doi: 10.1111/j.1467-9469.2007.00568.x

Inconsistency of the MLE for the Joint Distribution of Interval-Censored Survival Times and Continuous Marks

M. H. MAATHUIS and J. A. WELLNER

2 *Department of Statistics, University of Washington*

ABSTRACT. This paper considers the non-parametric maximum-likelihood estimator (MLE) for the joint distribution function of an interval-censored survival time and a continuous mark variable. We provide a new explicit formula for the MLE in this problem. We use this formula and the mark-specific cumulative hazard function of Huang & Louis (1998) to obtain the almost sure limit of the MLE. This result leads to necessary and sufficient conditions for consistency of the MLE, which imply that the MLE is inconsistent in general. We show that the inconsistency can be repaired by discretizing the marks. Our theoretical results are supported by simulations.

Key words: inconsistency, interval censoring, mark variable, non-parametric maximum likelihood, survival analysis

1. Introduction

Suppose that X is a survival time and Y is a continuous mark variable that may be correlated with X. Huang & Louis (1998) considered non-parametric estimation of the joint distribution of X and Y when X is subject to (random) right-censoring and the mark variable Y is observed if and only if X is uncensored. In many cases of interest, however, we can only observe an interval-censored version of the random variable X. For example, Hudgens, Maathuis & Gilbert (2007; henceforth HMG) analyzed an HIV vaccine trial in which X is the time of HIV infection and Y is a measure of the genetic distance between the infecting HIV virus and the virus in the vaccine. The participants of this trial were tested for HIV at several follow-up times. As a result, X was interval censored, that is, only known to be in a time interval determined by the follow-up times. Moreover, since the viral distance Y could only be determined for HIV positive individuals, Y was missing for all individuals who were HIV negative at their last follow-up visit.

Motivated by this example we consider the following model, which we refer to as the
1 'interval-censoredcontinuous-mark model'. Let $X > 0$ be a survival time and let $Y \in \mathbb{R}$ be a continuous mark variable. For a fixed integer $k \geq 1$, suppose that $\mathbf{T} = (T_1, \ldots, T_k)$ is a vector of observation times with distribution G. We assume that $0 < T_1 < \cdots < T_k$ and that $\mathbf{T}$ is independent of (X, Y). We cannot observe (X, Y) directly. Instead, our observed data are $W = (\mathbf{T}, \boldsymbol{\Delta}, Z)$, where

$$\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_{k+1}) \quad \text{with} \quad \Delta_j \equiv 1\{T_{j-1} < X \leq T_j\}, \quad j = 1, \ldots, k+1,$$

(with the convention that $T_0 \equiv 0$ and $T_{k+1} \equiv \infty$), and

$$Z = \Delta_+ Y \quad \text{with} \quad \Delta_+ \equiv \sum_{j=1}^{k} \Delta_j.$$

Note that the vectors $\mathbf{T}$ and $\boldsymbol{\Delta}$ determine a time interval $(T_{j-1}, T_j]$, $j = 1, \ldots, k+1$, that is known to contain the survival time X. The variable Z reflects that the mark variable Y is observed if and only if the survival endpoint is reached before the last observation time, i.e. if and only if $X \leq T_k$.

Our censoring model for X is called 'interval censoring case k', since each individual in the study has exactly k observation times $T_1,\ldots,T_k$ (see Groeneboom & Wellner, 1992, for case 1 and case 2 interval censoring, and Wellner, 1995, for case k interval censoring). Interval-censoring case 1 is also referred to as 'current status censoring', since we only observe the 'current status' of an individual at a single observation time. A model that allows the number of observation times to be random, and hence to vary across individuals in the study, is called 'mixed-case interval censoring' (see, for example, Schick & Yu, 2000; Van der Vaart & Wellner, 2000; Sun, 2006, page 12).

Our goal here is to study the non-parametric maximum-likelihood estimator (MLE) of the joint distribution F_0 of (X, Y) when the observations consist of $W_1,\ldots,W_n$ i.i.d. as W. In particular we focus on consistency issues, and we show, in fact, that the MLE is inconsistent in general.

There are several known examples of inconsistency of the non-parametric MLE. Barlow *et al.* (1972. pp. 255–258) showed that the MLE $\hat{F}_n$ for the class of star-shaped distributions (distributions on $[0, b)$ with $F(0)=0$ and $F(x)/x$ non-decreasing) is inconsistent, by showing that for sampling from the uniform distribution on $[0, 1]$ the MLE $\hat{F}_n(x) \rightarrow_{a.s.} x^2$. For distributions F with increasing failure rate average (IFRA), Boyles *et al.* (1985) showed that the MLE is inconsistent, and they identified the limit explicitly for sampling from a general continuous distribution function F. In the context of bivariate right-censored data, inconsistency of the non-parametric MLE for continuous bivariate distributions was pointed out by Tsai *et al.* (1986) and was also studied by Van der Laan (1996). For estimation of a distribution function on $\mathbb{R}$ based on left-truncated and case 1 interval-censored data, Pan & Chappell (1999) showed that the non-parametric MLE is inconsistent. Finally, Maathuis (2003, section 6.2) showed inconsistency of the MLE of the bivariate distribution of (X, Y) when X is subject to current status censoring and Y is observed exactly.

There are many more examples of inconsistent maximum likelihood estimators in parametric problems: see, for example, Neyman & Scott (1948), Bahadur (1958), Ferguson (1982), Ghosh & Yang (1995), Gupta *et al.* (1999), and the interesting review by Le Cam (1990).

To relate our inconsistency result to some of these earlier studies of inconsistency of the MLE, note that observation of W instead of (X, Y) can be regarded as observation of a (random) set A known to contain the unobservable (X, Y). We call such a set an *observed set*. In our model the observed sets can take two forms. When $\Delta_j=1$ for some $j \leq k$ (so $\Delta_+=1$), the observed set is a horizontal line segment:

$$A=(T_{j-1}, T_j] \times \{Z\}, \tag{1}$$

while when $\Delta_{k+1}=1$, or equivalently, when $\Delta_+=0$, the observed set is a half plane:

$$A=(T_k, \infty) \times \mathbb{R}. \tag{2}$$

The line segments that arise when $\Delta_+=1$ are an indicator of potential consistency problems for the MLE, since such line segments also occurred in the inconsistent MLEs studied by Van der Laan (1996) and Maathuis (2003, section 6.2). This prompted us to carefully study consistency of the MLE for interval-censored continuous-mark data.

Our work is also related to the classical competing-risks model, in which one studies the failure time X of a system that can fail from a (finite) number of J competing risks given by values of $Y \in \{1,\ldots,J\}$. The variable Y in this model can only be observed after the failure event happened, and is therefore a mark variable. Thus, the classical competing-risks model can be called a 'discrete-mark model', and can be viewed as the discrete counterpart of the continuous-mark model. The competing-risks model has been studied under various censoring assumptions for X. Aalen (1976, 1978) and Kalbfleisch & Prentice (1980, section 7.2, pp. 163–178) studied the MLE in this model when X is subject to right censoring. The generalization to interval-censored survival data with competing risks was considered by Hudgens *et al.* (2001) and Jewell *et al.* (2003). Jewell & Kalbfleisch (2004) studied computational issues of the MLE for current status data with competing risks, and Maathuis (2006) and Groeneboom *et al.* (2006a, 2006b) derived the asymptotic properties of the MLE in this model.

In the current paper we focus on the interval-censored continuous-mark model. In section 2 we derive a new formula for the MLE in this model, using connections with univariate right censored data. In section 3 we use this new formula and the mark-specific cumulative hazard function of Huang & Louis (1998) to derive the almost sure limit of the MLE. This result leads to necessary and sufficient conditions for consistency of the MLE that force a relation between the unknown distribution F_0 and the observation time distribution G. Since such a relation will typically not hold, it follows that the MLE is inconsistent in general. In section 4 we show that the inconsistency can be repaired by discretizing the marks, an operation that transforms the data into interval-censored competing-risks data. In section 5 we support our theoretical results by simulations of the MLE and the repaired MLE. Section 6 contains a discussion of some remaining issues. Technical proofs are collected in the appendix.

2. Explicit formula for the MLE

HMG noted a close connection between the MLE for univariate right-censored data and the MLE for interval-censored continuous-mark data. We use this connection in section 2.2 to derive a new explicit formula for the MLE for interval-censored continuous-mark data. But first, in section 2.1, we review univariate right-censored data in a way that shows the similarity between the two models.

2.1. Intermezzo: univariate right-censored data

Suppose that we want to estimate the distribution F_0 of a survival time X, and suppose that X is subject to right censoring. Thus, instead of n i.i.d. copies of X, we observe n i.i.d. copies of $(\min(X,T), 1\{X \le T\})$, where T is a random censoring time with distribution G. We assume that T is independent of X. It is well-known that the MLE $\hat{F}_n$ of F_0 in this model is given by the Kaplan–Meier estimator.

We now review the Kaplan–Meier estimator in a way that allows us to easily make a connection with interval-censored continuous-mark data. We first introduce some notation and terminology. Define $U \equiv \min(X,T)$ and $\Delta \equiv 1\{X \le T\}$, and let $(U_1,\Delta_1),\dots,(U_n,\Delta_n)$ denote n i.i.d. copies of (U,Δ). Recalling the discussion of observed sets in Section 1, each observation (U,Δ) defines an observed set A that is known to contain X: $A=\{U\}$ if $\Delta=1$, and $A=(U,\infty)$ if $\Delta=0$. Let $U_{(1)},\dots,U_{(n)}$ be the order statistics of $U_1,\dots,U_n$, and let $\Delta_{(i)}$ and $A_{(i)}$ be the corresponding values of Δ and A. We assume that all A_i with $\Delta_i=1$ are distinct, since this will be the case for the continuous-mark data. However, we allow ties in the Ts and Us provided that this assumption is not violated. We break such ties in U arbitrarily after ensuring that observations with $\Delta=1$ are ordered before those with $\Delta=0$.

By assuming that F has a density f with respect to some dominating measure μ, the likelihood (up to multiplicative terms depending only on G) is $L_n(F)=\Pi_{i=1}^n q(U_i,\Delta_i)$, where $q(u,\delta)=f(u)^{\delta}\{1-F(u)\}^{1-\delta}$. Since the first term of q is a density-type term, $L_n(F)$ can be made arbitrarily large by letting f peak at some value U_i with $\Delta_i=1$. This problem is usually solved by maximizing $L_n(F)$ over the class of distribution functions that have a density with

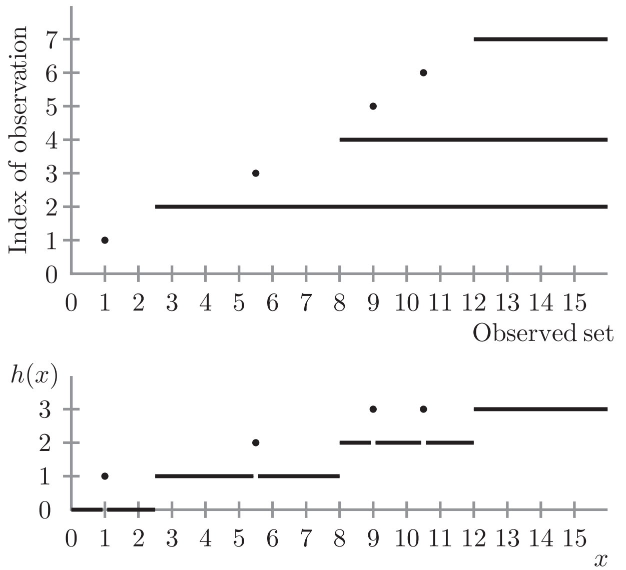

Fig. 1. Observed sets (upper panel) and the corresponding height map (lower panel) for univariate right censored data, based on the following 7 observations of (U, Δ): (1,1), (2.5,0), (5.5,1), (8,0), (9,1), (10.5,1) and (12,0). Note that the maximal intersections are given by the local maximum regions of the height map: $\{1\}$, $\{5.5\}$, $\{9\}$, $\{10.5\}$ and $(12, \infty)$.

respect to counting measure on the observed failure times. We can then write $L_n(F) = \prod_{i=1}^n P_F(A_i)$, where $P_F(A)$ is the probability of A under F.

It is well-known (Peto, 1973; Turnbull, 1976) that the MLE in censored data problems can only assign mass to a finite number of disjoint regions, called *maximal intersections* by Wong & Yu (1999). Maathuis (2005) introduced an efficient algorithm to compute the maximal intersections for d-variate interval-censored data. This algorithm is based on a height map $h: \mathbb{R}^d \to \mathbb{N}$ of the observed sets, where $h(x)$ is defined as the number of observed sets that contain x. Maathuis showed that the maximal intersections correspond exactly to the local maximum regions of the height map of the observed sets. (If there are ties in the observed sets, then these need to be resolved before applying the height map, see Maathuis, 2005.)

The height map $h: \mathbb{R} \mapsto \mathbb{N}$ for univariate right-censored data is illustrated in Figure 1. Note that $h(x)$ simply represents the number of observed sets $A_1, \ldots, A_n$ that overlap at the point x. It is clear that all sets $A_{(i)}$ with $i \in \mathcal{I} = \{i \in \{1, \ldots, n\} : \Delta_{(i)} = 1\}$, or in other words, all sets of the form $A_{(i)} = \{U_{(i)}\}$, are local maxima of the height map. Hence, all such sets are maximal intersections, and we denote these by $M_{(i)}$, $i \in \mathcal{I}$. This notation may seem redundant since $M_{(i)} = A_{(i)}$, but it will be useful in Section 2.2. Furthermore, if and only if $\Delta_{(n)} = 0$, the height map has an extra local maximum region $A_{(n)} = (U_{(n)}, \infty)$, resulting in an extra maximal intersection $M_{(n+1)} = (U_{(n)}, \infty)$. This situation is illustrated in Figure 1. Let $\bar{\mathcal{I}}$ be the collection of indices of all maximal intersections. Thus, $\bar{\mathcal{I}} = \mathcal{I}$ if $\Delta_{(n)} = 1$ and $\bar{\mathcal{I}} = \mathcal{I} \cup \{n+1\}$ if $\Delta_{(n)} = 0$.

Let p_i be the probability mass of maximal intersection $M_{(i)}$, $i \in \bar{\mathcal{I}}$. We can then write the likelihood in terms of the p_is:

$$\prod_{i=1}^n P(A_i) = \prod_{i=1}^n \left(\sum_{j \in \bar{\mathcal{I}}} p_j 1\{M_{(j)} \subseteq A_{(i)}\} \right) = \prod_{i=1}^n p_i^{\Delta_{(i)}} \left(\sum_{j \geq i+1, j \in \bar{\mathcal{I}}} p_j \right)^{1-\Delta_{(i)}}, \tag{3}$$

where the second equality follows from the fact that the data are ordered with respect to the variable $U = \min(X, T)$. The MLE $\hat{p}$ maximizes this expression under the constraints

$$\sum_{i \in \bar{\mathcal{I}}} p_i = 1 \quad \text{and} \quad p_i \geq 0 \quad \text{for all} \quad i \in \bar{\mathcal{I}}. \tag{4}$$

It is well-known that $\hat{p}$ is the Kaplan–Meier or product-limit estimator, given by

$$\hat{p}_i = \prod_{j=1}^{i-1} \left(1 - \frac{\Delta_{(j)}}{n-j+1}\right) \frac{\Delta_{(i)}}{n-i+1}, \quad i \in \mathcal{I},$$

and $\hat{p}_{n+1} = 1 - \sum_{i \in \mathcal{I}} \hat{p}_i$ if $\Delta_{(n)} = 0$ (see, for example, Shorack & Wellner, 1986, chapter 7, pp. 332–333). Equivalently, we can write

$$\sum_{j \geq i, j \in \bar{\mathcal{I}}} \hat{p}_j = \prod_{j \leq i-1} \left(1 - \frac{\Delta_{(j)}}{n-j+1}\right), \quad i \in \bar{\mathcal{I}}.$$

The vector $\hat{p}$ is uniquely determined. We obtain $\hat{F}_n(x)$ by summing all probability mass of $\hat{p}$ that falls in the interval $(0, x]$. It is well-known that $\hat{F}_n(x)$ is non-unique for $x > U_{(n)}$ if and only if $\Delta_{(n)} = 0$. This is caused by the fact that the MLE is indifferent to the distribution of mass within a maximal intersection, called 'representational non-uniqueness' by Gentleman & Vandal (2002). Since all maximal intersections $\{M_{(i)} : i \in \mathcal{I}\}$ are points, this non-uniqueness occurs if and only if $M_{(n+1)} = (U_{(n)}, \infty)$ exists, and this happens if and only if $\Delta_{(n)} = 0$.

2.2. Continuous-mark data: explicit formula for the MLE

We now return to the interval-censored continuous-mark model given in Section 1, and introduce some additional notation. Let $F_0(x, y) = P(X \leq x, Y \leq y)$ be the joint distribution of (X, Y), and let $F_{0X}(x) = F_0(x, \infty) = P(X \leq x)$ and $F_{0Y}(y) = F_0(\infty, y) = P(Y \leq y)$ be the marginal distributions of X and Y, respectively. Recall that G denotes the distribution of the observation times $\mathbf{T}$. We use subscripts to denote the marginal distributions of G. For example, G_1 is the distribution of T_1 and $G_{2,3}$ is the distribution of (T_2, T_3). For current status censoring ($k = 1$), we denote the observation time simply by T.

We study the MLE $\hat{F}_n$ of F_0, based on n i.i.d. copies $W_1, \ldots, W_n$ of W, where $W_i = (\mathbf{T}_i, \mathbf{\Delta}_i, Z_i)$, $\mathbf{T}_i = (T_{1i}, \ldots, T_{ki})$ and $\mathbf{\Delta}_i = (\Delta_{1i}, \ldots, \Delta_{k+1,i})$. We allow ties between the observation times of $\mathbf{T}_i$ and $\mathbf{T}_j$ for $i \neq j$.

The observed sets A in this model are given in equations (1) and (2). Recall that A is a line segment if $\Delta_+ = 1$ and that A is a half plane if $\Delta_+ = 0$. Assuming that F has a density f with respect to some dominating measure $\mu_X \times \mu_Y$, the likelihood (up to multiplicative terms only depending on G) is given by $L_n(F) = \prod_{i=1}^n q(W_i)$, where

$$q(w) = q(t, \delta, z) = \prod_{j=1}^{k} \left\{ \int_{(t_{j-1}, t_j]} f(s, z) \mu_X(\mathrm{d}s) \right\}^{\delta_j} \left(1 - F_X(t_k)\right)^{1-\delta_+},$$

and $F_X(x) = F(x, \infty)$ is the marginal distribution of X under F. Since the first term of q is a density-type term, $L_n(F)$ can be made arbitrarily large by letting $f(s, z)$ peak at $z = Z_i$ for some observation with $\Delta_{+i} = 1$. We therefore define the MLE $\hat{F}_n(x, y)$ to be the maximizer of $L_n(F)$ over the class $\mathcal{F}$ of all bivariate distribution functions that have a marginal density f_Y with respect to counting measure on the observed marks. We can then write $L_n(F) = \prod_{i=1}^n P_F(A_i)$.

Analogously to Maathuis (2005), we call the projection of A on the x-axis the *x-interval* of A. We denote the left endpoint and right endpoint of the x-interval by L and R:

$$L=\sum_{j=1}^{k+1}\Delta_j T_{j-1}, \quad R=\sum_{j=1}^{k+1}\Delta_j T_j.$$

Furthermore, we define a new variable U that will play an important role in our analysis:

$$U=\Delta_+ R+\Delta_{k+1}L. \tag{5}$$

Let $U_{(1)},\dots,U_{(n)}$ be the order statistics of $U_1,\dots,U_n$ and let $\boldsymbol{\Delta}_{(i)}=(\Delta_{1(i)},\dots,\Delta_{k+1,(i)})$, $Z_{(i)}$, $A_{(i)}$, $L_{(i)}$ and $R_{(i)}$ be the corresponding values of $\boldsymbol{\Delta}$, Z, A, L and R. We break ties in U arbitrarily after ensuring that observations with $\Delta_+=1$ are ordered before those with $\Delta_+=0$. Recall that the maximal intersections are the local maximum regions of the height map $h:\mathbb{R}^2\mapsto\mathbb{N}$ of the observed sets. Since Y is continuous, the observed sets $A_{(i)}$ with $i\in\mathcal{I}=\{i\in\{1,\dots,n\}:\Delta_{+(i)}=1\}$ are completely distinct with probability one. Hence, each such $A_{(i)}$ contains exactly one maximal intersection $M_{(i)}$ of the form:

$$\begin{aligned} &M_{(i)}=(D_{(i)},R_{(i)}]\times\{Z_{(i)}\}, \quad \text{where} \\ &D_{(i)}=\max\{\{L_{(j)}:j\notin\mathcal{I},j<i\}\cup\{L_{(i)}\}\}. \end{aligned} \tag{6}$$

To understand this expression, let $S_{(i)}$ be the collection of observed sets $A_{(j)}$ with $\Delta_{+(j)}=0$ and $L_{(i)}<L_{(j)}<R_{(i)}$. If $S_{(i)}=\emptyset$, then the height map is constant on $A_{(i)}$, and the complete set $A_{(i)}$ is a local maximum region. Hence, in this case $M_{(i)}=A_{(i)}$ and $D_{(i)}=L_{(i)}$. On the other hand, if $S_{(i)}\neq\emptyset$, then the height map is increasing on $A_{(i)}$ in the x-direction. Hence, in this case $M_{(i)}\subsetneq A_{(i)}$ and the left endpoint of $M_{(i)}$ is $\max\{L_{(j)}:A_{(j)}\in S_{(i)}\}$, which equals $\max\{L_{(j)}: j\notin\mathcal{I},j<i\}$. Note that the right endpoints of $M_{(i)}$ and $A_{(i)}$ are always identical. Moreover, note that the equations in (6) imply that the maximal intersections can be computed in $O(n\log n)$ time, since the most computationally intensive step consists of sorting the data. This is faster than the height map algorithm of Maathuis (2005), because of the special structure in the data.

Analogously to the situation for univariate right-censored data, there is an extra maximal intersection $M_{(n+1)}=A_{(n)}=(U_{(n)},\infty)\times\mathbb{R}$ if and only if $\Delta_{+(n)}=0$. Let $\bar{\mathcal{I}}$ be the collection of indices of all maximal intersections. Thus, $\bar{\mathcal{I}}=\mathcal{I}$ if $\Delta_{+(n)}=1$ and $\bar{\mathcal{I}}=\mathcal{I}\cup\{n+1\}$ if $\Delta_{+(n)}=0$. Let p_i be the probability mass of maximal intersection $M_{(i)}$, $i\in\bar{\mathcal{I}}$. Then the likelihood can be written as

$$\prod_{i=1}^n P(A_i)=\prod_{i=1}^n\left(\sum_{j\in\bar{\mathcal{I}}}p_j 1\{M_{(j)}\subseteq A_{(i)}\}\right)=\prod_{i=1}^n p_i^{\Delta_{+(i)}}\left(\sum_{j\geq i+1,j\in\bar{\mathcal{I}}}p_j\right)^{1-\Delta_{+(i)}}, \tag{7}$$

where the second equality follows from the fact that the data are ordered with respect to the variable U which was defined in (5). The MLE $\hat{p}$ maximizes this expression under the constraints (4). From the analogy with the likelihood (3) it follows immediately that

$$\hat{p}_i=\prod_{j=1}^{i-1}\left(1-\frac{\Delta_{+(j)}}{n-j+1}\right)\frac{\Delta_{+(i)}}{n-i+1}, \quad i\in\mathcal{I}, \tag{8}$$

and $\hat{p}_{n+1}=1-\sum_{i\in\mathcal{I}}\hat{p}_i$ if $\Delta_{+(n)}=0$. Equivalently, we can write

$$\sum_{j\geq i,j\in\bar{\mathcal{I}}}\hat{p}_j=\prod_{j\leq i-1}\left(1-\frac{\Delta_{+(j)}}{n-j+1}\right), \quad i\in\bar{\mathcal{I}}. \tag{9}$$

These formulas are different from (but equivalent to) the ones given in section 3.1 of HMG. The form given here has several advantages. First, the tail probabilities (9) can be computed in time complexity $O(n \log n)$, since sorting the data is the most computationally intensive step. Furthermore, the current form provides additional insights about the behavior of the MLE. In particular, it shows that the MLE can be viewed as a right endpoint imputation estimator (see Remark 1), and it allows for a derivation of the almost sure limit of the MLE (see section 3).

The vector $\hat{p}$ is uniquely determined. This was noted by HMG and also follows from our derivation here. We obtain $\hat{F}_n(x,y)$ by summing all probability mass of $\hat{p}$ that falls in the region $(0,x] \times (-\infty, y]$. We define a marginal MLE for the distribution of X by letting $\hat{F}_{Xn}(x) = \hat{F}_n(x, \infty)$. The estimators $\hat{F}_n$ and $\hat{F}_{Xn}$ can suffer considerably from representational non-uniqueness, since the maximal intersections $\{M_{(i)} : i \in \mathcal{I}\}$ are line segments, and the potential maximal intersection $M_{(n+1)}$ is a half plane. We let $\hat{F}_n^{\ell}$ denote the estimator that assigns all mass to the upper right corners of the maximal intersections, since it is a lower bound for the MLE. Similarly, we let $\hat{F}_n^{u}$ denote the estimator that assigns all mass to the lower left corners of the maximal intersections, since it is an upper bound for the MLE. The formulas for $\hat{F}_n^{\ell}$ and $\hat{F}_{Xn}^{\ell}$ can be written as follows:

$$1 - \hat{F}_{Xn}^{\ell}(x) = \prod_{U_{(i)} \le x} \left(1 - \frac{\Delta_{+(i)}}{n-i+1}\right), \tag{10}$$

$$\begin{aligned} \hat{F}_n^{\ell}(x,y) &= \sum_{i=1}^{n} \hat{p}_i 1\{U_{(i)} \le x, Z_{(i)} \le y\} \\ &= \sum_{U_{(i)} \le x} \prod_{U_{(j)} < U_{(i)}} \left(1 - \frac{\Delta_{+(j)}}{n-j+1}\right) \frac{\Delta_{+(i)} 1\{Z_{(i)} \le y\}}{n-i+1}, \end{aligned} \tag{11}$$

using (8), (9) and the definition of U in (5).

Remark 1. The MLE $\hat{F}_n^{\ell}$ can be viewed as a right endpoint imputation estimator. To see this, consider creating a new collection of observed sets $A'_{(i)}$:

$$A'_{(i)} = \begin{cases} \{U_{(i)}\} \times \{Z_{(i)}\} & \text{if } i \in \mathcal{I}, \\ A_{(i)} & \text{if } i \notin \mathcal{I}. \end{cases}$$

That is, for each $i = 1, \ldots, n$, we replace $A_{(i)}$ by its right endpoint if $\Delta_{+(i)} = 1$, while we leave it unchanged if $\Delta_{+(i)} = 0$. The intersection structures of $\{A_{(i)}\}_{i=1}^{n}$ and $\{A'_{(i)}\}_{i=1}^{n}$ are identical, meaning that $A_{(i)} \cap A_{(j)} = \emptyset$ if and only if $A'_{(i)} \cap A'_{(j)} = \emptyset$, for all $i, j \in \{1, \ldots, n\}$. Furthermore, the maximal intersections of $\{A'_{(i)}\}_{i=1}^{n}$ are $\{M'_{(i)} = A'_{(i)} : i \in \bar{\mathcal{I}}\}$. Hence, writing the likelihood for the imputed data in terms of p yields exactly the same likelihood as (7) . This implies that the maximizing vector $\hat{p}'$ is identical to the vector $\hat{p}$ for the original data. Moreover, the upper right corners of $\{M_{(i)}\}$, $i \in \bar{\mathcal{I}}$ and $\{M'_{(i)}\}$, $i \in \bar{\mathcal{I}}$ are identical. Since $\hat{F}_n^{\ell}$ assigns all mass to the upper right corners of the maximal intersections, it follows that $\widehat{F}_n^{\ell}$ is completely equivalent to the MLE for the modified data. Finally, note that the right endpoint imputation scheme imputes an x-value that is always at least as large as the unobserved X. This explains why the MLE $\widehat{F}_{Xn}^{\ell}$ tends to have a negative bias.

3. Inconsistency of the MLE

In this section we derive necessary and sufficient conditions for consistency of the MLEs $\hat{F}_{Xn}^{\ell}$ and $\hat{F}_n^{\ell}$ (Theorem 1). These conditions force a relation between the unknown distribution F_0 and the observation time distribution G. Since such a relation will typically not hold, it follows that $\hat{F}_n^{\ell}$ is inconsistent in general. Corollary 1 further strengthens this result when X

is subject to current status censoring, and shows that in that case $\hat{F}^{\ell}_{Xn}$ is inconsistent for any continuous choice of F_0 and G. Corollary 2 shows that the asymptotic biases of $\widehat{F}^{\ell}_{Xn}$ and $\hat{F}^{\ell}_{n}$ converge to zero as the number k of observation times per subject increases, at least for one particular distribution of $T_1, \ldots, T_k$.

The results in this section are based on deriving the limits $F^{\ell}_{X\infty}$ and F^{ℓ}_{∞} for the lower bounds $\hat{F}^{\ell}_{Xn}$ and $\hat{F}^{\ell}_{n}$ of the MLE. The reason for looking at these lower bounds is that $\hat{F}^{\ell}_{Xn}$ and $\hat{F}^{\ell}_{n}$ can be expressed in simple closed forms; see (10) and (11). Moreover, in many cases representational non-uniqueness disappears in the limit, so that the limits of $\hat{F}_{Xn}$ and $\hat{F}_{n}$ are unique and equal to $F^{\ell}_{X\infty}$ and F^{ℓ}_{∞}. Necessary and sufficient conditions for uniqueness of the limit are: (i) all maximal intersections $M_{(i)}$, $i \in \mathcal{I}$, converge to points, and (ii) $\sum_{i \in \mathcal{I}} \hat{p}_i \to 1$ as $n \to \infty$. These conditions are satisfied in Examples 1 and 2 in section 5. If these conditions fail, then the upper bounds $F^{u}_{X\infty}$ and F^{u}_{∞} can be obtained from their lower bounds by reassigning mass from the upper right corners of the maximal intersections to the lower left corners. This occurs in Examples 3 and 4 in section 5, and further details can be found in Maathuis (2006, section 9.4).

In order to derive $F^{\ell}_{X\infty}$ and F^{ℓ}_{∞} we start by rewriting (10) and (11) in terms of stochastic processes. We introduce the following notation:

$$\begin{aligned} \mathbb{H}_n(x) &= \mathbb{P}_n 1\{U \le x\}, \quad x \ge 0, \\ \mathbb{V}_n(x, y) &= \mathbb{P}_n \Delta_+ 1\{U \le x, Z \le y\}, \quad x \ge 0, y \in \mathbb{R}, \\ \mathbb{V}_{Xn}(x) &\equiv \mathbb{V}_n(x, \infty) = \mathbb{P}_n \Delta_+ 1\{U \le x\}, \quad x \ge 0, \end{aligned} \tag{12}$$

where U is defined in (5) and $\mathbb{P}_n f(X) = n^{-1} \sum_{i=1}^n f(X_i)$. Furthermore, let

$$\hat{\Lambda}_n(x, y) = \int_{[0,x]} \frac{\mathbb{V}_n(\mathrm{d}s, y)}{1 - \mathbb{H}_n(s-)} \quad \text{and} \quad \hat{\Lambda}_{Xn}(x) \equiv \hat{\Lambda}_n(x, \infty) = \int_{[0,x]} \frac{\mathbb{V}_{Xn}(\mathrm{d}s)}{1 - \mathbb{H}_n(s-)}. \tag{13}$$

Since

$$\hat{\Lambda}_n(\mathrm{d}x, y) = \frac{\mathbb{P}_n \Delta_+ 1\{U = x, Z \le y\}}{\mathbb{P}_n 1\{U \ge x\}} \quad \text{and} \quad \hat{\Lambda}_{Xn}(\mathrm{d}x) = \frac{\mathbb{P}_n \Delta_+ 1\{U = x\}}{\mathbb{P}_n 1\{U \ge x\}},$$

we can write (10) and (11) in terms of $\hat{\Lambda}_{Xn}$ and $\hat{\Lambda}_n$:

$$1 - \hat{F}^{\ell}_{Xn}(x) = \prod_{s \le x} \{1 - \hat{\Lambda}_{Xn}(\mathrm{d}s)\}, \tag{14}$$

$$\hat{F}^{\ell}_{n}(x, y) = \int_{s \le x} \prod_{u < s} \{1 - \hat{\Lambda}_{Xn}(\mathrm{d}u)\} \hat{\Lambda}_n(\mathrm{d}s, y). \tag{15}$$

Note that (14) is analogous to the Kaplan–Meier estimator for right-censored data, and that (15) is analogous to equation (3.3) of Huang & Louis (1998). However, our functions $\hat{\Lambda}_{Xn}$ and $\hat{\Lambda}_n$ are defined differently, since they are based on the variable U. This difference lies at the root of the inconsistency problems of the MLE.

The limits of the processes $\mathbb{H}_n$, $\mathbb{V}_n$, $\mathbb{V}_{Xn}$, $\hat{\Lambda}_n$, $\hat{\Lambda}_{Xn}$, $\hat{F}^{\ell}_{n}$ and $\hat{F}^{\ell}_{Xn}$ are given in the appendix (Lemmas 1–3) and are denoted by H, V, V_X, Λ_∞, $\Lambda_{X\infty}$, F^{ℓ}_{∞} and $F^{\ell}_{X\infty}$, respectively. Corollaries 3–5 in the Appendix provide various alternative ways to express F^{ℓ}_{∞}.

We are now ready to give necessary and sufficient conditions for consistency of $\hat{F}^{\ell}_{Xn}$ and $\hat{F}^{\ell}_{n}$, after introducing the following notation:

$$H(x) = V_X(x) + \int_{[0,x]} \{1 - F_{0X}(s)\} \, \mathrm{d}G_k(s), \tag{16}$$

$$V(\mathrm{d}x, y) = \sum_{j=1}^{k} F_0(x, y)\,\mathrm{d}G_j(x) - \sum_{j=2}^{k} \int_{[0,x]} F_0(s, y)\,\mathrm{d}G_{j-1,j}(s, x), \tag{17}$$

$$V_X(\mathrm{d}x) = \sum_{j=1}^{k} F_{0X}(x)\,\mathrm{d}G_j(x) - \sum_{j=2}^{k} \int_{[0,x]} F_{0X}(s)\,\mathrm{d}G_{j-1,j}(s, x), \tag{18}$$

See (22)–(24) in the appendix. Moreover, throughout this section we let τ be such that $H(\tau) < 1$, we define $0/0 = 0$ and $f(x-) = \lim_{t \uparrow x} f(t)$ for any function $f : \mathbb{R} \mapsto \mathbb{R}$.

Theorem 1

The MLE is inconsistent in general. The MLE $\hat{F}^{\ell}_{Xn}$ is consistent for F_{0X} on $(0, \tau]$ if and only if the following condition holds for all $x \in (0, \tau]$:

$$\Lambda_{X\infty}(x) \equiv \int_{[0,x]} \frac{V_X(\mathrm{d}s)}{1 - H(s-)} = \int_{[0,x]} \frac{F_{0X}(\mathrm{d}s)}{1 - F_{0X}(s-)} \equiv \Lambda_{0X}(x). \tag{19}$$

The MLE $\hat{F}^{\ell}_{n}$ is consistent for F_0 on $(0, \tau] \times \mathbb{R}$ if and only if the following condition holds for all $x \in (0, \tau]$, $y \in \mathbb{R}$:

$$\Lambda_{\infty}(x, y) \equiv \int_{[0,x]} \frac{V(\mathrm{d}s, y)}{1 - H(s-)} = \int_{[0,x]} \frac{F_0(\mathrm{d}s, y)}{1 - F_{0X}(s-)} \equiv \Lambda_0(x, y). \tag{20}$$

Finally, let $x_0 \in (0, \tau]$ with $F_{X\infty}(x_0) > 0$. Then $\hat{F}^{\ell}_{n}(x_0, y)/\hat{F}^{\ell}_{Xn}(x_0)$ is consistent for $F_{0Y}(y)$ if X and Y are independent.

Proof. The one-to-one correspondence between a univariate distribution function and its cumulative hazard function implies that $\hat{F}^{\ell}_{Xn}$ is consistent for F_{0X} if and only if $\Lambda_{X\infty}$ [(26) in the appendix] equals the cumulative hazard function Λ_{0X} of F_{0X}. This gives condition (19). Similarly, it follows that $\hat{F}^{\ell}_{n}(x, y)$ is consistent for $F_0(x, y)$ if and only if Λ_{∞} [(25) in the appendix] equals the mark-specific cumulative hazard function Λ_0 of F_0. This gives condition (20). The final claim of the theorem follows from (32) in the appendix. □

Note that conditions (19) and (20) are difficult to interpret, since F_{0X} and F_0 enter on both sides of the equations when we plug in expressions (16)–(18) for $H(s-)$, $V(\mathrm{d}s, y)$ and $V_X(\mathrm{d}s)$. However, it is clear that the conditions force a relation between the unknown distribution F_0 and the observation time distribution G. Such a relation will typically not hold and cannot be assumed since F_0 is unknown. Hence, it follows that the MLE is inconsistent in general. The following corollary further strengthens this result when X is subject to current status censoring.

Corollary 1

Let X be subject to current status censoring, and let F_{0X} and G be continuous. Then the MLE $\hat{F}^{\ell}_{Xn}$ is inconsistent for any choice of F_{0X} and G.

Proof. Let $\gamma = \inf\{x : F_{0X}(x) > 0\} < \tau$. Since X is subject to current status censoring and since the distributions G and F_{0X} are continuous, condition (19) can be rewritten as

$$\int_{(\gamma,x]} \frac{\mathrm{d}G(s)}{1 - G(s)} = \int_{(\gamma,x]} \frac{\mathrm{d}F_{0X}(s)}{F_{0X}(s)\{1 - F_{0X}(s)\}}, \quad x \in (\gamma, \tau].$$

This integral equation is solved by

$$-\log\{1-G(x)\}+C=\log\left\{\frac{F_{0X}(x)}{1-F_{0X}(x)}\right\}, \quad x\in(\gamma,\tau].$$

This yields $F_{0X}(x)=[1+\exp(-C)\{1-G(x)\}]^{-1}$ for $x\in(\gamma,\tau)$. Since there is no finite C such that $F_{0X}(\gamma)=0$ holds, it follows that condition (19) fails for all continuous distributions G and F_{0X}. □

Finally, we show that the asymptotic bias of the MLE converges to zero as the number k of observation times per subject increases, for at least one particular distribution of $\mathbf{T}=(T_1,\ldots,T_k)$, namely if $T_1,\ldots,T_k$ are distributed as the order statistics of a uniform sample on $[0,\theta]$. The proof of this result is given in the appendix.

Corollary 2

Let X be subject to interval censoring case k, and let the elements $T_1,\ldots,T_k$ of $\mathbf{T}$ be the order statistics of k independent uniform random variables on $[0,\theta]$. Let $V^k(x,y)$, $V_X^k(x)$, $H^k(x)$, $\Lambda_\infty^k(x,y)$ and $\Lambda_{X\infty}^k(x)$ denote the limits defined in Lemmas 1 and 2, using the superscript k to denote the dependence on k. Then

$$\Lambda_{X\infty}^k(x)=\int_{[0,x]}\frac{V_X^k(\mathrm{d}s)}{1-H^k(s-)}\to\int_{[0,x]}\frac{F_{0X}(\mathrm{d}s)}{1-F_{0X}(s-)}=\Lambda_{0X}(x), \quad k\to\infty,$$

$$\Lambda_\infty^k(x,y)=\int_{[0,x]}\frac{V^k(\mathrm{d}s,y)}{1-H^k(s-)}\to\int_{[0,x]}\frac{F_0(\mathrm{d}s,y)}{1-F_{0X}(s-)}=\Lambda_0(x,y), \quad k\to\infty,$$

for all continuity points of Λ_{0X} and Λ_0 with $x<\theta$ and $y\in\mathbb{R}$.

4. Repaired MLE via discretization of marks

We now define a simple repaired estimator $\tilde{F}_n(x,y)$ which is consistent for $F_0(x,y)$ for y on a grid. The idea behind the estimator is that one can define discrete competing risks based on a continuous random variable. Doing so transforms interval-censored continuous-mark data into interval-censored data with competing risks.

To describe the method, we let $K>0$ and define a grid $-\infty\equiv y_0<y_1<\cdots<y_K<y_{K+1}\equiv\infty$. Next, we introduce a new random variable $C\in\{1,\ldots,K+1\}$:

$$C=\sum_{j=1}^{K+1} j1\{y_{j-1}<Y\le y_j\}.$$

We can determine the value of C for all observations with an observed mark. Hence, we can transform the observations $(\mathbf{T},\mathbf{\Delta},Z)$ into $(\mathbf{T},\mathbf{\Delta},Z^*)$, where $Z^*=\Delta_+C$. This gives interval-censored data with $K+1$ competing risks.

Since the observed sets for interval-censored data with competing risks form a partition of the space $\mathbb{R}_+\times\{1,\ldots,K+1\}$, Hellinger consistency of the MLE follows from theorems 9 and 10 of Van der Vaart & Wellner (2000). Under some additional regularity conditions, we can derive local and uniform consistency from the Hellinger consistency; see Maathuis (2006, section 4.2). This means that we can consistently estimate the sub-distribution functions $F_{0j}(x)=P(X\le x, C=j)=P(X\le x, y_{j-1}<Y\le y_j)$, $x\in\mathbb{R}_+$. Hence, we can consistently estimate $F_0(x,y_j)=\sum_{\ell=1}^{j}F_{0\ell}(x)$ for $x\in\mathbb{R}_+$ and y_j on the grid.

Note that the introduction of the variable C causes more overlap between observed sets, since previously non-overlapping horizontal line segments may overlap if they are assigned the same value of C. As a result, the repaired MLE has smaller maximal intersections in the x-direction. Hence, the repaired MLE is affected less by representational non-uniqueness on the x-axis. This is visible in Examples 3 and 4 in section 5.

The repaired MLE can be computed with one of the algorithms described in Groeneboom *et al.* (2006a, section 2.4). It may be tempting to choose K large, such that $F_0(x, y)$ can be estimated for y on a fine grid. However, this may result in a poor estimator. To obtain a good estimator one should choose the grid such that there are ample observations for each value of C. In practice, one can start with a coarse grid, and then refine the grid as long as the estimator stays close to the one computed on the coarse grid.

In principle it is possible to estimate the entire joint distribution function $F_0(x, y)$ for (x, y) in the interior of the support of the distribution of the observation times under smoothness assumptions on F_0. This would proceed by letting both K and the y_js defining the partition all depend on n in such a way that $K = K_n \to \infty$,

$$\max_{1\le j\le K_n-1}(y_{j+1,n} - y_{j,n}) \to 0, \qquad \text{and} \qquad n \min_{1\le j\le K_n-1}(y_{j+1,n} - y_{j,n}) \to \infty,$$

as $n \to \infty$. It would even be possible to choose K_n and $\{y_{j,n}\}$ depending on the data via model-selection methods (see, for example, Birgé & Massart, 1997; Barron, Birgé & Massart, 1999), but these further developments are beyond the scope of the present paper and will be investigated in detail elsewhere.

Maathuis (2006) and Groeneboom, *et al.* (2006a, 2006b) showed that the MLE for current status data with competing risks converges at rate $n^{1/3}$ to a new self-induced limiting distribution. This result implies that one can use subsampling to construct pointwise confidence intervals for the sub-distribution functions (Politis *et al.*, 1999). This method is also valid for the repaired MLE for current status data with continuous marks, and can be used for the construction of pointwise confidence intervals for $F_0(x, y)$ for y on the grid. The limiting distribution of the MLE for more general forms of interval censoring with competing risks has not yet been established, and in such cases the use of sub-sampling is therefore not yet justified.

Jewell *et al.* (2003) and Maathuis (2006, chapter 7) studied estimation of a family of smooth functionals of the sub-distribution functions for current status data with competing risks. Jewell *et al.* (2003) suggested that their 'naive estimator' yields asymptotically efficient estimators for these smooth functionals, and Maathuis (2006) showed that the same is true for the MLE. These results extend to the repaired MLE for current status data with continuous marks. Asymptotic properties of estimators of smooth functionals for more general forms of interval censoring with competing risks are currently still unknown.

5. Examples

In this section we support the theoretical results of sections 3 and 4 by simulations. In particular, we show support for our claims that $\hat{F}_n^{\ell} \to_{\text{a.s.}} F_{\infty}^{\ell}$, $\hat{F}_n^{u} \to_{\text{a.s.}} F_{\infty}^{u}$ and $\tilde{F}_n \to_{\text{a.s.}} F_0$. Moreover, we show that the difference between the true underlying distribution F_0 and the limits of the MLE F_{∞}^{ℓ} and F_{∞}^{u} can be considerable. We give four examples that cover a wide range of scenarios. They include cases where X and Y are independent (Example 1) or dependent (Examples 2–4), where X is subject to interval-censoring case 1 (Examples 1 and 2) or case 2 (Examples 3 and 4), and where the distribution of $\mathbf{T}$ is continuous (Examples 1–3) or discrete (Example 4).

Example 1. Let X and Y be independent, with $X \sim \text{Unif}(0, 1)$ and $Y \sim \exp(1)$. Let X be subject to current status censoring with observation time $T \sim \text{Unif}(0, 0.5)$ independent of (X, Y).

Example 2. Let $X \sim \mathrm{Unif}(0,1)$, and let $Y|X$ be exponentially distributed with mean $2/(2X+1)$. Let X be subject to current status censoring with observation time $T \sim \mathrm{Unif}(0,1)$ independent of (X, Y).

Example 3. Let $X \sim \mathrm{Unif}(0,2)$, and let $Y \equiv X$. Let X be subject to interval censoring case 2 with observation times (T_1, T_2), independent of (X, Y) and uniformly distributed over $\{(t_1,t_2): 0 \le t_1 \le 1, 1 \le t_2 \le 2\}$.

Example 4. Let (X, Y) be uniformly distributed over $\{(x,y): 0 \le x \le y \le 1\}$. Let X be subject to interval censoring case 2 with observation times (T_1, T_2) independent of (X, Y). Let the distribution of (T_1, T_2) be discrete: $G\{(0.25,0.5)\}=0.3$, $G\{(0.25,0.75)\}=0.3$ and $G\{(0.5,0.75)\}=0.4$.

For each example we derived the limits F_∞^ℓ and F_∞^u of the MLE, using Lemma 3. Details of these derivations are given in Maathuis (2006, section 9.4). We also computed the MLEs $\hat{F}_n^\ell$ and $\hat{F}_n^u$ and the repaired MLEs $\tilde{F}_n^\ell$ and $\tilde{F}_n^u$ for a simulated data set of size $n=10,000$. For the repaired MLE we used an equidistant grid with $K=20$ points as shown in Figure 4.

The results are given in Figures 2–4. These figures show that the MLEs $\hat{F}_n^\ell$ and $\hat{F}_n^u$ are indeed very close to our derived limits F_∞^ℓ and F_∞^u. On the other hand, the repaired MLEs $\tilde{F}_n^\ell$ and $\tilde{F}_n^u$ are very close to the true underlying distribution F_0. Moreover, the results show that there can be a very significant difference between the limit of the MLE and the true underlying distribution F_0.

We now discuss the simulation results in more detail. Figure 2 considers estimation of the joint distribution F_0. It shows the contour lines of the MLE $\hat{F}_n^\ell$, its limit F_∞^ℓ, and the true underlying distribution F_0. Note that $\hat{F}_n^\ell$ and F_∞^ℓ are almost indistinguishable, while there is a clear difference between F_∞^ℓ and F_0. The results for the upper limits $\hat{F}_n^u$ and F_∞^u are similar and not shown. Results for the repaired MLE are not shown since this estimator only takes values for y on a grid.

Figure 3 considers estimation of the marginal distribution F_{0X}. We see that the MLEs $\hat{F}_{Xn}^\ell$ and $\hat{F}_{Xn}^u$ are close to the derived limits $F_{X\infty}^\ell$ and $F_{X\infty}^u$. Moreover, note that $\hat{F}_{Xn}^\ell$ tends to be below F_{0X}. This can be understood via Remark 1, which explains that $\hat{F}_n^\ell$ can be viewed as a right endpoint estimator, and hence tends to have a negative bias. Note that the repaired MLE $\tilde{F}_n$ closely follows F_{0X}.

Figure 4 considers estimation of $F_0(x_0, y)$ for fixed x_0. The function $F_0(x_0, y)$ is often estimated as an alternative for F_{0Y}, since F_{0Y} is heavily affected by representational non-uniqueness if the support of $T_1, \ldots, T_k$ is strictly contained in the support of X, a situation that often occurs in practice. The values of x_0 were chosen to show a range of scenarios for the behavior of the MLE, and we see that $\hat{F}_n(x_0, y)$ can be much too large, much too small and non-unique. The repaired MLE $\tilde{F}_n$ is again close to the underlying distribution.

Note that our examples are not linked to any specific application. For readers who are interested in a comparison between the MLE and the repaired MLE in a practical situation, we refer to HMG. They provide such a comparison for the HIV/AIDS vaccine trial data VAX004 (Flynn *et al.*, 2005), as well as for simulated data that mimic the vaccine data. They show a difference between the MLE and the repaired MLE in this setting, but the size of the difference is quite small. This can be explained by Corollary 2, since the time between successive follow-up visits is relatively short (about 6 months) and the infection rate is low. Much larger differences can be expected in, for example, cross-sectional HIV studies, where there is only one observation time per person.

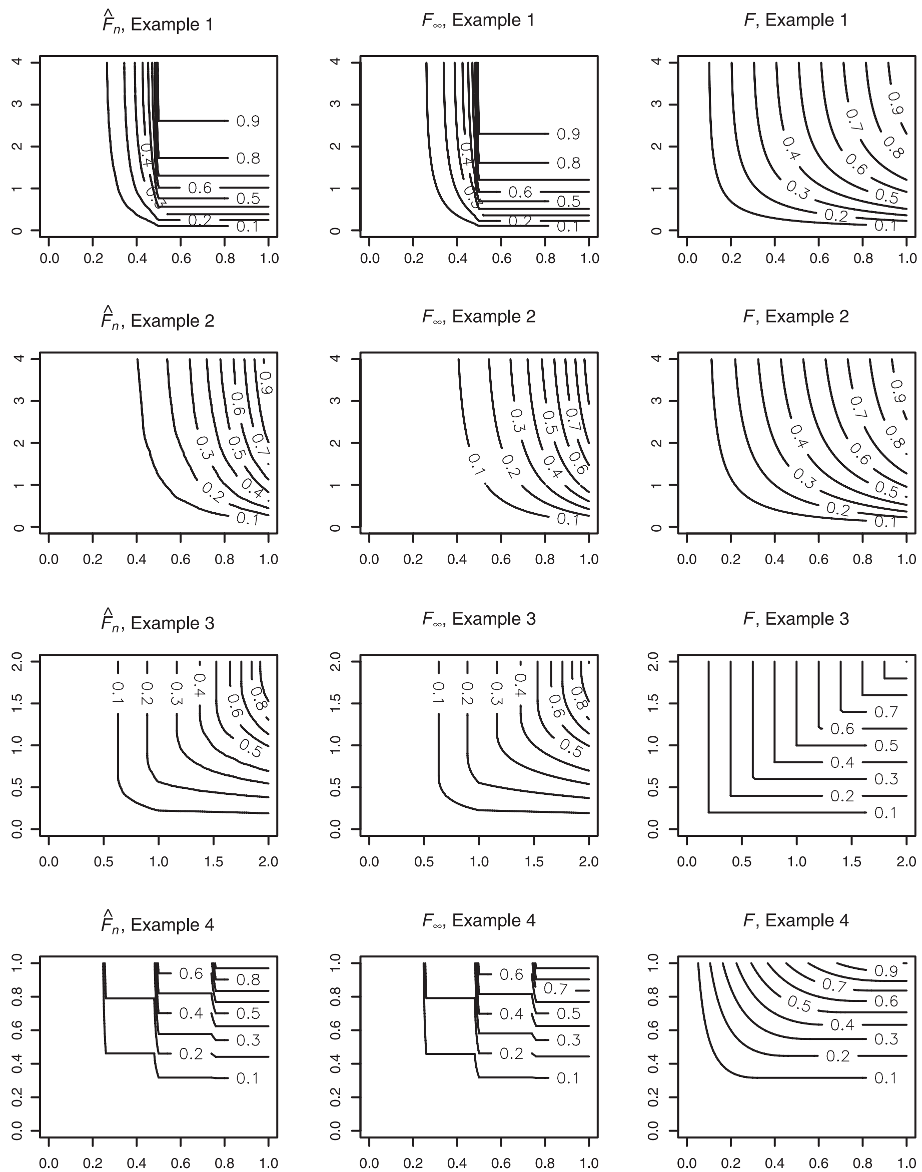

Fig. 2. Contour lines of the bivariate functions $\hat{F}_n^{\ell}$ (left column), F_∞^{ℓ} (middle column) and F_0 (right column) for Examples 1–4. All functions were computed on an equidistant grid with grid size 0.02, and sample size $n = 10,000$.

6. Discussion

We studied the MLE of the bivariate distribution of an interval-censored survival time and a continuous mark variable. We derived the almost sure limit of the MLE, and showed that the MLE is inconsistent in general. We proposed a simple method to repair the inconsistency, and illustrated the behavior of the inconsistent and repaired MLE in four examples.

We were prompted to investigate consistency of the MLE in the interval-censored continuous-mark model, since the observed sets in this model can take the form of line segments.

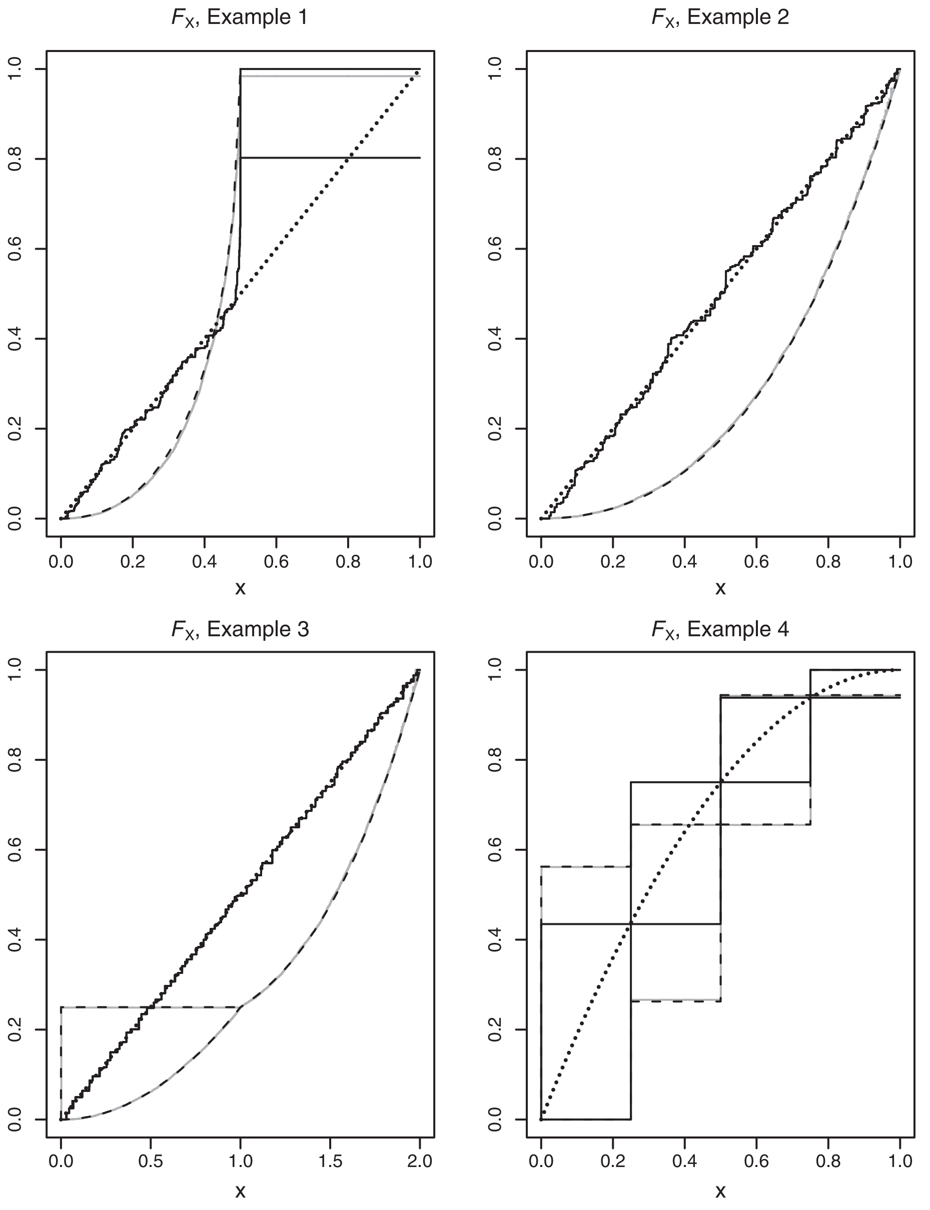

Fig. 3. Estimation of F_{0X} in Examples 1–4. Dotted: the true underlying distribution F_{0X}. Solid grey: the MLEs $\hat{F}^{\ell}_{Xn}$ and $\hat{F}^{u}_{Xn}$. Dashed: the limits $F^{\ell}_{X\infty}$ and $F^{u}_{X\infty}$ of the MLE. Solid black: the repaired MLEs $\tilde{F}^{\ell}_{Xn}$ and $\tilde{F}^{u}_{Xn}$, using the equidistant grid with $K=20$ shown in Figure 4. In all cases $n=10,000$.

Such line segments are an indicator of consistency problems for the MLE, since the MLE for bivariate censored data has been found to be inconsistent before when such line segments were present [Van der Laan (1996) and Maathuis (2003, section 6.2)]. In this sense our results do not come as a surprise, and they confirm the idea that the presence of line segments is indicative of consistency problems of the MLE.

There are, however, interesting differences in the underlying reasons for inconsistency in the above mentioned models. The inconsistency of the MLE in the model considered by

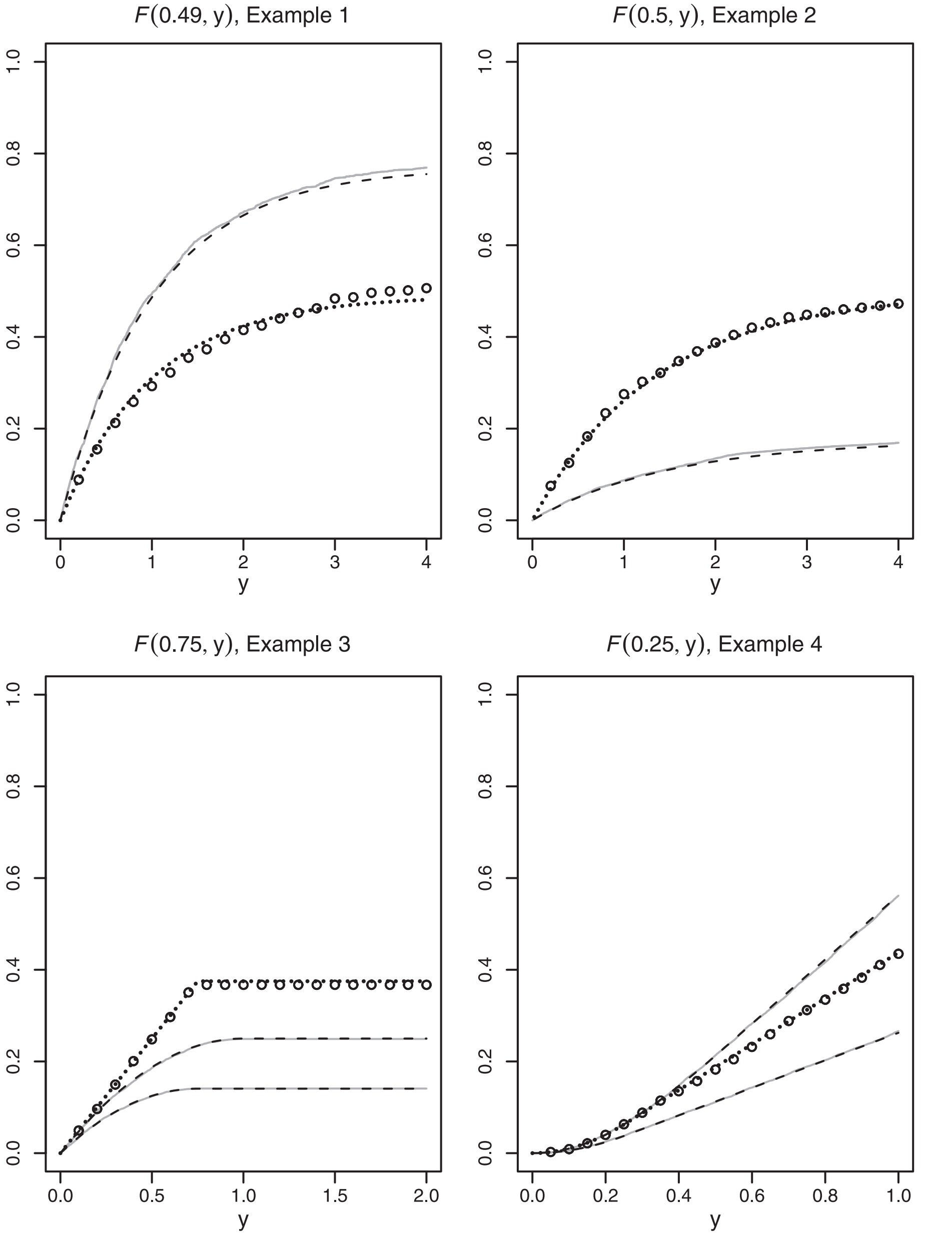

Fig. 4. Estimation of $F_0(x_0, y)$ in Examples 1–4, for fixed x_0 and $y \in \mathbb{R}$. Dotted: the true underlying distribution $F_0(x_0, y)$. Solid grey: the MLEs $\hat{F}_n^{\ell}(x_0, y)$ and $\hat{F}_n^{u}(x_0, y)$. Dashed: the limits $F_{\infty}^{\ell}(x_0, y)$ and $F_{\infty}^{u}(x_0, y)$ of the MLE. Circles: the repaired MLE $\tilde{F}_n^{\ell}(x_0, y) = \tilde{F}_n^{u}(x_0, y)$, using an equidistant grid with $K = 20$. In all cases $n = 10,000$.

Maathuis (2003) could be explained by representational non-uniqueness of the MLE. This is not the case for the interval-censored continuous-mark model, where the MLE is typically inconsistent even if its limit is fully unique. Rather, the inconsistency in the interval-censored continuous-mark model can be explained by the fact that the cumulative hazard functions that define the MLE in (10) and (11) do not converge to the true underlying cumulative hazard functions.

Finally, we provide a more detailed discussion of the connections between the current paper and the paper by HMG, since these papers have been heavily influenced by each other. HMG started studying the interval-censored continuous-mark model, to analyze data from the first Phase III HIV/AIDS vaccine trial VAX004 (Flynn *et al.*, 2005). We suspected inconsistency of the MLE in this model, and investigated this issue more closely. This study has resulted in the current paper. In turn, our paper has influenced the work of HMG and their analysis of the VAX004 data.

There are also some differences between the models in the two papers. HMG considered a slightly more complicated interval-censored continuous-mark model, assuming that X is mixed-case interval-censored (as discussed in section 1) instead of case k interval-censored. They showed that our results in sections 3 and 4 can be generalized to that situation. Thus, the MLE is typically inconsistent in this model as well, and this inconsistency can be repaired by discretizing the marks. HMG also considered a complication regarding the mark variable Y. In addition to assuming that Y is missing for all individuals who did not experience the failure event, they allowed Y to be missing with some probability $p \in (0, 1)$ for individuals who did experience the failure event. In this case there is no closed form available for the MLE. It is therefore more difficult to study consistency issues, and consistency of the MLE in this model is currently still an open problem. However, due to the presence of line segments we expect inconsistency, and this conjecture is supported by simulation results of HMG. HMG therefore included our repaired MLE in the analysis of the VAX004 data.

Acknowledgements

This research was supported by NSF grant DMS-0203320. We would like to thank Piet Groeneboom and Michael Hudgens for helpful discussions and comments. We also owe thanks to an anonymous referee and an associate editor for useful suggestion concerning the presentation of our results.

References

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3**, 15–27.

Aalen, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Statist.* **6**, 534–545.

Bahadur, R. R. (1958). Examples of inconsistency of maximum likelihood estimates. *Sankhyā* **20**, 207–210.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. & Brunk, H. D. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression*. John Wiley & Sons, New York.

Barron, A., Birgé, L. & Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113**, 301–413.

Birgé, L. & Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for lucien le*
3 *cam*, 55–87, Springer, New York.

Boyles, R. A., Marshall, A. W. & Proschan, F. (1985). Inconsistency of the maximum likelihood estimator of a distribution having increasing failure rate average. *Ann. Statist.* **13**, 413–417.

Ferguson, T. S. (1982). An inconsistent maximum likelihood estimate. *J. Amer. Statist. Assoc.* **77**, 831–834.

Flynn, N. M., Forthal, D. N., Harro, C. D., Judson, F. N., Mayer, K. H. & Para, M. F. (2005). The rgp120 HIV Vaccine Study Group. Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J. Infect. Dis.* **191**, 654–665.

Gentleman, R. & Vandal, A. C. (2002). Nonparametric estimation of the bivariate CDF for arbitrarily censored data. *Can. J. Statist.* **30**, 557–571.

Ghosh, M. & Yang, M.-C. (1995). Inconsistent MLE's for multiple-choice questionnaire Rasch models. *Parisankhyan Samikkha* **2**, 31–40.

Gill, R. D. & Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18**, 1501–1555.

Groeneboom, P., Maathuis, M. H. & Wellner, J. A. (2006a). Current status data with competing risks: consistency and rates of convergence of the MLE. Technical Report 500, Department of Statistics, University of Washington. Submitted to Annals of Statistics. Available at http://www.stat.washington.edu/marloes/papers.

Groeneboom, P., Maathuis, M. H. & Wellner, J. A. (2006b). Current status data with competing risks: limiting distribution of the MLE. Technical Report 501, Department of Statistics, University of Washington. Submitted to Annals of Statistics. Available at http://www.stat.washington.edu/marloes/papers.

Groeneboom, P. & Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Birkhäuser Verlag, Basel.

Gupta, A. K., Székely, G. J. & Zsigri, G. (1999). An inconsistent location MLE. *Math. Methods Statist.* **8**, 119–120.

Huang, Y. & Louis, T. A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika* **85**, 785–798.

Hudgens, M. G., Maathuis, M. H. & Gilbert, P. B. (2007). Nonparametric estimation of the joint distribution of a survival time subject to interval censoring and a continuous mark variable. *Biometrics*,
4 in press.

Hudgens, M. G., Satten, G. A. & Longini, I. M. (2001). Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics* **57**, 74–80.

Jewell, N. P. & Kalbfleisch, J. D. (2004). Maximum likelihood estimation of ordered multinomial parameters. *Biostatistics* **5**, 291–306.

Jewell, N. P., Van der Laan, M. J. & Henneman, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika* **90**, 183–197.

Kalbfleisch, J. D. & Prentice, R. L. (1980). *The statistical analysis of failure time data*. Wiley, New York.

Le Cam, L. (1990). Maximum likelihood: an introduction. *Int. Statist. Rev.* **58**, 153–171.

Maathuis, M. H. (2003). *Nonparametric maximum likelihood estimation for bivariate censored data*. Master's thesis, Delft University of Technology, The Netherlands. Available at http://www.stat.washington.edu/marloes/papers.

Maathuis, M. H. (2005). Reduction algorithm for the MLE for the distribution function of bivariate interval censored data. *J. Comp. Graph. Statist.* **14**, 352–362.

Maathuis, M. H. (2006). *Nonparametric estimation for current status data with competing risks*. Ph.D.
5 thesis, University of Washington.

Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32.

Pan, W. & Chappell, R. (1999). A note on inconsistency of NPMLE of the distribution function from left truncated and case I interval censored data. *Lifetime Data Anal.* **5**, 281–291.

Peto, R. (1973). Experimental survival curves for interval-censored data. *App. Stat.* **22**, 86–91.

Politis, D. N., Romano, J. P. & Wolf, M. (1999). *Subsampling*. Springer-Verlag, New York.

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

Shorack, G. R. & Wellner, J. A. (1986). *Empirical processes with applications to statistics*. John Wiley & Sons, New York.

Sun, J. (2006). *The statistical analysis of interval-censored failure time data*. Springer, New York.

Tsai, W.-Y., Leurgans, S. & Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14**, 1351–1365.

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

Van der Laan, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.* **24**, 596–627.

Van der Vaart, A. W. & Wellner, J. A. (2000). Preservation theorems for Glivenko-Cantelli and uniform
6 Glivenko-Cantelli classes. In *High dimensional probability II*, 115–133, Birkhäuser, Boston.

Wellner, J. A. (1995). Interval censoring, case 2: alternative hypotheses. In *Analysis of censored data*, Vol. 27 *of IMS Lecture Notes – Monograph Series*, 271–291, IMS, Hayward.

Wong, G. Y. & Yu, Q. (1999). Generalized MLE of a joint distribution function with multivariate interval-censored data. *J. Multivariate Anal.* **69**, 155–166.

Received October 2006, in final form April 2007

M. H. Maathuis, University of Washington, Department of Statistics, Campus Box 354322, Seattle, WA 98195-4322, USA.
E-mail: marloes@stat.washington.edu

Appendix

This section contains several technical lemmas and proofs that are needed for the results in section 3. Lemma 1 gives the almost sure limits H, V and V_X of the processes $\mathbb{H}_n$, $\mathbb{V}_n$, $\mathbb{V}_{Xn}$ that were defined in (12). Lemma 2 provides the almost sure limits Λ_∞ and $\Lambda_{X\infty}$ of the processes $\hat{\Lambda}_n$ and $\hat{\Lambda}_{Xn}$ that were defined in (13). Lemma 3 gives the almost sure limits F_∞^ℓ and $F_{X\infty}^\ell$ of the MLEs $\hat{F}_n^\ell$ and $\hat{F}_{Xn}^\ell$ that were given in (10) and (11). Corollary 3 provides an alternative way to express F_∞^ℓ. Corollaries 4 and 5 specialize this result to two special cases, namely the case that X and Y are independent, and the case that X is subject to current status censoring. Finally, we provide a proof of Corollary 2.

Lemma 1
For $I \subseteq \mathbb{R}^d$ with $d \geq 1$, and let $\mathcal{D}(I)$ be the space of cadlag functions on I. Furthermore, let $\|\cdot\|_\infty$ be the supremum norm on $(\mathcal{D}(\mathbb{R}_+), \mathcal{D}(\mathbb{R}_+), \mathcal{D}(\mathbb{R}_+ \times \mathbb{R}))$. Then

$$\|(\mathbb{H}_n - H, \mathbb{V}_{Xn} - V_X, \mathbb{V}_n - V)\|_\infty \to_{\text{a.s.}} 0, \tag{21}$$

where

$$V(x,y) = \sum_{j=1}^k \int_{[0,x]} F_0(t,y)\,\mathrm{d}G_j(t) - \sum_{j=2}^k \int_{0\leq s\leq t\leq x} F_0(s,y)\,\mathrm{d}G_{j-1,j}(s,t), \tag{22}$$

$$V_X(x) = \sum_{j=1}^k \int_{[0,x]} F_{0X}(t)\,\mathrm{d}G_j(t) - \sum_{j=2}^k \int_{0\leq s\leq t\leq x} F_{0X}(s)\,\mathrm{d}G_{j-1,j}(s,t), \tag{23}$$

$$H(x) = V_X(x) + \int_{[0,x]} \{1 - F_{0X}(s)\}\,\mathrm{d}G_k(s), \tag{24}$$

and $G_{j-1,j}$ and G_k are defined in the beginning of section 2.2.

Proof. Equation (21) follows immediately from the Glivenko–Cantelli theorem, with $H(x) = E(1\{U \leq x\})$, $V(x,y) = E(\Delta_+ 1\{U \leq x, Z \leq y\})$ and $V_X(x) = V(x,\infty) = E(\Delta_+ 1\{U \leq x\})$. We now express H, V and V_X in terms of F_0 and G. Note that the events $[\Delta_j = 1]$, $j = 1,\ldots,k+1$, are disjoint. Furthermore, note that $U = T_j$ and $Z = Y$ on $[\Delta_j = 1]$, $j = 1,\ldots,k$, and $U = T_k$ on $[\Delta_{k+1} = 1]$. Hence,

$$\begin{aligned} V(x,y) &= E(\Delta_+ 1\{U \leq x, Z \leq y\}) = \sum_{j=1}^k P(\Delta_j = 1, Y \leq y, T_j \leq x) \\ &= \sum_{j=1}^k P(X \in (T_{j-1}, T_j], Y \leq y, T_j \leq x) \\ &= \sum_{j=1}^k \int_{0\leq s\leq t\leq x} \{F_0(t,y) - F_0(s,y)\}\,\mathrm{d}G_{j-1,j}(s,t). \end{aligned}$$

Using $T_0=0$, $X>0$ and $G(\{0<T_1<\cdots<T_k\})=1$, this can be written as

$$\sum_{j=1}^{k}\int_{[0,x]} F_0(t,y)\,\mathrm{d}G_j(t)-\sum_{j=2}^{k}\int_{0\le s\le t\le x} F_0(s,y)\,\mathrm{d}G_{j-1,j}(s,t).$$

Taking $y=\infty$ yields the expression for $V_X(x)$. The expression for H follows similarly, using

$$H(x)=E1\{U\le x\}=\sum_{j=1}^{k}P(\Delta_j=1,T_j\le x)+P(\Delta_{k+1}=1,T_k\le x).$$

Lemma 2
Let $\|\cdot\|_\infty$ *be the supremum norm on* $(\mathcal{D}[0,\tau],\mathcal{D}([0,\tau]\times\mathbb{R}))$. *Then*

$$\|(\hat\Lambda_{Xn}-\Lambda_{X\infty},\hat\Lambda_n-\Lambda_\infty)\|_\infty\to_{\text{a.s.}}0,$$

where

$$\Lambda_\infty(x,y)=\int_{[0,x]}\frac{V(\mathrm{d}s,y)}{1-H(s-)},\quad x\in[0,\tau],y\in\mathbb{R},\tag{25}$$

$$\Lambda_{X\infty}(x)=\Lambda_\infty(x,\infty)=\int_{[0,x]}\frac{V_X(\mathrm{d}s)}{1-H(s-)},\quad x\in[0,\tau].\tag{26}$$

Proof. This proof is similar to the discussion on page 1536 of Gill & Johansen (1990). For all $x\ge 0$, let $\mathbb{H}_n^-(x)\equiv\mathbb{H}_n(x-)$. Consider the mappings

$$(\mathbb{H}_n^-,\mathbb{V}_{Xn},\mathbb{V}_n)\to(\{1-\mathbb{H}_n^-\}^{-1},\mathbb{V}_{Xn},\mathbb{V}_n)\to(\hat\Lambda_{Xn},\hat\Lambda_n)$$

on the spaces

$$\begin{aligned}(\mathcal{D}_-[0,\tau],\mathcal{D}[0,\tau],\mathcal{D}([0,\tau]\times\mathbb{R}))&\to(\mathcal{D}_-[0,\tau],\mathcal{D}[0,\tau],\mathcal{D}([0,\tau]\times\mathbb{R}))\\&\to(\mathcal{D}[0,\tau],\mathcal{D}([0,\tau]\times\mathbb{R})),\end{aligned}$$

where $\mathcal{D}_-(0,\tau]$ is the space of 'caglad' (left-continuous with right limits) functions on $(0,\tau]$. The first mapping is continuous with respect to the supremum norm when we restrict the domain of its first argument to elements of $\mathcal{D}_-[0,\tau]$ that are bounded by say $\{1+H(\tau)\}/2<1$. Strong consistency of $\mathbb{H}_n^-$ ensures that it satisfies this bound with probability one for n large enough. The second mapping is continuous with respect to the supremum norm by the Helly–Bray lemma. Combining the continuity of these mappings with Lemma 1 yields the result of the theorem. □

Lemma 3
Let $\|\cdot\|_\infty$ *be the supremum norm on* $(D[0,\tau],\mathcal{D}([0,\tau]\times\mathbb{R}))$. *Then*

$$\|(\hat F^\ell_{Xn}-F^\ell_{X\infty},\hat F^\ell_n-F^\ell_\infty)\|_\infty\to_{\text{a.s.}}0,$$

where

$$F^\ell_{X\infty}(x)=1-\prod_{s\le x}\{1-\Lambda_{X\infty}(\mathrm{d}s)\},\tag{27}$$

$$F^\ell_\infty(x,y)=\int_{u\le x}\prod_{s<u}\{1-\Lambda_{X\infty}(\mathrm{d}s)\}\Lambda_\infty(\mathrm{d}u,y).\tag{28}$$

Proof. To derive the almost sure limit of $\hat{F}^{\ell}_{Xn}$, consider the mapping

$$\hat{\Lambda}_{Xn} \rightarrow \prod_{s \leq x} \{1 - \hat{\Lambda}_{Xn}(\mathrm{d}s)\} = 1 - \hat{F}^{\ell}_{Xn}(x) \tag{29}$$

on the space $\mathcal{D}[0, \tau]$ to itself. This mapping is continuous with respect to the supremum norm when its domain is restricted to functions of uniformly bounded variation (Gill & Johansen, 1990, theorem 7). Note that, for $s \in [0, \tau]$, $\hat{\Lambda}_{Xn}(s) \leq 1/\{1 - \mathbb{H}_n(\tau)\} < 2/\{1 - H(\tau)\}$ with probability one for n large enough. Together with the monotonicity of $\hat{\Lambda}_{Xn}$ this implies that with probability one $\hat{\Lambda}_{Xn}$ is of uniformly bounded variation on $[0, \tau]$, for n large enough. The almost sure limit of $\widehat{F}^{\ell}_{Xn}$ now follows by combining Lemma 2 and the continuity of (29).

To derive the almost sure limit of $\hat{F}^{\ell}_{n}$ consider the mapping

$$(\hat{\Lambda}_{Xn}, \hat{\Lambda}_n) \rightarrow \int_{u \leq x} \prod_{s < u} \{1 - \hat{\Lambda}_{Xn}(\mathrm{d}s)\} \hat{\Lambda}_n(\mathrm{d}u, y) = \hat{F}^{\ell}_n(x, y)$$

on the space $(\mathcal{D}[0, \tau], \mathcal{D}([0, \tau] \times \mathbb{R}))$ to $\mathcal{D}([0, \tau] \times \mathbb{R})$. This mapping is continuous with respect to the supremum norm when its domain is restricted to functions of uniformly bounded variation Huang & Louis, 1998, theorem 1). Note that $\hat{\Lambda}_n(x, y) \leq \hat{\Lambda}_{Xn}(x)$, so that with probability one the pair $(\hat{\Lambda}_n, \hat{\Lambda}_{Xn})$ is uniformly bounded for n large enough. The result then follows as in the first part of the proof.

Corollary 3
For $x \in [0, \tau], y \in \mathbb{R}$, we can write

$$F^{\ell}_{\infty}(x, y) = \int_{[0,x]} \frac{\Lambda_{\infty}(\mathrm{d}s, y)}{\Lambda_{X\infty}(\mathrm{d}s)} \mathrm{d}F^{\ell}_{X\infty}(s) = \int_{[0,x]} \frac{V(\mathrm{d}s, y)}{V_X(\mathrm{d}s)} \mathrm{d}F^{\ell}_{X\infty}(s). \tag{30}$$

Proof. Combining equations (27) and (28) yields

$$F^{\ell}_{\infty}(x, y) = \int_{[0,x]} \{1 - F^{\ell}_{X\infty}(s-)\} \Lambda_{\infty}(\mathrm{d}s, y). \tag{31}$$

Taking $y = \infty$ gives $F^{\ell}_{X\infty}(x) = F^{\ell}_{\infty}(x, \infty) = \int_{[0,x]} \{1 - F^{\ell}_{X\infty}(s-)\} \Lambda_{X\infty}(\mathrm{d}s)$, so that $\mathrm{d}F^{\ell}_{X\infty}(s) = \{1 - F^{\ell}_{X\infty}(s-)\} \Lambda_{X\infty}(\mathrm{d}s)$. Combining this with (31) yields the first equality of (30). The second equality follows from the identities

$$\Lambda_{\infty}(\mathrm{d}s, y) = V(\mathrm{d}s, y)/\{1 - H(s-)\},$$

$$\Lambda_{X\infty}(\mathrm{d}s) = V_X(\mathrm{d}s, y)/\{1 - H(s-)\}.$$

Corollary 4
Let X and Y be independent. Then

$$F^{\ell}_{\infty}(x, y) = F^{\ell}_{X\infty}(x) F_{0Y}(y), \quad x \in [0, \tau], \; y \in \mathbb{R}. \tag{32}$$

Proof. If X and Y are independent, (17) and (18) yield $V(\mathrm{d}s, y) = F_{0Y}(y) V_X(\mathrm{d}s)$. Substituting this into (30) gives the result. □

Corollary 5
Let X be subject to current status censoring ($k = 1$). Then

$$F^{\ell}_{\infty}(x, y) = \int_{[0,x]} P(Y \leq y | X \leq s) \, \mathrm{d}F^{\ell}_{X\infty}(s), \quad x \in [0, \tau], \; y \in \mathbb{R}.$$

Proof. For $k=1$ equations (17) and (18) reduce to $V(\mathrm{d}s, y) = F_0(s, y)\,\mathrm{d}G(s)$ and $V_X(\mathrm{d}s) = F_{0X}(s)\,\mathrm{d}G(s)$. Hence, $V(\mathrm{d}s, y)/V_X(\mathrm{d}s) = F_0(s, y)/F_{0X}(s) = P(Y \le y | X \le s)$. Substituting this into (30) completes the proof. □

Proof of corollary 2

Since the observation times are the order statistics of k i.i.d. uniform random variables, the marginal densities g_j, $j=1,\ldots,k$ and the joint densities $g_{j-1,j}$, $j=2,\ldots,k$ are known (see, for example, Shorack & Wellner, 1986, p. 97). Summing them over j yields:

$$\sum_{j=1}^{k} g_j(t) = \frac{k}{\theta} 1_{[0,\theta]}(t) \sum_{j-1=0}^{k-1} \binom{k-1}{j-1} \left(\frac{t}{\theta}\right)^{j-1} \left(1 - \frac{t}{\theta}\right)^{k-1-(j-1)} = \frac{k}{\theta} 1_{[0,\theta]}(t),$$

$$\sum_{j=2}^{k} g_{j-1,j}(s, t) = \frac{k(k-1)}{\theta^2} 1_{[0 \le s \le t \le \theta]} \left(1 - \frac{t-s}{\theta}\right)^{k-2}.$$

Let $x < \theta$. Plugging the above expressions for g_j and $g_{j-1,j}$ into (22), and using Fubini's theorem to rewrite the second term of (22), we get

$$\begin{aligned} V^k(x, y) &= \frac{k}{\theta} \int_{[0,x]} F_0(t, y)\,\mathrm{d}t - \int\int_{0 \le s \le t \le x} F_0(s, y) \frac{k(k-1)}{\theta^2} \left(1 - \frac{t-s}{\theta}\right)^{k-2} \mathrm{d}s\,\mathrm{d}t \\ &= \frac{k}{\theta} \int_{[0,x]} F_0(s, y) \left(1 - \frac{x-s}{\theta}\right)^{k-1} \mathrm{d}s = \int_{[0,x]} F_0(s, y)\,\mathrm{d}Q_x^k(s), \end{aligned}$$

where, for $s \le x$,

$$Q_x^k(s) = \int_0^s \frac{k}{\theta} \left(1 - \frac{x-r}{\theta}\right)^{k-1} \mathrm{d}r = \left(1 - \frac{x-s}{\theta}\right)^k - \left(1 - \frac{x}{\theta}\right)^k.$$

Thus, as $k \to \infty$, $Q_x^k(s)$ converges weakly to the distribution function with mass 1 at x. Plugging in $y = \infty$ in $V^k(x, y)$ yields $V_X^k(x) = \int_{[0,x]} F_{0X}(s)\,\mathrm{d}Q_x^k(s)$. Furthermore, plugging in the expressions for V_X^k and G_k in (24) gives

$$H^k(x) = \int_{[0,x]} F_{0X}(s)\,\mathrm{d}Q_x^k(s) + \int_{[0,x]} (1 - F_{0X}(s)) \frac{k}{\theta} \left(\frac{s}{\theta}\right)^{k-1} \mathrm{d}s.f$$

Hence, for $x < \theta$ we have $V^k(x, y) \to F_0(x, y)$, $V_X^k(x) \to F_{0X}(x)$ and $1 - H^k(x) \to 1 - F_{0X}(x)$ as $k \to \infty$ for continuity points of the limits. The corollary then follows from the extended Helly–Bray theorem. □

Author Query Form

Journal: SJOS

Article: 06-172

Dear Author,

During the preparation of your manuscript for publication, the questions listed below have arisen. The numbers pertain to the numbers in the margin of the proof. Please attend to these matters and return the form with this proof.

Many thanks for your assistance.

Refs.	Query	Remarks
Q1	Au: Please check if the hyphenations of interval-censored, right censored, and Continuous-mark (model/dak) are ok.	
Q2	Au: Please provide forenames in full.	
Q3	Au: Provide editors for Birge & Massout, 1997	
Q4	Au: Please update Hudgens *et al.*, 2007.	
Q5	Au: Provide location of university.	
Q6	Provide editors for Van der Vaart & Wellner, 2000.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	
Insert in text the matter indicated in the margin		New matter followed by or
Delete	/ through single character, rule or underline or through all characters to be deleted	or
Substitute character or substitute part of one or more word(s)	/ through letter or through characters	new character / or new characters /
Change to italics	— under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	/ through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	,
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	/ through character or where required	
Reduce space between characters or words	between characters or words affected	