

$$d_2 \equiv \int_{D_n} \{\hat{\Lambda}^0_n(t) - \theta_0\}\, d\{V_n(t) - \theta_0\, G_n(t)\} = \int_{D_n} \{\hat{\Lambda}^0_n(t) - \theta_0\}^2\, d\, G_n(t)\,.$$

We will only show the latter. Let J_n denote the set of indices i such that $T_{(i)}$ belongs to D_n, ordered from smallest to largest. Partition J_n into consecutive blocks of indices $B_1, \ldots, B_k$ such that, on each B_j, we have that $\hat{\Lambda}^{(0)}_n(T_{(i)})$ is constant for all $i \in B_j$. Denote the constant value on B_j by v_j. There is potentially one block B_l on which $\hat{\Lambda}^{(0)}_n$ is equal to θ_0. On every other B_j, we have

$$v_j = \frac{n^{-1} \sum_{m \in B_j} w_m\, \bar{N}_m}{n^{-1} \sum_{m \in B_j} w_m} \equiv \frac{\sum_{m \in B_j} w_m\, \bar{N}_m}{\sum_{m \in B_j} w_m}\,.$$

This is an easy consequence of the characterization of the constrained solution. We can now write

$$\begin{aligned}
d_2 &= \sum_{j \neq l} \sum_{i \in B_j} \{\hat{\Lambda}^{(0)}_n(T_{(i)}) - \theta_0\}\, (n^{-1}\, w_i\, \bar{N}_i - \theta_0\, n^{-1}\, w_i) \\
&= \sum_{j \neq l} (v_j - \theta_0) \left(\sum_{i \in B_j} \frac{w_i\, \bar{N}_i}{n} - \theta_0 \sum_{i \in B_j} \frac{w_i}{n} \right) \\
&= \sum_{j \neq l} (v_j - \theta_0) \left(\sum_{i \in B_j} \frac{w_i}{n} \right) \left(\frac{\sum_{i \in B_j} w_i\, \bar{N}_i}{\sum_{i \in B_j} w_i} - \theta_0 \right) \\
&= \sum_{j \neq l} (v_j - \theta_0)^2 \sum_{i \in B_j} \frac{w_i}{n} = \sum_{j \neq l} \sum_{i \in B_j} \{\hat{\Lambda}^{(0)}_n(T_{(i)}) - \theta_0\}^2\, n^{-1}\, w_i \\
&= \int_{D_n} \{\hat{\Lambda}^0_n(t) - \theta_0\}^2\, d\, G_n(t)\,.
\end{aligned}$$

□

References

Banerjee, M. & Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699–731.

Banerjee, M. & Wellner, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32**, 405–24.

De Gruttola, V. & Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* **45**, 1–11.

Delgado, M. A., Rodriguez-poo, J. & Wolf, M. (2001). Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. *Econ. Lett.* **73**, 241–50.

Groeneboom, P. (1996). Inverse problems in statistics. Proceedings of the St. Flour Summer School in Probability, *Lecture Notes in Mathematics 1648*, Ed. P. Bernard, pp. 67–164, New York: Springer Verlag.

Groeneboom, P. & Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Boston, MA: Birkhäuser.

Groeneboom, P. & Wellner, J. A. (2001). Computing Chernoff's distribution. *J. Comp. Graph. Statist.* **10**, 388–400.

Jewell, N. P. & van der Laan, M. J. (1995). Generalizations of current status data with applications. *Lifetime Data Anal.* **1**, 101–9.

Jewell, N. P., van der Laan, M. J. & Henneman, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika* **90**, 183–97.

JONGBLOED, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comp. Graph. Statist.* **7**, 310–21.

LOADER, C. (1999). Bandwidth selection: classical or plug-in? *Ann. Statist.* **27**, 415–38.

POLITIS, D. N., ROMANO, J. P. & WOLF, M. (1999). *Subsampling*, New York: Springer Verlag.

ROBERTSON, T., WRIGHT, F. T. & DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. New York: Wiley.

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

SHIBOSKI, S. (1998). Generalized additive models for current status data. *Lifetime Data Anal.* **4**, 29–50.

SONG, S. (2004). Estimation with univariate "mixed case" interval censored data. *Statist. Sinica.* **14**, 269–82.

SUN, J. & KALBFLEISCH, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica.* **5**, 279–90.

VANDAL, A. C., GENTLEMAN, R. & LIU, X. (2005). Constrained estimation and likelihood intervals for censored data. *Can. J. Statist.* **33**, 71–84.

WELLNER, J. A. (2003). Gaussian white noise models: some results for monotone functions. In *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, IMS Lecture Notes-Monograph Series, Vol 43, Ed. J. E. Kolassa and D. Oakes, pp. 87–104, Haywood, CA: Institute of Mathematical Statistics.

WELLNER, J. A. & ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779–814.

[*Received June 2005. Revised May 2006*]