Electronic Journal of Statistics
Vol. 7 (2013) 1783–1805
ISSN: 1935-7524
DOI: 10.1214/13-EJS824

The bivariate current status model

Piet Groeneboom

Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
e-mail: P.Groeneboom@tudelft.nl, *url:* http://dutiosc.twi.tudelft.nl/~pietg/

Abstract: For the univariate current status and, more generally, the interval censoring model, distribution theory has been developed for the maximum likelihood estimator (MLE) and smoothed maximum likelihood estimator (SMLE) of the unknown distribution function, see, e.g., [10, 7, 4, 5, 6, 9, 13] and [11]. For the bivariate current status and interval censoring models distribution theory of this type is still absent and even the rate at which we can expect reasonable estimators to converge is unknown.

We define a purely discrete plug-in estimator of the distribution function which locally converges at rate $n^{1/3}$, and derive its (normal) limit distribution. Unlike the MLE or SMLE, this estimator is not a proper distribution function. Since the estimator is purely discrete, it demonstrates that the $n^{1/3}$ convergence rate is in principle possible for the MLE, but whether this actually holds for the MLE is still an open problem.

We compare the behavior of the plug-in estimator with the behavior of the MLE on a sieve and the SMLE in a simulation study. This indicates that the plug-in estimator and the SMLE have a smaller variance but a larger bias than the sieved MLE. The SMLE is conjectured to have a $n^{1/3}$-rate of convergence if we use bandwidths of order $n^{-1/6}$. We derive its (normal) limit distribution, using this assumption. Finally, we demonstrate the behavior of the MLE and SMLE for the bivariate interval censored data of [1], which have been discussed by many authors, see e.g., [18, 3, 2] and [15].

AMS 2000 subject classifications: Primary 62G05, 62N01; secondary 62G20.

Keywords and phrases: Bivariate current status, bivariate interval censoring, maximum likelihood estimators, maximum smoothed likelihood estimators, cube root n estimation, asymptotic distribution.

Received September 2012.

1. Introduction

We consider the bivariate current status model, also called the bivariate interval censoring, case 1, model. This means that our observations consist of a quadruple $(T, U, \Delta_1, \Delta_2)$, where

$$\Delta_1 = 1_{\{X \le T\}},\, \Delta_2 = 1_{\{Y \le U\}}, \tag{1.1}$$

and (X, Y) is independent of the observation (T, U). We want to estimate the distribution function F_0 of the 'hidden' random vector (X, Y).

A maximum likelihood estimator $\hat{F}_n$ of F_0, the distribution function of (X,Y), maximizes the expression

$$\sum_{i=1}^{n}\{\Delta_{i1}\Delta_{i2}\log F(T_i,U_i)+\Delta_{i1}(1-\Delta_{i2})\log\{F(T_i,\infty)-F(T_i,U_i)\}\\ +(1-\Delta_{i1})\Delta_{i2}\log\{F(\infty,U_i)-F(T_i,U_i)\}\\ +(1-\Delta_{i1})(1-\Delta_{i2})\log\{1-F(\infty,U_i)-F(T_i,\infty)+F(T_i,U_i)\}\}$$

over all bivariate distribution functions F. Another formulation is that $\hat{F}_n$ maximizees

$$\int\delta_1\delta_2\log F(u,v)\,d\mathbb{P}_n+\int\delta_1(1-\delta_2)\log\{F_1(u)-F(u,v)\}\,d\mathbb{P}_n\\ +\int(1-\delta_1)\delta_2\log\{F_2(v)-F(u,v)\}\,d\mathbb{P}_n\\ +\int(1-\delta_1)(1-\delta_2)\log\{1-F_1(u)-F_2(v)+F(u,v)\}\,d\mathbb{P}_n$$

over F, where F_1 and F_2 are the first and second marginal dfs of F, respectively, and $\mathbb{P}_n$ is the empirical measure of the observations $(T_i,U_i,\Delta_{i1},\Delta_{i2})$, $i=1,\dots,n$.

One looks for a solution of the form

$$\hat{F}_n=\sum_{j=1}^{m}\alpha_j 1_{[\tau_j,\infty)},\ \sum_{j=1}^{m}\alpha_j\le 1,\ \alpha_j>0,\ 1\le j\le m=m_n,$$

where we denote by $[\tau_j,\infty)$ an infinite rectangle $[t_{j1},\infty)\times[t_{j2},\infty)$, where $\tau_j=(t_{j1},t_{j2})$. Then the (Fenchel or Kuhn-Tucker) duality conditions for the solution are:

$$\int_{[\mathbf{t},\infty)}\frac{\delta_1\delta_2}{F(u,v)}\,d\mathbb{P}_n+\int_{[t_1,\infty)\times[0,t_2)}\frac{\delta_1(1-\delta_2)}{F_1(u)-F(u,v)}\,d\mathbb{P}_n\\ +\int_{[0,t_1)\times[t_2,\infty)}\frac{(1-\delta_1)\delta_2}{F_2(v)-F(u,v)}\,d\mathbb{P}_n\\ +\int_{[0,t_1)\times[0,t_2)}\frac{(1-\delta_1)(1-\delta_2)}{1-F_1(u)-F_2(v)+F(u,v)}\,d\mathbb{P}_n\\ \le 1, \tag{1.2}$$

for all $\mathbf{t}=(t_1,t_2)\in\mathbb{R}^2$, where $\mathbb{P}_n$ is the empirical df of the observations $(u_i,v_i,\delta_{i1},\delta_{i2})$. We must have equality in (1.2), if $\mathbf{t}=\tau_j$, $j=1,\dots,m_n$, is a point of mass of the solution (i.e., the constraints are active!), where the rectangles $[\tau_j,\infty)$ are the generators of the solution.

An R package, called ‘MLEcens’ is available for computing the MLE. The algorithm determines the maximal intersection rectangles where the MLE has mass via a preliminary reduction algorithm, and next computes the mass of the

MLE in these rectangles, using the support reduction algorithm of [12]. The reduction algorithm is described in [15]. The R package uses as an example a data set, studied in [1], which is actually not of the current status type but has interval censoring, case 2, data. We shall discuss these data in section 5. The MLE for this data set is also discussed in section 7.3.3 of [18], who also refers to [3] and [2] for discussions of the computation of the MLE for this data set. The computation of the rectangles where the MLE puts mass is also treated in [17], where also minimax lower bounds for the estimation rate of the MLE and consistency of the MLE are derived.

There is an extensive discussion on where to put the mass, once one has determined rectangles which can have positive mass, see, e.g. [18], section 7.3, [3, 2] and [15]. The algorithm for computing these rectangles, proposed in [15], seems at present to be the fastest.

It is in our view somewhat doubtful whether all the energy spent on computing these maximal intersection rectangles and the ensuing question of whether one should place the mass of the MLE at the left lower corner or the right upper corner of the rectangles is really worth the effort. One could also specify in advance a set of points where one allows mass to be placed. In this way one obtains an MLE on a sieve, where the sieve consists of distributions having discrete mass at these points. The bottleneck in the computation of the MLE for the bivariate interval censoring problem is not the determination of the maximal intersection rectangles, but the computation of the mass the MLE puts on these rectangles, since there usually are very many!

The latter phenomenon shows up in particular in simulations. As an example, simulating data from the distribution with density $f(x, y) = x + y$ on the unit square, with a uniform observation distribution, we got for sample size $n = 5000$ about $5 \cdot 10^5$ possible rectangles where the masses could be placed, which is (at present) an almost prohibitive number if one wants to do simulations of the limit behavior of the MLE on an ordinary table computer or laptop. Ultimately, the discussion on these matters should in our view be determined by insights into the distribution theory of the MLE or the MLE on the chosen sieve.

In section 2 we show that a purely discrete plug-in estimator locally attains the $n^{1/3}$ rate. We also determine its asymptotic (normal) distribution. In section 3 we study the local limit behavior of the smoothed maximum likelihood estimator and derive its asymptotic distribution under the assumption that it can asymptotically can be represented by an integral in the observation space, proceeding along similar lines as in the one-dimensional case (see, e.g., [5, 9] and [8]). Section 4 presents a simulation study, comparing the behavior of the MLE, the SMLE and the plug-in estimator. The results seem to be in accordance with Theorems 2.1 and 3.1 in sections 2 and 3, respectively. Section 5 extends the MLE and SMLE to a more general setting of interval censoring and applies the methods on a data set in [1], which has been discussed by many authors, see e.g., [18, 3, 2] and [15]. The paper ends with some concluding remarks on faster rates for the SMLE, which can be attained if one uses higher order kernels.

2. A purely discrete $n^{1/3}$-rate estimator for the bivariate current status model

Basically, the MLE for the 1-dimensional current status model is the monotone derivative of the cusum diagram

$$(\mathbb{G}_n(t), V_n(t))\,,\ t \in I, \qquad V_n(t) = \int_{u \le t} \delta \, d\mathbb{P}_n(u, \delta),$$

where I is the observation interval and $\mathbb{G}_n$ the empirical distribution function of the observations $T_1, \dots, T_n$. So it can be considered to be a monotone version of the 'derivative' $dV_n(t)/d\mathbb{G}_n(t)$. Note that if we replace V_n and $\mathbb{G}_n$ by their deterministic equivalents, the derivative becomes

$$\frac{F_0(t)g(t)}{g(t)} = F_0(t),$$

so is indeed the object we want to estimate.

For the simplest bivariate current status model, which is sometimes called the 'in-out' model, we only have the information whether the hidden variable is below and to the left of the observation point (T_i, U_i) or not. In this case we could also define

$$V_n(t,u) = \int_{v \le t,\, w \le u} \delta \, d\mathbb{P}_n(v, w, \delta),$$

where $\delta = 1$ represents the situation that the hidden variable is below and to the left of (v, w). If the empirical observation distribution is again denoted by $\mathbb{G}_n$, we this time want to estimate the 'derivative' $dV_n(t,u)/d\mathbb{G}_n(t,u)$, since, replacing V_n and $\mathbb{G}_n$ by their deterministic equivalents, the derivative becomes

$$\frac{F_0(t,u)g(t,u)}{g(t,u)} = F_0(t,u).$$

So we want to find a version of the derivative $dV_n(t,u)/d\mathbb{G}_n(t,u)$, under the (shape) restriction that it is a bivariate distribution function.

However, a natural cusum diagram for this situation does not seem to exist. But we can define a 2-dimensional 'Fenchel process', incorporating the duality conditions for a solution of the optimization problem. Analogously to the 1-dimensional current status model, the Fenchel duality conditions for the isotonic least squares (LS)estimate, minimizing

$$\sum_{i=1}^{n} \left\{\Delta_i - F(T_i, U_i)\right\}^2, \qquad \Delta_i = 1_{\{X_i \le T_i,\, Y_i \le U_i\}}$$

over all bivariate distribution functions F, where the (X_i, Y_i) are the hidden variables, are:

$$\int_{v \ge t,\, w \ge u} \left\{\delta - F(v,w)\right\} \, d\mathbb{P}_n(v, w, \delta) \le 0, \tag{2.1}$$

with equality if (t,u) is a point of mass of the solution. So we have to deal with a process

$$(t,u)\mapsto \int_{v\ge t,\,w\ge u} F(v,w)\,d\mathbb{G}_n(v,w) \tag{2.2}$$

which has to lie above the process

$$(t,u)\mapsto \int_{v\ge t,\,w\ge u} \delta\,d\mathbb{P}_n(v,w,\delta),$$

with points of touch at points of mass of F. Denoting temporarily the process (2.2) by Q_n, we get that the isotonic least squares estimator can (formally) be denoted by $dQ_n(t,u)/d\mathbb{G}_n(t,u)$ (which no longer necessarily coincides with the MLE!). Note, however, that the function Q_n is not necessarily close to a convex or concave function, so here the analogy with 1-dimensional current status breaks down. But it must have the property that its 'derivative' w.r.t. $d\mathbb{G}_n$ must be a distribution function, which is analogous to the fact that the derivative of the convex minorant of the cusum diagram must be a distribution function in the 1-dimensional case.

For the real current status model the situation is more complicated, since we then have to deal with 4 regions instead of 2. From (1.2) we get:

$$\begin{aligned}
&\int_{[\mathbf{t},\infty)} \frac{\delta_1\delta_2}{\hat F_n(u,v)}\,d\,\mathbb{P}_n + \int_{[t_1,\infty)\times[0,t_2)} \frac{\delta_1(1-\delta_2)}{\hat F_{n1}(u)-\hat F_n(u,v)}\,d\,\mathbb{P}_n\\
&+\int_{[0,t_1)\times[t_2,\infty)} \frac{(1-\delta_1)\delta_2}{\hat F_{n2}(v)-\hat F_n(u,v)}\,d\,\mathbb{P}_n\\
&+\int_{[0,t_1)\times[0,t_2)} \frac{(1-\delta_1)(1-\delta_2)}{1-\hat F_{n1}(u)-\hat F_{n2}(v)+\hat F_n(u,v)}\,d\,\mathbb{P}_n\\
&\le 1,
\end{aligned}$$

where $\mathbf{t}=(t_1,t_2)$, with equality if (t_1,t_2) is a point of mass of $\hat F_n$.

It has been conjectured that the MLE in the bivariate current status model converges locally at rate $n^{1/3}$, just as in the 1-dimensional current status model (with smooth underlying distribution functions). [17] proves a minimax lower bound of order $n^{-1/3}$. It would be somewhat surprising if the 1-dimensional rate would be preserved in dimension 2, since in general one gets lower rates for density estimators if the dimension gets up, and the estimation of the distribution function in the current status model is similar to density estimation problems, as argued above.

To show that it is in principle possible to attain the local rate $n^{1/3}$, we construct a purely discrete estimator, converging locally at rate $n^{1/3}$. We restrict ourselves for simplicity to distributions with support $[0,1]^2$ in the remainder of this section, but the generalization to more general rectangles is obvious. We have the following result, which is proved in the Appendix.

Theorem 2.1. *Consider an interior point (t, u), and define the square A_n, with midpoint (t, u), by:*

$$A_n = [t - h_n, t + h_n] \times [u - h_n, u + h_n].$$

Moreover, suppose that the observation distribution G is twice continuously differentiable at (t, u) with a strictly positive density $g(t, u)$ at (t, u), and that F_0 is twice continuously differentiable at (t, u). Moreover, suppose

$$\lim_{n\to\infty} h_n^2 n^{1/3} = c > 0. \tag{2.3}$$

Then the estimator

$$\tilde{F}_n(t, u) \stackrel{def}{=} \frac{\int_{A_n} \delta_1 \delta_2 \, d\mathbb{P}_n(v, w, \delta_1, \delta_2)}{\int_{A_n} d\mathbb{G}_n(v, w)}, \tag{2.4}$$

where $\mathbb{G}_n$ is the empirical distribution function of the observations (T_i, U_i) and $\mathbb{P}_n$ is the empirical distribution function of the observations

$$(T_i, U_i, \Delta_{i1}, \Delta_{i2}), \; i = 1, \dots, n,$$

satisfies:

$$n^{1/3} \left\{ \tilde{F}_n(t, u) - F_0(t, u) \right\} \xrightarrow{\mathcal{D}} N\left(\beta, \sigma^2\right),$$

where $N(\beta, \sigma^2)$ is a normal distribution with first moment

$$\beta = c \left\{ \tfrac{1}{6} \left\{ \partial_1^2 F_0(t, u) + \partial_2^2 F_0(t, u) \right\} + \frac{\partial_1 F_0(t, u) \partial_1 g(t, u) + \partial_2 F_0(t, u) \partial_2 g(t, u)}{3g(t, u)} \right\},$$

and variance

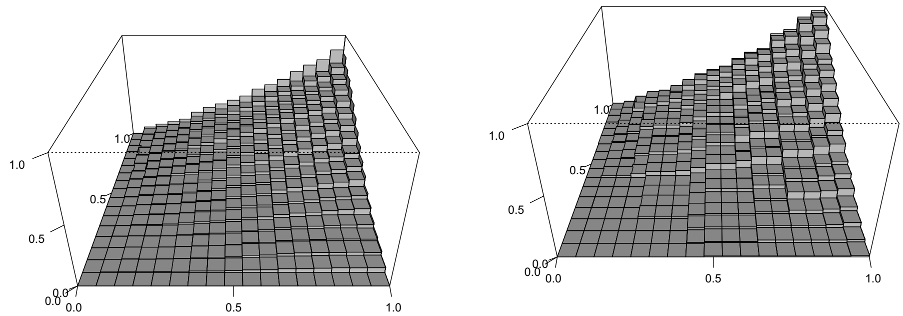
$$\sigma^2 = \frac{F_0(t, u) \{1 - F_0(t, u)\}}{4cg(t, u)}.$$

We now allow as possible points of mass the points $(t_{n1}, u_{n1}), \dots, (t_{n,m_n}, u_{n,m_n})$, running through a rectangular grid, where the distances between the points on the x- and y-axis are of order $n^{-1/3}$, and define the estimate $\tilde{F}_n$ at each point (t_{ni}, u_{ni}) as in Theorem 2.1. Next we define the masses p_{ni} at the points (t_{ni}, u_{ni}) by the equations

$$\sum_{j: t_{nj} \le t_{ni}, \, u_{nj} \le u_{ni}} p_{nj} = \tilde{F}_n(t_{ni}, u_{ni}), \; i = 1, \dots, m_n.$$

Note that the estimate $\tilde{F}_n$ we obtain in this way is not necessarily a distribution function and that the masses p_{nj} can have negative values.

Also note that we get roughly order $n^{1/3} \times n^{1/3}$ equations in this way, which turned out to be solvable, although it is not clear beforehand that the system one gets is non-singular. Nevertheless, one can build up the system from left



(a) Plug-in estimator

(b) MLE on the points of mass of the plug-in estimate

Fig 1. *The plug-in estimator $\tilde{F}_n$ and the MLE, for a sample of size $n = 5000$ of bivariate current status data, where the hidden variables have a distribution with density $f_0(x, y) = x + y$, and the observation distribution is uniform on $[0, 1]^2$.*

and below up to the right and above, where one gets more an more values in the corresponding matrix, so it seems likely that in general the solution exists. This seems a point for further research. A picture of $\tilde{F}_n$, together with the MLE, computed on the sieve of points of mass of the plug-in estimator, is shown in Figure 1. The sieved MLE is a proper (discrete) distribution function, so all its masses are nonnegative.

Assuming the support of the distribution with function F_0 to be $[0, 1]^2$, we treat the points near the upper and right boundary in a special way in computing $\tilde{F}_n$. If, for example $t_{ni} > 1-h_n$, where $h_n \sim n^{-1/6}$ and (t_{ni}, u_{ni}) is a point of the grid, we put the δ_{j1} corresponding to the contribution of observations (T_j, U_j) such that $T_j \geq 2-t_{ni}-h_n$, equal to 1. We treat the second coordinate in the same way. This reduces the bias we otherwise would get, with an underestimation of the distribution function near the right and upper boundary. The idea to treat the points near the boundary in this way is inspired by, but different from, the reflection method proposed by [16]. For points near the left and lower boundary, we follow a similar procedure. If, for example $t < h_n$, where $h_n \sim n^{-1/6}$ and (t, u) is a point of the grid, we put the δ_{j1} corresponding to the contribution of observations (T_j, U_j) such that $T_j \leq h_n - t$, equal to 0. The bias near the boundary is actually of order $O(h_n)$ in this way and we do not attain the $O(h_n^2)$ for interior points, though.

3. The smoothed maximum likelihood estimator

Throughout this section we will assume for simplicity that the support of the distribution of the 'unobservables' is $[0, 1]^2$ and that we want to estimate the corresponding distribution function F_0 on $[0, 1]^2$. The generalization to more general rectangles is obvious.

Let K be a symmetric non-negative kernel, for example the triweight kernel

$$K(x) = \tfrac{35}{32}\left(1-x^2\right)^3 1_{[-1,1]}(x),$$

and $K_h(x) = h^{-1}K(x/h)$, for $h>0$. Moreover, let the integrated kernel $I\!K$ be defined by

$$I\!K(x) = \int_{-\infty}^{x} K(y)\,dy.$$

We follow the approach for the 1-dimensional case, discussed in the references [4] to [9].

At an interior point (t,u), not too close to the boundary, the smoothed maximum likelihood estimator (SMLE) is just defined by

$$\hat F_{nh}^{(SML)}(t,u) = \int I\!K_h(t-v) I\!K_h(u-w)\,d\hat F_n(v,w), \qquad I\!K_h(x) = I\!K(x/h). \tag{3.1}$$

To prevent the negative bias at the right and upper boundary of the support, we also perform a correction by extending the definition of $I\!K$ near the upper and right boundary by:

$$I\!K^b(x) = \int_x^{\infty} K(y)\,dy \tag{3.2}$$

and defining

$$\begin{aligned}&\hat F_{nh}^{(SML)}(t,u)\\ &= \int \left\{I\!K_h(t-v) + I\!K_h^b(2-t-v)\right\}\left\{I\!K_h(u-w) + I\!K_h^b(2-u-w)\right\}\,d\hat F_n(v,w),\end{aligned} \tag{3.3}$$

and $I\!K_h^b(x) = h^{-1} I\!K^b(x/h)$. This definition of the (integrated) boundary kernel is based on the reflection boundary correction method for density estimates, proposed in [16]. Note that the definitions (3.1) and (3.3) coincide if $\max(t,u) \le 1-h$.

We next define the score function in the hidden space:

$$\kappa_{(t,u)}(x,y) = \left\{I\!K_h(t-x) + I\!K_h^b(2-t-y)\right\}\left\{I\!K_h(u-x) + I\!K_h^b(2-u-y)\right\},$$

Scores in the observation space are given by

$$\theta_{F_0}(v,w,\delta_1,\delta_2) = E\left\{a(X,Y)\,\big|\,(T,U,\Delta_1,\Delta_2) = (v,w,\delta_1,\delta_2)\right\}, \tag{3.4}$$

where a is a score in the hidden space. We have, for example

$$E\left\{a(X,Y)\,\big|\,(T,U,\Delta_1,\Delta_2) = (v,w,1,1)\right\} = \frac{\int_{x\le v,\,y\le w} a(x,y)\,dF_0(x,y)}{F_0(v,w)}$$

With this notation, we want to solve the equation

$$\begin{aligned}
&E\left\{\theta_{F_0}(T,U,\Delta_1,\Delta_2)\,\big|\,(X,Y)=(x,y)\right\}\\
&=\int_{v\ge x,\,w\ge y}\theta_{F_0}(v,w,1,1)\,g(v,w)\,dv\,dw+\int_{v\ge x,\,w<y}\theta_{F_0}(v,w,1,0)\,g(v,w)\,dv\,dw\\
&\quad+\int_{v<x,\,w\ge y}\theta_{F_0}(v,w,0,1)\,g(v,w)\,dv\,dw\\
&\qquad+\int_{v<x,\,w<y}\theta_{F_0}(v,w,0,0)\,g(v,w)\,dv\,dw\\
&=\kappa_{(t,u)}(x,y). \qquad (3.5)
\end{aligned}$$

Defining

$$\phi_{F_0}(x,y)=\int_{v\le x,\,w\le y}a(v,w)\,dF_0(v,w), \qquad (3.6)$$

where a is a score function in the hidden space, and differentiating (3.5) w.r.t. x and y, we now obtain the equation:

$$\begin{aligned}
&\frac{\phi_{F_0}(x,y)}{F_0(x,y)}-\frac{\phi_{F_0}(x,1)-\phi_{F_0}(x,y)}{F_0(x,1)-F_0(x,y)}-\frac{\phi_{F_0}(1,y)-\phi_{F_0}(x,y)}{F_0(1,y)-F_0(x,y)}\\
&\qquad\qquad-\frac{\phi_{F_0}(x,1)+\phi_{F_0}(1,y)-\phi_{F_0}(x,y)}{1-F_0(x,1)-F_0(1,y)+F_0(x,y)}\\
&=g(x,y)^{-1}\frac{\partial^2\kappa_{(t,u)}(x,y)}{\partial x\partial y}\\
&=\frac{\{K_h(t-x)+K_h(2-t-x)\}\{K_h(u-y)+K_h(2-u-y)\}}{g(x,y)}
\end{aligned}$$

This equation has the solution

$$\begin{aligned}
&\phi_{F_0}(x,y)\\
&=\frac{\{K_h(t-x)+K_h(2-t-x)\}\{K_h(u-y)+K_h(2-u-y)\}}{g(x,y)}\\
&\quad\cdot\left\{\frac{1}{F_0(x,y)}+\frac{1}{F_0(x,1)-F_0(x,y)}+\frac{1}{F_0(1,y)-F_0(x,y)}\right.\\
&\qquad\qquad\left.+\frac{1}{1-F_0(x,1)-F_0(1,y)+F_0(x,y)}\right\}^{-1} \qquad (3.7)
\end{aligned}$$

Note that the solution satisfies:

$$\phi_{F_0}(1,y)=\phi_{F_0}(x,1)=0,\ x,y\in[0,1].$$

This suggests that the asymptotic behavior of the SMLE is given by:

$$\int\theta_{F_0}(x,y,\delta_1,\delta_2)\,d\left(\mathbb{P}_n-P\right)(x,y,\delta_1,\delta_2), \qquad (3.8)$$

where

$$\begin{aligned}
&\theta_{F_0}(v,w,\delta_1,\delta_2)\\
&= \frac{\delta_1\delta_2\phi_{F_0}(x,y)}{F_0(x,y)} - \frac{\delta_1(1-\delta_2)\phi_{F_0}(x,y)}{F_0(x,1)-F_0(x,y)} - \frac{(1-\delta_1)\delta_2\phi_{F_0}(x,y)}{F_0(1,y)-F_0(x,y)}\\
&\qquad + \frac{(1-\delta_1)(1-\delta_2)\phi_{F_0}(x,y)}{1-F_0(1,y)-F_0(x,1)+F_0(x,y)},
\end{aligned}$$

leading at interior points (t,u) to an asymptotic variance, given by:

$$\begin{aligned}
&\frac{1}{n}\int_{(x,y)\in[0,1]^2}\left\{\frac{1}{F_0(x,y)}+\frac{1}{F_0(x,1)-F_0(x,y)}+\frac{1}{F_0(1,y)-F_0(x,y)}\right.\\
&\qquad\qquad\left.+\frac{1}{1-F_0(1,y)-F_0(x,1)+F_0(x,y)}\right\}^{-1}\\
&\qquad\cdot\frac{\{K_h(t-x)+K_h(2-t-x)\}^2\{K_h(u-y)+K_h(2-u-y)\}^2}{g(x,y)}\,dx\,dy\\
&\sim \left(nh^2\right)^{-1}\left\{\frac{1}{F_0(t,u)}+\frac{1}{F_0(t,1)-F_0(t,u)}+\frac{1}{F_0(1,t)-F_0(t,u)}\right.\\
&\qquad\left.+\frac{1}{1-F_0(1,u)-F_0(t,1)+F_0(t,u)}\right\}^{-1} g(t,u)^{-1}\left\{\int K(v)^2\,dv\right\}^2.
\end{aligned}$$

Assume that $\max(t,u)\le 1-h$. Then the bias is given by:

$$\begin{aligned}
&\int \mathbb{K}_h(t-v)\mathbb{K}_h(u-w)f_0(v,w)\,dv\,dw - F_0(t,u)\\
&= \int\left\{\int K_h(t-v)\int_0^v f_0(x,w)\,dx\,dv\right\}\mathbb{K}_h(u-w)\,dw - F_0(t,u)\\
&= \int K_h(t-v)K_h(u-w)F_0(v,w)\,dv\,dw - F_0(t,u)\\
&= \int K(v)K(w)\left\{F_0(t-hv,u-hw)-F_0(t,u)\right\}\,dv\,dw\\
&= \tfrac12\left\{\partial_1^2F_0(t,u)+\partial_2^2F_0(t,u)\right\}h^2\left\{\int x^2K(x)\,dx\right\}^2 + o\left(h^2\right).
\end{aligned}$$

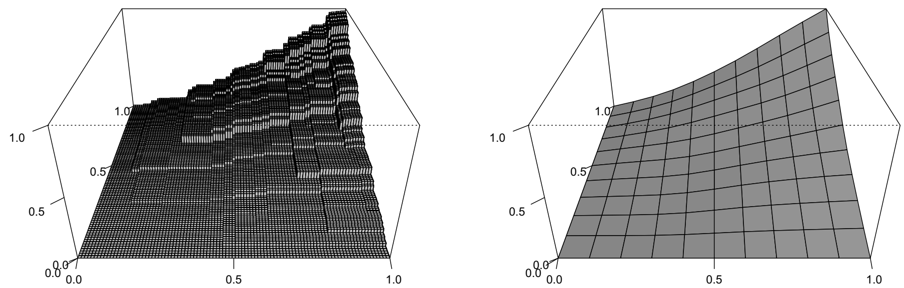
The SMLE is compared with the MLE in Figure 2. Using the assumption that $\hat F_n^{(SML)}(t,u)-F_0(t,u)$ has the asymptotic representation (3.8), we get the following result.

Theorem 3.1 (Conjecture)**.** *Under the conditions of Theorem 2.1 we have for each point $(t,u)\in(0,1)^2$ satisfying these conditions, if $c=\lim_{n\to\infty} n^{1/3}h_n^2$,*

$$n^{1/3}\left\{\hat F_{n,h_n}^{(SML)}(t,u)-F_0(t,u)\right\}\xrightarrow{\mathcal{D}} N\left(\beta,\sigma^2\right),$$

where $N(\beta,\sigma^2)$ is a normal distribution with first moment

$$\beta=\tfrac12 c\left\{\partial_1^2F_0(t,u)+\partial_2^2F_0(t,u)\right\}\left\{\int x^2K(x)\,dx\right\}^2,$$



(a) MLE (b) SMLE

FIG 2. *The MLE and SMLE for for a sample of size* $n = 1000$ *of bivariate current status data, where the hidden variables have a distribution with density* $f_0(x, y) = x + y$*, and the observation distribution is uniform on* $[0, 1]^2$.

and variance

$$\sigma^2 = c^{-1}\left\{\frac{1}{F_0(t,u)} + \frac{1}{F_0(t,1) - F_0(t,u)} + \frac{1}{F_0(1,t) - F_0(t,u)} + \frac{1}{1 - F_0(1,u) - F_0(t,1) + F_0(t,u)}\right\}^{-1} g(t,u)^{-1}\left\{\int K(v)^2\,dv\right\}^2.$$

Remark 3.1. Note that choosing $h_n \asymp n^{-1/6}$ is the asymptotically optimal choice (modulo constants) since the variance is of order $1/(nh_n^2)$ and the bias of order h_n^2, unless the bias is of order $o(h_n^2)$ (as happens when F_0 is the uniform distribution function on $[0, 1]^2$). Also note that the bias term, caused by the interaction of the observation distribution G and the distribution function F_0, which entered into the bias term of Theorem 2.1, plays no role here.

4. A simulation study

In order to compare the behavior of the three estimators, we took 1000 samples from the distribution with distribution function

$$F_0(x, y) = \tfrac{1}{2}xy(x + y),\ x, y \in [0, 1]^2,$$

and generated bivariate current status data from this with respect to the uniform distribution on $[0, 1]^2$. The samples size taken were $n = 100, 500, 1000$ and 5000, respectively. So we have observations (T_i, U_i) from the uniform distribution in $[0, 1]^2$, and for each such pair we get from the corresponding pair (X_i, Y_i), drawn from F_0, independently w.r.t. (T_i, U_i), the indicators

$$\Delta_{i1} = 1_{\{X_i \le T_i\}} \text{ and } \Delta_{i2} = 1_{\{Y_i \le U_i\}}.$$

Since simulations for sample size 5000 with the ‘real’ MLE, based on the maximal intersection rectangles, would have taken prohibitively long, we instead computed the MLE on a sieve, constructed in the following way. For each

Table 1

Estimated values of the standard deviations times $n^{1/3}$ for three estimators of $F_0(t,u)$ at a number of values of (t,u), where $F_0(t,u)=\frac{1}{2}tu(t+u)$. The values corresponding to ∞ are the asymptotic values, deduced from Theorems 2.1 and 3.1. The asymptotic values for the MLE are unknown

	$u=0.6$			
t	n	MLE	SMLE	Plug-in
0.2	100	0.251	0.133	0.185
	500	0.257	0.132	0.152
	1000	0.247	0.128	0.142
	5000	0.246	0.123	0.126
	∞	–	0.130	0.107
0.4	100	0.379	0.212	0.198
	500	0.367	0.194	0.176
	1000	0.360	0.190	0.179
	5000	0.357	0.178	0.156
	∞	–	0.182	0.162
0.6	100	0.475	0.263	0.226
	500	0.412	0.223	0.221
	1000	0.450	0.238	0.215
	5000	0.441	0.218	0.205
	∞	–	0.203	0.206
0.8	100	0.550	0.276	0.290
	500	0.508	0.253	0.265
	1000	0.503	0.255	0.266
	5000	0.514	0.236	0.240
	∞	–	0.183	0.236

(a) $n^{1/3}$ times SD

	$u=0.6$			
t	n	MLE	SMLE	Plug-in
0.2	100	0.003	0.043	0.205
	500	−0.001	0.037	0.263
	1000	−0.037	0.029	0.264
	5000	−0.0002	0.030	0.230
	∞	–	0.044	0.133
0.4	100	−0.004	0.061	0.155
	500	−0.006	0.056	0.169
	1000	−0.022	0.048	0.164
	5000	−0.006	0.049	0.167
	∞	–	0.056	0.166
0.6	100	0.115	0.099	0.015
	500	0.003	0.084	0.202
	1000	−0.036	0.080	0.190
	5000	0.007	0.072	0.188
	∞	–	0.067	0.200
0.8	100	0.118	0.124	−0.152
	500	−0.018	0.112	−0.078
	1000	0.006	0.117	−0.090
	5000	−0.0002	0.116	−0.006
	∞	–	0.078	0.233

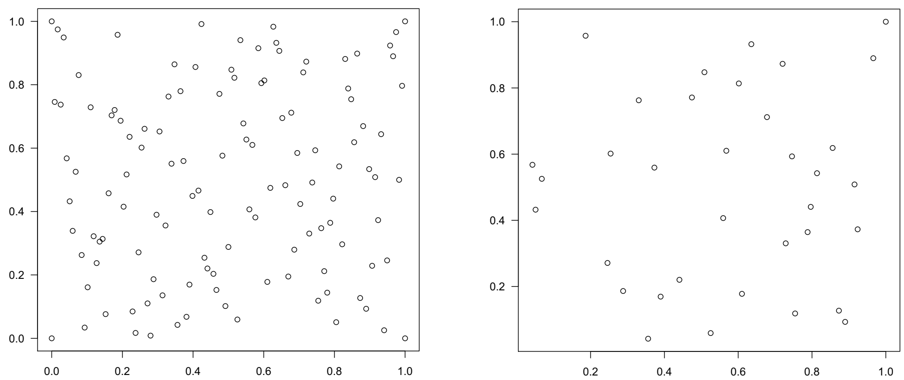
(b) $n^{1/3}$ times bias

sample of size n, we distributed $m_n=\left[n^{2/3}\right]$ points on the unit square, by letting their x- and y-coordinates be multiples of $n^{-1/3}$ and permuting these coordinates randomly, according to the uniform distribution on permutations. Here $[x]$ denotes the largest integer $\le x$.

So we start with order $n^{2/3}$ points on which the sieved MLE can place its mass, to which we add the vertices of the unit square to ensure a finite log likelihood, and compute the MLE which only is allowed to have mass at these points. This set of points is shown in Figure 3, corresponding to sample size $n=1000$, and the reduced set of points on which the MLE actually puts positive mass is also shown in this picture.

The rather different set of points on which the plug-in estimate puts its mass is shown in Figure 4. In fact, inspection of the set of points on which the MLE, based on the maximal intersection rectangles puts its mass, shows that that such sets are rather similar to Figure 3b and not similar to Figure 4. By the rather irregular structure of sets like Figure 3b the bias of the MLE is reduced w.r.t. the plug-in estimator which is constant on squares with sides of order $n^{-1/3}$. We note, however, that the number of points in Figure 3a is the same as in Figure 4 (the number is 121).

We took the bandwidth h_n for the SMLE equal to $n^{-1/6}$ and we also used this as binwidth for the plug-in estimator. Table 1a shows that there is no indication that $n^{1/3}$ times the standard deviation of the MLE is increasing with sample



(a) Initial set of points of possible mass (b) Mass points of MLE

FIG 3. *A set of points on which the MLE is allowed to put its mass and the actual set of mass points of the MLE for a sample of size* $n = 1000$.

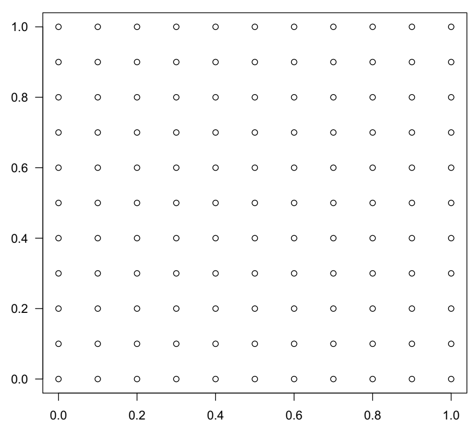


FIG 4. *The set of points on which the plug-in estimator puts its mass for* $n = 1000$.

size, and Table 1b suggests that the bias times $n^{1/3}$ is vanishing (as is also true in the one-dimensional case!), as $n \to \infty$, so the hypothesis that the rate of convergence of the MLE is $n^{1/3}$ is not contradicted.

However, if the rate of the MLE is actually of order $n^{1/3}(\log n)^{-1/3}$, for example, we can probably not detect this in the present way. The theory for the MSLE and plug-in estimators seems to be confirmed by the simulations, in particular at the points $(0.4, 0.6)$ and $(0.6, 0, 6)$, where the boundary effects are still not active. Note that if, for example, $n = 500$, the bandwidths for the SMLE are equal to $500^{-1/6} \approx 0.3549537$, so the boundary kernel starts getting active for the points $(0.2, 0.6)$ and $(0.8, 0.6)$. This is still true for sample size $n = 5000$, where the bandwidth is $5000^{-1/6} \approx 0.2418271$. It is clear from Table 1a that

Table 2
Estimated values of the standard deviations times $n^{1/3}$ for three estimators of $F_0(t,u)$ at a number of values of (t,u), where $F_0(t,u)=tu$. The values corresponding to ∞ are the asymptotic values, deduced from Theorems 2.1 and 3.1. The asymptotic values for the MLE are unknown

	$u=0.6$			
t	n	MLE	SMLE	Plug-in
0.2	100	0.371	0.199	0.388
	500	0.367	0.173	0.286
	1000	0.363	0.145	0.262
	5000	0.360	0.121	0.195
	∞	–	0	0
0.4	100	0.458	0.266	0.511
	500	0.438	0.220	0.391
	1000	0.447	0.197	0.332
	5000	0.438	0.159	0.257
	∞	–	0	0
0.6	100	0.480	0.294	0.576
	500	0.486	0.239	0.433
	1000	0.491	0.221	0.362
	5000	0.470	0.178	0.301
	∞	–	0	0
0.8	100	0.517	0.303	0.602
	500	0.497	0.242	0.438
	1000	0.490	0.226	0.386
	5000	0.479	0.177	0.303
	∞	–	0	0

(a) $n^{1/3}$ times SD

	$u=0.6$			
t	n	MLE	SMLE	Plug-in
0.2	100	−0.032	−0.009	−0.007
	500	0.001	0.037	0.005
	1000	−0.027	−0.006	−0.008
	5000	0.006	0.024	−0.009
	∞	–	0	0
0.4	100	−0.016	−0.009	−0.028
	500	0.018	0.007	0.004
	1000	−0.022	−0.001	0.001
	5000	−0.008	−0.003	0.004
	∞	–	0	0
0.6	100	0.106	0.023	−0.025
	500	0.011	0.022	0.014
	1000	−0.004	0.005	−0.008
	5000	0.027	0.010	0.004
	∞	–	0	0
0.8	100	0.092	0.033	−0.011
	500	−0.015	0.037	−0.001
	1000	0.012	0.012	0.008
	5000	0.014	−0.011	0.006
	∞	–	0	0

(b) $n^{1/3}$ times bias

the MLE has a bigger variance than the other two estimators, but on the other hand the bias of the MLE is usually smaller than that of the other estimators.

If the second order partial derivatives of the distribution function F_0 vanish at (t,u), as happens, for example, with the uniform distribution, it is possible to achieve higher rates of convergence with the SMLE and the plug-in estimator. We demonstrate this for the uniform distribution function F_0, where we keep the bandwidth constant and equal to 0.4 for the SMLE and equal to 0.2 for the plug-in estimator (to avoid the boundary correction). In this case the bias times $n^{1/3}$ tends to zero for the SMLE and the plug-in estimator, see Table 2b. If we keep the bandwidth constant, the variance of the SMLE and plug-in estimator should be of order n^{-1} and these estimators should therefore actually attain a parametric rate of convergence in this case. We expect the MLE to have again rate $n^{1/3}$ in this case however, which is also (to a certain extent) suggested by Table 2.

5. More general bivariate interval censoring

The purpose of this section is to show that the MLE and SMLE, discussed in the preceding sections in the context of the bivariate current status model, can also be used for more general interval censored data. As mentioned earlier, the current status model is the simplest case of the interval censoring model. For the bivariate interval censoring, case 2, model, the data are of the form

$$\left(T_{i1}, U_{i1}, T_{i2}, U_{i2}, \Delta_{i1}^{(1)}, \Delta_{i2}^{(1)}, \Delta_{i1}^{(2)}, \Delta_{i2}^{(2)}\right),\ i = 1, \dots, n,$$

where

$$\Delta_{i1}^{(1)} = 1_{\{X_{i1} \le T_{i1}\}}, \quad \Delta_{i1}^{(2)} = 1_{\{T_{i1} < X_{i1} \le U_{i1}\}}, \quad \Delta_{i2}^{(1)} = 1_{\{X_{i2} \le T_{i2}\}},$$
$$\Delta_{i2}^{(2)} = 1_{\{T_{i2} < X_{i2} \le U_{i2}\}},$$

Defining

$$\Delta_{i1}^{(3)} = 1 - \Delta_{i1}^{(1)} - \Delta_{i1}^{(2)}, \qquad \Delta_{i2}^{(3)} = 1 - \Delta_{i2}^{(1)} - \Delta_{i2}^{(2)},$$

and defining the corresponding generic values $\delta_i^{(j)}$ similarly, we can define the measure $dV_n^{(ij)}$ by

$$dV_n^{(ij)} = \delta_1^{(i)}\delta_1^{(j)}\delta_2^{(i)}\delta_2^{(j)}\, d\mathbb{P}_n\left(t, u, v, w, \delta_1^{(i)}, \delta_2^{(i)}, \delta_1^{(j)}, \delta_2^{(j)}\right),\ 1 \le i, j \le 3,$$

and an MLE of F is then obtained by maximizing

$$\begin{aligned}
&\ell(F) \\
&= \int \log F(t,v)\, dV_n^{(11)} \\
&+ \int_{t \le u,\, v \le w} \log\{F(u,w) - F(t,w) - F(u,v) + F(t,v)\}\ dV_n^{(22)} \\
&+ \int_{u \ge t} \log\{F(u,v) - F(t,v)\}\ dV_n^{(21)} + \int_{w \ge v} \log\{F(t,w) - F(t,v)\}\ dV_n^{(12)} \\
&+ \int_{u \ge t} \log\{F_1(u) - F_1(t) - F(u,w) + F(t,w)\}\ dV_n^{(23)} \\
&\quad + \int_{w \ge v} \log\{F_2(w) - F_2(v) - F(u,w) + F(u,v)\}\ dV_n^{(32)} \\
&\quad + \int_{w \ge v} \log\{F_1(t) - F(t,w)\}\ dV_n^{(13)} + \int_{u \ge t} \log\{F_2(v) - F(u,v)\}\ dV_n^{(31)} \\
&\qquad + \int \log\{1 - F_1(u) - F_2(w) + F(u,w)\}\ dV_n^{(33)}
\end{aligned} \tag{5.1}$$

over bivariate distribution functions F. The Fenchel duality conditions become:

$$\begin{aligned}
&\int_{t \ge x,\, v \ge y} \frac{dV_n^{(11)}}{F(t,v)} + \int_{t < x \le u,\, v < y \le w} \frac{dV_n^{(22)}}{F(u,w) - F(t,w) - F(u,v) + F(t,v)} \\
&\quad + \int_{t < x \le u,\, v \ge y} \frac{dV_n^{(21)}}{F(u,v) - F(t,v)} + \int_{t \ge x,\, v < y \le w} \frac{dV_n^{(12)}}{F(t,w) - F(t,v)} \\
&\quad + \int_{t < x \le u,\, w \ge y} \frac{dV_n^{(23)}}{F_1(u) - F_1(t) - F(u,w) + F(t,w)}
\end{aligned}$$

$$+\int_{u\ge x,\,v<y\le w}\frac{dV_n^{(32)}}{F_2(w)-F_2(v)-F(u,w)+F(u,v)}$$
$$+\int_{t<x\le u,\,v\ge y}\frac{dV_n^{(13)}}{F_1(t)-F(t,w)}+\int_{t\ge x,\,v<y\le w}\frac{dV_n^{(31)}}{F_2(v)-F(u,v)}$$
$$+\int_{u<x,\,w<y}\frac{dV_n^{(33)}}{1-F_1(u)-F_2(w)+F(u,w)}\le 1, \tag{5.2}$$

with equality if (x,y) is a point of mass of dF.

For computational purposes (but probably not for the development of distribution theory!) it is more convenient not to distinguish between the measures $V_n^{(ij)}$ and just to introduce rectangles to which the unobservable observations are known to belong, where we represent the (one-sided) unbounded rectangles by finite rectangles with upper or lower bounds outside the range of the observed data. In this set-up we simply have to maximize

$$\sum f_i \log H_i' p, \tag{5.3}$$

where $p=(p_1,\dots,p_m)'$ is a vector of probability masses at possible points of mass (x_j,y_j) and H_i is a vector of length m, consisting of ones and zeros, where the component H_{ij} is equal to 1 if the point (x_j,y_j) is contained in the rectangle

$$[L_{i1},R_{i1}]\times[L_{i2},R_{i2}],$$

and is zero, otherwise, and where the f_i denote the multiplicities at the ith observation point. The masses p_j should be nonnegative and sum to 1. This optimization can easily be handled by using iterative quadratic minimization and the support reduction algorithm, documented in [12]. In fact, the treatment is completely analogous to the treatment of the Aspect experiment for quantum statistics, discussed there.

The data of [1] are given in Table 3, where the rectangles to which the hidden observations are known to belong are denoted by $[L_{i1},R_{i1}]\times[L_{i2},R_{i2}]$, $i=1,\dots,n$. The frequencies of the hidden observations belonging to these rectangles are given in the 5th and 10th column. There are 87 observation rectangles and the total sample size, taking the frequencies into account, is 204. The table is also given in [18], Table 7.1, p. 165, but there the rectangles are slightly changed from the data in [1] by lowering the left bounds by 1 if they are larger than zero. Since we do not see a pressing reason for doing that, we just give the data here as they were given by [1]. If the upper bound R_{ij} is unknown, we put $R_{ij}=\infty$ and if the lower bound L_{ij} is unknown, we put $L_{ij}=-\infty$.

The maximal intersection rectangles where the MLE will put its mass are given in Table 4a. They can be computed, for example, by applying the reduction algorithm, used in the R package MLEcens.

To facilitate the comparison with the existing literature, we will only discuss the MLE, based on the preliminary reduction to rectangles which can have mass, and not follow the procedure we used for computing the MLE on a sieve in the

TABLE 3
The Betensky-Finkelstein data

L_{i1}	R_{i1}	L_{i2}	R_{i2}	frequency		L_{i1}	R_{i1}	L_{i2}	R_{i2}	frequency
0	3	0	–	3		6	–	6	–	3
0	3	3	–	1		6	–	9	–	1
0	3	6	–	3		9	–	0	–	2
0	6	6	–	1		9	–	9	–	3
0	3	9	–	1		9	–	12	–	1
0	3	12	–	5		12	–	0	–	5
0	3	15	–	5		12	–	6	–	1
0	6	15	–	1		12	–	9	–	4
3	3	3	–	1		12	–	12	–	10
3	3	6	–	1		15	–	0	–	3
3	3	9	–	3		15	–	3	–	1
3	6	9	–	2		15	–	6	–	1
3	6	12	–	3		15	–	9	–	2
3	3	15	–	2		15	–	12	–	8
3	6	15	–	2		15	–	15	–	9
3	6	18	–	1		18	–	0	–	1
3	3	21	–	1		18	–	6	–	1
6	6	0	–	2		18	–	9	–	1
6	9	0	–	1		18	–	12	–	1
6	9	9	–	1		18	–	15	–	3
6	6	12	–	1		18	–	18	–	6
6	9	12	–	2		21	–	15	–	1
6	6	15	–	1		–	0	0	–	9
6	9	15	–	1		–	0	3	–	3
6	6	18	–	1		–	0	6	–	10
6	9	18	–	2		–	0	9	–	6
9	9	0	–	1		–	0	12	–	8
9	12	0	–	2		–	0	15	–	5
9	9	9	–	2		–	0	18	–	4
9	12	9	–	1		–	0	21	–	1
9	12	12	–	3		0	–	0	3	1
9	9	15	–	1		6	–	0	6	1
9	12	24	–	1		6	–	6	6	1
9	9	27	–	1		12	–	0	3	1
12	12	0	–	1		12	–	0	6	1
12	15	0	–	1		15	–	0	3	1
12	15	6	–	1		21	–	15	15	1
12	15	15	–	1		3	–	–	0	1
12	15	21	–	1		9	–	–	0	1
0	–	0	–	6		12	–	–	0	1
3	–	0	–	2		0	3	0	6	1
6	–	0	–	1		3	6	6	12	1
6	–	3	–	2		9	9	9	9	1
–	0	–	0	1						

simulation from the density $f(x,y) = x+y$ on $[0,1]^2$. We will use the convention of putting the mass of the MLE in the right upper corner of these rectangles, and compute the MLE by the support reduction algorithm of [12]. The result is shown in Table 4, where the masses of the MLE are given. It is seen that this is in close correspondence with Table 7.2 on p. 166 of [18], apart from the slightly different definition of the rectangles. The R package MLEcens also gives this result (in all 9 decimals).

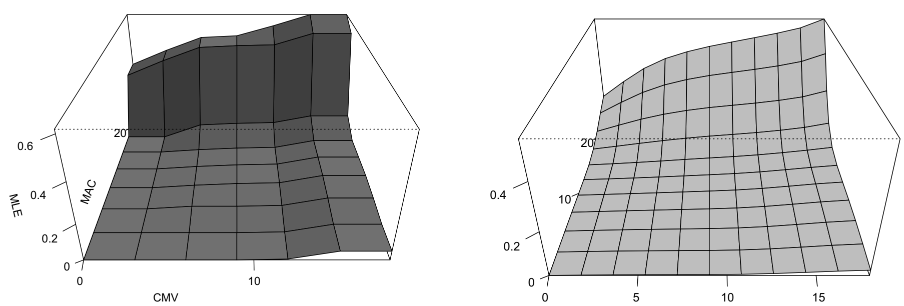
TABLE 4
Maximal intersection rectangles and masses of MLE

L_{j1}	R_{j1}	L_{j2}	R_{j2}
0	0	0	0
0	0	21	–
3	3	21	–
6	6	6	6
6	6	18	–
9	9	9	9
9	9	27	–
12	12	0	0
12	12	24	–
15	15	0	0
15	15	21	–
21	–	15	15
21	–	18	–

(a) Canonical rectangles

L_{j1}	L_{j2}	mass MLE
0	0	0.013676984
0	21	0.307533525
3	21	0.087051863
6	6	0.014940282
6	18	0.062521573
9	9	0.010009349
9	27	0.071073995
12	0	0.004836043
12	24	0.053334241
15	0	0.042456241
15	21	0.021573343
21	15	0.044427509
21	18	0.266565054

(b) Masses of MLE



(a) MLE (b) SMLE

FIG 5. *MLE and SMLE for the Betensky-Finkelstein data, restricted to the interval* $[0, 18] \times [0, 24]$.

The SMLE for bivariate interval censoring is again defined by (3.1). A picture of the MLE and the SMLE is shown in Figure 5 and the picture of the level curves in Figure 7. For the meaning of the codings CMV (cytomegalovirus) and MAC (mycobacterium avium complex), see [1] or [18], section 7.3. Both the MSLE and the MSLE indicate that CMV shedding occurs prior to MAC colonization.

It can be seen from this picture that the steep increase of the first marginal df of the MLE and SMLE, shown in Figure 6, is due to the 'ridge' for the larger values of the second coordinate. The levels of both estimates are shown in Figure 7. It seems to me that the SMLE might be the more sensible estimate, also in view of the representational non-uniqueness of the MLE, which is somewhat 'washed out' by the SMLE.

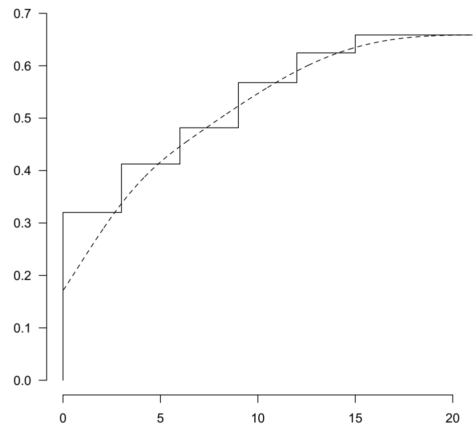
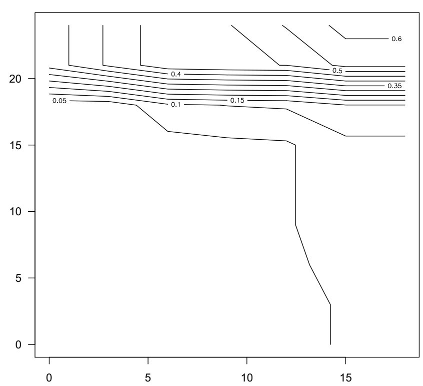
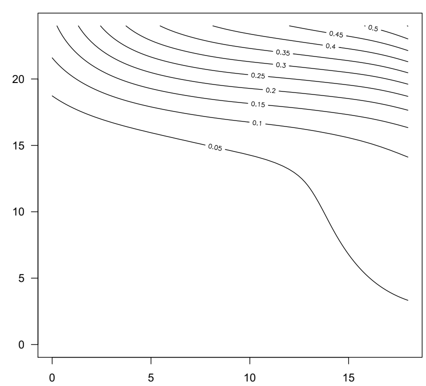


FIG 6. *The first marginal df of the data set in [1], computed on the interval* $[0, 21)$ *(the largest observation point on the first coordinate is* 21*). The solid curve gives the first marginal df of the MLE and the dashed curve the first marginal of the SMLE, taking bandwidth* $n^{-1/6}$.



(a) Level plot of the MLE



(b) Level plot of the SMLE

FIG 7. *Contourplot of the MLE and SMLE for the Betensky-Finkelstein data, restricted to the interval* $[0, 18] \times [0, 24]$.

6. Concluding remarks

In the preceding, three estimators for the bivariate current status model were studied: the maximum likelihood estimator (MLE), which in the simulation study was restricted to the MLE on a sieve, the smoothed maximum likelihood estimator (SMLE), obtained by integrating a kernel w.r.t. the masses of the MLE, and a purely discrete plug-in estimator. The SMLE and plug-in seem to attain the $n^{1/3}$ rate, with asymptotically normal distributions.

It might be somewhat surprising that the SMLE have the same rate, whereas the natural rates in the one-dimensional case are $n^{1/3}$ and $n^{2/5}$, respectively, see [9]. But in the bivariate case the variance is of order $1/(nh^2)$ and the bias of order h^2, if a bandwidth h is taken in both directions and the kernel is of the usual symmetric and positive type. This makes the optimal choice of bandwidth of order $n^{-1/6}$, leading to a rate of order $n^{1/3}$.

We concentrated on compact support, but since we focused on local estimation, this restriction does not seem essential, except for the SMLE, since there the boundary played an important role. Further research on this matter seems needed.

It is also possible to attain again the local $n^{2/5}$ rate in the bivariate case, but then one has to take recourse to higher order kernels K with the property

$$\int u^2 K(u)\, du = 0.$$

One also has to take bandwidths of order $n^{-1/10}$ instead of order $n^{-1/6}$ in this case, which makes the judicious choice of boundary kernels even more important. Moreover, one has to strengthen the conditions of the theorems to the existence of 4th derivatives. However, if one is willing to do that, it is easy to strengthen Theorem 3.1 by letting the kernel $I\!\!K$ be based on, for example, the kernel

$$K_1(u) = \tfrac{315}{512}\left(1-u^2\right)^3 (11u^2-3)1_{[-1,1]}(u)$$

which is the fourth order kernel corresponding to the Triweight kernel

$$K(u) = \tfrac{35}{32}\left(1-u^2\right)^3 1_{[-1,1]}(u).$$

But one loses the property that the resulting estimate is necessarily a distribution function, since the kernel K_1 is no longer positive. For the choice of higher order kernels, see, e.g., [14].

7. Appendix

Proof of Theorem 2.1. We use the representation

$$\begin{aligned}
&\frac{\int_{A_n}\delta_1\delta_2\, d\mathbb{P}_n(v,w,\delta_1,\delta_2)}{\int_{A_n} d\mathbb{G}_n(v,w)} - F_0(t,u)\\
&= \frac{\int_{A_n}\delta_1\delta_2\, d\mathbb{P}_n - E\left\{\int_{A_n}\delta_1\delta_2\, d\mathbb{P}_n \,\middle|\, (T_i,U_i),\, i=1,\dots,n\right\}}{\int_{A_n} d\mathbb{G}_n(v,w)}\\
&\quad + \frac{E\left\{\int_{A_n}\delta_1\delta_2\, d\mathbb{P}_n \,\middle|\, (T_i,U_i),\, i=1,\dots,n\right\}}{\int_{A_n} d\mathbb{G}_n(v,w)} - \frac{n^{-1}\sum_{i=1}^n F_0(t,u)1_{A_n}(T_i,U_i)}{\int_{A_n} d\mathbb{G}_n(v,w)}.
\end{aligned} \tag{7.1}$$

The numerator of the first term on the right-hand side can be written:

$$n^{-1}\sum_{i=1}^n \{\Delta_{i1}\Delta_{i2}-F_0(T_i,U_i)\}\,1_{A_n}(T_i,U_i).$$

This is the sum of i.i.d. random variables, and

$$\text{var}\left(\{\Delta_{11}\Delta_{12}-F_0(T_1,U_1)\}\,1_{A_n}(T_1,U_1)\right)\sim 4n^{-1/3}g(t,u)F_0(t,u)\,\{1-F_0(t,u)\}\,,$$

as $n\to\infty$. Hence:

$$n^{1/3}\frac{\int_{A_n}\delta_1\delta_2\,d\mathbb{P}_n - E\left\{\int_{A_n}\delta_1\delta_2\,d\mathbb{P}_n \;\middle|\; (T_i,U_i),\,i=1,\dots,n\right\}}{\int_{A_n} d\mathbb{G}_n(v,w)}\xrightarrow{\mathcal{D}} N(0,\sigma^2),$$

where

$$\sigma^2=\frac{F_0(t,u)\,\{1-F_0(t,u)\}}{4g(t,u)}\,.$$

The numerator of the second term on the right-hand side of (7.1) can be written:

$$n^{-1}\sum_{i=1}^n \{F_0(T_i,U_i)-F_0(t,u)\}\,1_{A_n}(T_i,U_i).$$

Note that, putting $h=h_n\sim n^{-1/6}$,

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n E\,\{F_0(T_i,U_i)-F_0(t,u)\}\,1_{A_n}(T_i,U_i)\\
&=\int_{\max\{|v-t|,|w-u|\}\le h}\{F_0(v,w)-F_0(t,u)\}\;g(v,w)\,dv\,dw\\
&=\partial_1F_0(t,u)\int_{\max\{|v-t|,|w-u|\}\le h}(v-t)\,g(v,w)\,dv\,dw\\
&\qquad+\partial_2F_0(t,u)\int_{\max\{|v-t|,|w-u|\}\le h}(w-u)\,g(v,w)\,dv\,dw\\
&\qquad+\tfrac12\partial_1^2F_0(t,u)\int_{\max\{|v-t|,|w-u|\}\le h}(v-t)^2\,g(v,w)\,dv\,dw\\
&\qquad+\tfrac12\partial_2^2F_0(t,u)\int_{\max\{|v-t|,|w-u|\}\le h}(w-u)^2\,g(v,w)\,dv\,dw+o\left(n^{-2/3}\right)\\
&=\partial_1F_0(t,u)\partial_1g(t,u)\int_{\max\{|v-t|,|w-u|\}\le h}(v-t)^2\,dv\,dw\\
&\qquad+\partial_2F_0(t,u)\partial_2g(t,u)\int_{\max\{|v-t|,|w-u|\}\le h}(w-u)^2\,dv\,dw\\
&\qquad+\tfrac12\partial_1^2F_0(t,u)\int_{\max\{|v-t|,|w-u|\}\le h}(v-t)^2\,g(v,u)\,dv\,dw
\end{aligned}$$

$$+\tfrac{1}{2}\partial_2^2 F_0(t,u)\int_{\max\{|v-t|,|w-u|\}\le h}(w-u)^2\,g(v,u)\,dv\,dw+o\left(n^{-2/3}\right)$$
$$=\tfrac{4}{3}\left\{\partial_1F_0(t,u)\partial_1 g(t,u)+\partial_2F_0(t,u)\partial_2 g(t,u)\right\}h^4$$
$$+\tfrac{2}{3}\left\{\partial_1^2F_0(t,u)+\partial_2^2F_0(t,u)\right\}g(t,u)h^4+o\left(n^{-2/3}\right).$$

Moreover,

$$\operatorname{var}\left(n^{-1}\sum_{i=1}^n\left\{F_0(T_i,U_i)-F_0(t,u)\right\}1_{A_n}(T_i,U_i)\right)$$
$$=O\left(n^{-1}\int_{\max\{|t-t_0|,|u-u_0|\}\le h}\left\{F_0(t,u)-F_0(t_0,u_0)\right\}^2\,g(t,u)\,dt\,du\right)$$
$$=O\left(n^{-5/6}\right).$$

The result now follows. □

References

[1] R. A. Betensky and D. M. Finkelstein, *A non-parametric maximum likelihood estimator for bivariate interval censored data*, Statist. Med. **18** (1999), no. 22, 3089–3100.

[2] Kris Bogaerts and Emmanuel Lesaffre, *A new, fast algorithm to find the regions of possible support for bivariate interval-censored data*, J. Comput. Graph. Statist. **13** (2004), no. 2, 330–340. MR2063988

[3] Robert Gentleman and Alain C. Vandal, *Nonparametric estimation of the bivariate CDF for arbitrarily censored data*, Canad. J. Statist. **30** (2002), no. 4, 557–571. MR1964427 (2004b:62090)

[4] R. B. Geskus and P. Groeneboom, *Asymptotically optimal estimation of smooth functionals for interval censoring. I*, Statist. Neerlandica **50** (1996), no. 1, 69–88. MR1381209 (97k:62039)

[5] R. B. Geskus and P. Groeneboom, *Asymptotically optimal estimation of smooth functionals for interval censoring. II*, Statist. Neerlandica **51** (1997), no. 2, 201–219. MR1466426 (99d:62015)

[6] Ronald Geskus and Piet Groeneboom, *Asymptotically optimal estimation of smooth functionals for interval censoring, case* 2, Ann. Statist. **27** (1999), no. 2, 627–674. MR1714713 (2000j:60044)

[7] P. Groeneboom, *Lectures on inverse problems*, Lectures on probability theory and statistics (Saint-Flour, 1994), Lecture Notes in Math., vol. 1648, Springer, Berlin, 1996, pp. 67–164. MR1600884 (99c:62092)

[8] P. Groeneboom, *Nonparametric (smoothed) likelihood and integral equations*, Discussion paper. To appear in the Journal of Statistical Planning and Inference, 2012.

[9] P. Groeneboom, G. Jongbloed, and B. I. Witte, *Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model*, Ann. Statist. **38** (2010), 352–387.

[10] P. Groeneboom and J.A. Wellner, *Information bounds and nonparametric maximum likelihood estimation*, DMV Seminar, vol. 19, Birkhäuser Verlag, Basel, 1992. MR1180321 (94k:62056)

[11] Piet Groeneboom, *Likelihood ratio type two-sample tests for current status data*, Scand. J. Stat. **39** (2012), no. 4, 645–662.

[12] Piet Groeneboom, Geurt Jongbloed, and Jon A. Wellner, *The support reduction algorithm for computing non-parametric function estimates in mixture models*, Scand. J. Statist. **35** (2008), no. 3, 385–399. MR2446726 (2009m:62115)

[13] Piet Groeneboom and Tom Ketelaars, *Estimators for the interval censoring problem*, Electron. J. Stat. **5** (2011), 1797–1845. MR2870151

[14] M. C. Jones and D. F. Signorini, *A comparison of higher-order bias kernel density estimators*, J. Amer. Statist. Assoc. **92** (1997), no. 439, 1063–1073. MR1482137

[15] Marloes H. Maathuis, *Reduction algorithm for the NPMLE for the distribution function of bivariate interval-censored data*, J. Comput. Graph. Statist. **14** (2005), no. 2, 352–362. MR2160818

[16] Eugene F. Schuster, *Incorporating support constraints into nonparametric estimators of densities*, Comm. Statist. A—Theory Methods **14** (1985), no. 5, 1123–1136. MR797636 (86m:62078)

[17] S. Song, *Estimation with bivariate interval censored data*, Ph.D. dissertation, University of Washington, Seattle, USA, 2001.

[18] Jianguo Sun, *The statistical analysis of interval-censored failure time data*, Statistics for Biology and Health, Springer, New York, 2006. MR2287318 (2007h:62007)