Bivariate Survival Curve Estimation using Nonparametric Smoothing Techniques

By

Ronald C. Pruitt

Technical Report No. 543

School of Statistics

University of Minnesota

7 March 1990

[1]Research partially supported by National Institutes of General Medical Sciences Grant GM-42921-01.

Abstract

The generalized maximum likelihood estimator (GMLE) has been studied for bivariate censored data by several authors. The estimator has been shown to be consistent for data from discrete distributions, but can be inconsistent for data from absolutely continuous distributions. The GMLE is not unique for this problem, and it is contained in a larger class of self-consistent estimators. Here we provide smoothed versions of a self-consistent estimator which are consistent for absolutely continuous data. The self-consistency property is not retained under bivariate right censoring but is retained under univariate censoring. The estimator is a proper survival function, affine equivariant, redistributive, and intuitively appealing. A version of the estimator is easily understood based only on the concepts of univariate product limit estimation and self-consistency.

1 Introduction.

We consider the problem of estimating a bivariate survival curve with bivariate right censored data. Many estimators have been proposed for this problem, including those of Muñoz (1980), Campbell (1981), Langberg and Shaked (1982), Campbell and Földes (1982), Hanley and Parnes (1983), Tsai, Leurgans, and Crowley (1986), and Dabrowska (1988), but none of them are entirely satisfactory.

All of the estimators proposed above are in one way or another generalizations of the empirical distribution function to deal with censoring. A concept we consider important in this regard is the distribution of mass of these estimators when applied to sample data. We call an estimator *redistributive* if it associates mass $1/n$ with each censored observation. The technical definition is postponed until Section 2, but we can illustrate the definition simply for univariate data. The univariate product limit estimator is redistributive, since mass $1/n$ is associated with each observation when the estimator is considered via Efron's (1967) redistribute to the right algorithm for computing it. It seems reasonable to hope that estimators based on the empirical distribution function would be redistributive. Only the GMLE and the estimator proposed here are redistributive among this group.

The estimators of Langberg and Shaked (1982), Campbell and Földes (1982), and Dabrowska (1988) usually fail to be survival functions for data arising from absolutely continuous distributions. They can be applied to discretized data to remove this problem for large enough sample sizes, but none of the estimators has been studied under these conditions, and no arguments to use anything other than the GMLE, which is unique for the discrete problem (for large enough sample sizes), have been advanced.

Muñoz (1980), Campbell (1981), and Hanley and Parnes (1983) all discuss the GMLE for this problem. This estimator is known to be consistent if the distribution being estimated is purely discrete (Campbell, 1981), but may be inconsistent for continuous data (see Leurgans, Tsai, and Crowley, 1982). The estimator is also not unique for samples taken from absolutely continuous distributions. In this paper, we establish the uniform consistency of the GMLE under a restricted censoring scheme in which either both or neither of the variables are censored and the data come from absolutely continuous distributions.

Tsai, Leurgans, and Crowley (1986) proposed an estimator, here called the TLC estimator, using nonparametric smoothing techniques relying on a decomposition of a bivariate survival function, and showed it to be uniformly consistent. Although nonparametric smoothing techniques are used, it seems likely that the estimates of the survival function converge at the usual $n^{-1/2}$ rate, just as integrated density estimates converge faster asymptotically than the density estimates themselves. The estimator developed here shares all these features with the TLC estimator. The negative features of the TLC estimator are that it is not redistributive, is not affine equivariant, and only assigns mass to uncensored observations and singly censored observations with the uncensored value smaller than the censored value. Further comments on these points are postponed until Section 5. The decomposition on which the estimator is based is also not very intuitive: this is not necessarily a negative feature of the estimator, but it is an area in which we feel the estimator proposed here is more attractive.

The goal is to construct an estimator which avoids the negative features of those given above. To do this we make use of the idea of self-consistency formulated by Efron (1967). Uniform convergence and weak convergence of self-consistent estimators was considered in a general context by Tsai and Crowley (1985). It has been pointed out (see correction notice) that their results apply only under very restrictive conditions which are not satisfied even for the standard univariate examples. In Section 2, we show that uniform convergence can be obtained under less restrictive conditions. This result is shown to apply to the univariate examples in Section 3. The bivariate case can't be covered since the bivariate GMLE, which is self-consistent, can be inconsistent. But we use the general result to show the GMLE is uniformly consistent in the special case when all censoring occurs in both variables or neither. We then attack the general case using nonparametric smoothing techniques to handle observations with only one variable censored and then applying the previous result to handle observations with both variables censored. This is done in Section 4. We make some comments in a final section.

2 Background.

Let the unobservable survival times of interest be given by (X_1, X_2) and a nuisance censoring variable be given by (C_1, C_2). We consider bivariate right censoring, where the observable variables are Y_1, Y_2, D_1, and D_2, where $Y_j = X_j \wedge C_j$ and $D_j = 1[Y_j = X_j]$. It has usually been assumed that (X_1, X_2) and (C_1, C_2) are independent to ensure the identifiability of the survival function. This assumption is stronger than necessary for identifiability (see Pruitt, 1990), but we will assume it for the results of this paper. In Section 5 we indicate some generalizations. For univariate data the subscripts will be omitted.

We consider the problem in a more general framework. Let $(\vec{X}_1, \vec{C}_1), \ldots, (\vec{X}_n, \vec{C}_n)$ be iid random vectors distributed as $(\vec{X}, \vec{C})$, where $\vec{X}$ and $\vec{C}$ are independent, and $F_x(F_c)$ is the distribution function of $\vec{X}(\vec{C})$. The X portion of the data is of interest and the C portion is not. Let $\mathcal{X} \times \mathcal{C}$ and $\mathcal{Y}$ be two sample spaces and let $\mathcal{M}$ be a many-to-one mapping from $\mathcal{X} \times \mathcal{C}$ to $\mathcal{Y}$. The observed data are $\vec{Y}_1 = \mathcal{M}(\vec{X}_1, \vec{C}_1), \ldots, \vec{Y}_n$. The mapping $\mathcal{M}$ can be extended to a mapping $\tilde{\mathcal{M}}$ from the space of product probability measures on $\mathcal{X} \times \mathcal{C}$ to the space of probability measures on $\mathcal{Y}$. For degenerate distributions $\vec{X}(\vec{C}, \vec{Y})$ with distribution function $F_x(F_c, F_y)$ and $\mathcal{M}(\vec{X}, \vec{C}) = \vec{Y}$, define $\tilde{\mathcal{M}}(F_x, F_c) = F_y$. We then extend the definition of $\tilde{\mathcal{M}}$ in the usual way through simple distributions to general distributions.

Associated with each observation $\vec{Y}$ is a set E where $\vec{Y}$ carries the same information as $\vec{X} \in E$. We will call E the set associated with $\vec{Y}$. An estimator of F_x based on $\vec{Y}_1, \ldots, \vec{Y}_n$ is called *redistributive* if mass $1/n$ can be associated with each observation $\vec{Y}_i$ and the mass associated with $\vec{Y}_i$ is assigned entirely within the set E_i. An estimator based on the empirical distribution function of the incomplete data which is not redistributive cannot have the empirical distribution of the complete data in its pre-image under $\tilde{\mathcal{M}}$.

We next consider the notion of self-consistency. Write

$$H(G_y, G_x)(\vec{t}) = -G_x(\vec{t}) + \int \mathrm{E}_{G_x}[1(\vec{X} \leq \vec{t}|\vec{Y} = \vec{y}]dG_y(\vec{y}), \tag{2.1}$$

and note H is a function defined on $\mathcal{X}$. Let $A \subseteq \mathcal{X}$ be specified and let $\| F \|_\infty = \sup_{t \in A} |F(t)|$ for any function F defined on $\mathcal{X}$. The set A of interest will depend on the specific example, and unless otherwise specified theorems hold for any A and all suprema are taken over the set A. An estimator $\hat{F}_x$ based on F_y is called self-consistent if $\| H(F_y, \hat{F}_x) \|_\infty = 0$. For a definition of self-consistency when $\vec{X}$ and $\vec{C}$ are not assumed independent see Pruitt (1990). Note that $\| H(F_y, F_x) \|_\infty = 0$ and F_y is in the image of $\tilde{\mathcal{M}}$ if and only if F_x is the first component of a pre-image of F_y under $\tilde{\mathcal{M}}$. Hence questions about the uniqueness of solutions to the self-consistency equation reduce to questions about the identifiability of F_x

from F_y. This does not contradict the nonuniqueness of self-consistent estimators for the bivariate problem, where the nonuniqueness arises because sample distribution functions are typically not in the image of $\tilde{\mathcal{M}}$ (see Pruitt, 1990).

We will provide a proof of the consistency of self-consistent estimators in this framework. This is closely related to the work of Tsai and Crowley (1985). In their revised proof of the almost sure consistency of self-consistent estimators they assume H is continuously differentiable, a condition H does not satisfy in standard examples. Here we show that to prove consistency only continuity assumptions about H are needed. We define H as an extended real-valued function. Let $c/0 = \infty$ for any c, and let $\infty - \infty = 0$.

Theorem 2.1 *Let $\mathcal{G} = \{G_x : \| H(F_y, G_x) \|_\infty < \infty\}$. Make the following assumptions:*

A0) *Define $H(G_y, G_x)$ for all G_y and G_x so that $\| H(F_y, F_x) \|_\infty = 0$ if and only if F_x is the first component of a pre-image of F_y under $\tilde{\mathcal{M}}$. This will usually be by an extension of (2.1).*

A1) *F_y^n is a sequence of distribution functions with $\| F_y^n - F_y \|_\infty \to 0$ a.s.*

A2) *For this sequence $\sup_{G_x} \| H(F_y^n, G_x) - H(F_y, G_x) \|_\infty \to 0$ almost surely.*

A3) *The solution of $H(F_y, G_x) = 0, F_x$, is unique.*

A4) *For any sequence of distributions G_n with $\| G_n - G_x \|_\infty \to 0$ almost surely and $G_x \in \mathcal{G}$, we have $\| H(F_y, G_n) - H(F_y, G_x) \|_\infty \to 0$ almost surely.*

A5) *There exist δ_0 and ϵ_0 such that*

$$\delta_0 = \inf_{G_x \notin Q_{\epsilon_0}} \| H(F_y, G_x) \|_\infty > 0,$$

where $Q_\epsilon = \{G_x : \| G_x - F_x \|_\infty < \epsilon\}$.

Then $\| \hat{F}_x^n - F_x \|_\infty \to 0$ almost surely.

Assumption 5 will usually be quite easy to prove. Assumptions 0,1, and 3 essentially define the problem, and assumptions 2 and 4 are the critical ones. That these assumptions hold for the univariate right censoring and interval censoring problems is established in Section 3. It is not particularly important to make assumption 0 explicit in the univariate case, but this becomes one of the problems in the bivariate case. Assumptions 2 and 4 both fail in the bivariate right censoring case, and Section 4 shows one way around this problem.

Before proceeding to the proof we give an auxiliary lemma.

Lemma 2.2 *Conditions A3), A4), and A5) imply, for every* $\epsilon > 0$,

$$\delta = \inf_{G_x \notin Q_\epsilon} \| H(F_y, G_x) \|_\infty > 0.$$

Proof: Obtain δ_0 and ϵ_0 from A5). If $\epsilon \geq \epsilon_0$ then $\delta \geq \delta_0 > 0$. If $\epsilon < \epsilon_0$, then

$$\inf_{G_x \in \bar{Q}_{\epsilon_0} \setminus Q_\epsilon} \| H(F_y, G_x) \|_\infty = \delta^* > 0$$

where $\bar{Q}_{\epsilon_0} = \{G_x : \| G_x - F_x \|_\infty \leq \epsilon_0\}$. Since $\bar{Q}_{\epsilon_0} \setminus Q_\epsilon$ is compact, the infimum is attained by A4) and is non-zero by A3). □

We now prove Theorem 2.1.

Proof: [of Theorem 2.1] Fix $\epsilon > 0$. Let

$$M_\epsilon = \{G_x : \| H(F_y, G_x) \|_\infty < \epsilon\},$$

$$\text{and } Q_\epsilon = \{G_x : \| F_x - G_x \|_\infty < \epsilon\}.$$

By Lemma 2.2, there exists $\gamma > 0$ with $M_\gamma \subset Q_\epsilon$ such that $F_x^* \notin M_\gamma$ implies $\| H(F_y, F_x^*) \|_\infty > \gamma$. Let Ω be the set where the convergence in A2) holds, and choose N such that $n > N$ implies

$$\sup_{G_x} \| H(F_y, G_x) - H(F_y^n, G_x) \|_\infty < \gamma/2$$

for all sequences in Ω. Then $n > N$ implies $\| H(F_y^n, F_x^*) \|_\infty > \gamma/2$ for all $F_x^* \notin Q_\epsilon$. For $n > N$, $\hat{F}_x^n \in Q_\epsilon$ for any solution $\hat{F}_x^n$ of $H(F_y^n, G_x) = 0$. Hence $\| \hat{F}_x^n - F_x \|_\infty \to 0$ almost surely. □

3 The one dimensional problem.

3.1 Right censoring.

We wish to show that Theorem 2.1 applies to the usual right censored univariate data case. The results are straightforward, but the bivariate case follows the same arguments so these are presented for clarity. Here we will take F_y^n to be any sequence of distributions with $\| F_y^n - F_y \|_\infty \to 0$. For example we may take the empirical distribution of the incomplete data. The distribution F_y may equivalently be specified by the subsurvival functions: $S_u(t) = P(Y > t, D = 1)$ and $S_c(t) = P(Y > t, D = 0)$. For A0) we need to worry about the definition of H when a censored data point, Y, is observed and $G_x(Y) = 1$. Note that (2.1) is equivalent to

$$H(F_y, G_x)(t) = \bar{G}_x(t) - \int \mathrm{E}_{G_x}[1(X > t)|Y = \vec{y}]dF_y(\vec{y}),$$

where $\bar{G}_x(t) = 1 - G_x(t)$. This can be extended to all G_x by defining H by

$$H(F_y, G_x)(t) = \bar{G}_x(t) - S_u(t) + \int \frac{\bar{G}_x(t \vee y)}{\bar{G}_x(y)} DS_c(y^-)$$

which is a Riemann-Stieltjes integral, recalling $c/0$ is defined to be infinity for any c. For notational convenience, define

$$f(y; t, G_x) = \frac{\bar{G}_x(t \vee y)}{\bar{G}_x(y)},$$

Assumptions 1) and 3) are well known: Tsai (1986) has shown that A3) holds for $A = [0, T)$ where $T = \sup\{t | S_u(t) + S_c(t) > 0\}$ and $A = \Re_+$ if $T = \infty$. We state the remaining three assumptions as lemmas.

Lemma 3.1 *For the univariate right censoring problem,*

$$\sup_{G_x} \| H(F_y^n, G_x) - H(F_y, G_x) \|_\infty \to 0$$

almost surely, for any sequence F_y^n with $\| F_y^n - F_y \|_\infty \to 0$ a.s.

Proof: Note that the required convergence holds for $G_x \notin \mathcal{G}$. To see this let $Y(Y_n)$ have distribution function $F_y(F_y^n)$. For $G_x \notin \mathcal{G}$, there exists t_0 with $\bar{G}_x(t_0) = 0$ but $P[Y \geq$

$t_0, D = 0] > 0$, and hence for large enough n, $P[Y_n \geq t_0, D = 0] > 0$ and $\| H(F_y^n, G_x) \|_\infty = \infty$. For $G_x \in \mathcal{G}$, let Ω be the set where the convergence of F_y^n holds. Note that on Ω, both S_u^n and S_c^n, the empirical versions of S_u and S_c converge uniformly in t. Suppose $\epsilon_u(n)$ and $\epsilon_c(n)$ are such that

$$\| S_u - S_u^n \|_\infty < \epsilon_u(n) \qquad \text{and} \qquad \| S_c - S_c^n \|_\infty < \epsilon_c(n),$$

and $\epsilon_u(n)$ and $\epsilon_c(n)$ both converge to zero as $n \to \infty$. Then

$$\begin{aligned}
\sup_{G_x \in \mathcal{G}} \| H(F_y^n, G_x) - H(F_y, G_x) \|_\infty &= \sup_{G_x \in \mathcal{G}} \sup_t |H(F_y^n, G_x)(t) - H(F_y, G_x)(t)| \\
&\leq \epsilon_u(n) + \sup_{G_x \in \mathcal{G}} \sup_t \left| \int f(y; t, G_x)(DS_c^n(y^-) - DS_c(y^-)) \right| \\
&\leq \epsilon_u(n) + \epsilon_c(n) \sup_{G_x \in \mathcal{G}} \sup_t \left| \int Df(y; t, G_x) \right| \\
&\leq \epsilon_u(n) + \epsilon_c(n).
\end{aligned}$$

□

Lemma 3.2 *For any sequence of distributions G_n with $\| G_n - G_x \|_\infty \to 0$ almost surely and $G_x \in \mathcal{G}$, we have $\| H(F_y, G_n) - H(F_y, G_x) \|_\infty \to 0$ almost surely.*

Proof: Let Ω be the set where $\| G_n - G_x \|_\infty \to 0$. On Ω

$$\| H(F_y, G_n) - H(F_y, G_x) \|_\infty \leq \| G_x - G_n \|_\infty + \int \sup_t |f(y; t, G_x) - f(y; t, G_n)| DS_c(y^-)$$

By bounded convergence it suffices to show that the integrand converges to zero pointwise (in y) on the support of S_c, where $\bar{G}_x(y) > 0$ since $G_x \in \mathcal{G}$. Now

$$\begin{aligned}
\sup_t |f(y; t, G_x) - f(y; t, G_n)| &\leq 1(\bar{G}_x(y) > 0, \bar{G}_n(y) = 0) \\
&\quad + \sup_t \left| \frac{\bar{G}_x(t \vee y)}{\bar{G}_x(y)} - \frac{\bar{G}_n(t \vee y)}{\bar{G}_n(y)} \right| 1(\bar{G}_x(y) > 0, \bar{G}_n(y) > 0).
\end{aligned}$$

The first term converges to zero. For the last term, fix $0 < \epsilon < 0.5$. Take N_1 so large that $|\bar{G}_n(y) - \bar{G}_x(y)| < 2^{-1}\epsilon \bar{G}_x(y)$ for $n > N_1$. Take $N_2 > N_1$ so large that $\sup_t |\bar{G}_x(t) - \bar{G}_n(t)| < 2^{-1}\epsilon(1 - \epsilon)$ whenever $n > N_2$. Then by checking the maximum positive and negative values for the absolute value in the final term it can be shown to be smaller than ϵ on Ω for $n > N_2$. Hence the lemma follows. □

Lemma 3.3 *There exist δ and ϵ such that for $F_x^* \notin Q_\epsilon \equiv \{G_x : \| G_x - F_x \|_\infty < \epsilon\}$ we have* $\| H(F_y, F_x^*) \|_\infty > \delta > 0$.

Proof: Let $\gamma = S_u(0) = P[D = 1] > 0$, and let $\delta = \gamma/4$. Let $\epsilon = 1 - 2\delta$. Assume we can find t_0 such that $F_x^*(t_0) > \epsilon - \delta + F_x(t_0)$. Then

$$\begin{aligned}
&\int \mathrm{E}_{F_x^*}[1(X \le t_0)|\vec{Y} = \vec{y}]dF_y(\vec{y}) - F_x(t_0) = \\
&= \int \mathrm{E}_{F_x^*}[1(X \le t_0)|\vec{Y} = \vec{y}]dF_y(\vec{y}) - \int \mathrm{E}_{F_x}[1(X \le t_0)|\vec{Y} = \vec{y}]dF_y(\vec{y}) \\
&\le S_c(0) \\
&= 1 - \gamma
\end{aligned}$$

so that $H(F_y, F_x^*)(t_0) < -\delta$. The case when $F_x^*(t_0) < F_x(t_0) - \epsilon + \delta$ is similar. □

This shows that the univariate product limit estimator is uniformly consistent on A, where $A = [0, T)$ and $T = \sup\{t | S_u(t) + S_c(t) > 0\}$.

3.2 Interval censoring.

It can also be shown that Theorem 2.1 holds for the double censoring case considered by Turnbull (1974,1976), Chang and Yang (1987), and Chang (1990). In fact, if A3) is assumed the results can be applied to more general censoring models (see Turnbull, 1976 and Gill, 1989). For example if we assume that the incomplete observations arise as information that X is in some measurable set E, similar results can be obtained. A completely general scheme would allow any measurable set E to occur, but here we just consider a simple generalization of double censoring which allows us to keep the incomplete observations as real valued vectors.

We consider four types of incomplete observations: uncensored, right censored, left censored, and interval censored with both endpoints finite. That is we observe $X \in E$ where E can be of the forms: $\{e\}, (e, \infty), (-\infty, e]$, or $(e_1, e_2]$. Which endpoints are open and closed are not dictated by the method, we could with only notational inconvenience allow the 9 possible types given by allowing inclusion and exclusion of endpoints in the types above. The observables Y and D take on possible values $D = 0, Y = y$ corresponding to $X \in (y, \infty)$; $D = 1, Y = y$ corresponding to $X = y$;$D = 2, Y = y$ corresponding to

$X \in -\infty, y]$; and $D = 3, \vec{Y} = (y_1, y_2)$ corresponding to $X \in (y_1, y_2]$. Without loss of generality assume $P[Y_1 < Y_2] = 1$, since the class $D = 1$ exists. We assume the censoring mechanism satisfies

$$P[X \in F|Y, D] = \frac{P[X \in EF]}{P[X \in E]},$$

where (Y, D) corresponds to the event $X \in E$.

For this situation Theorem 2.1 can be shown to apply if we assume A3) is satisfied, or equivalently that F_x is identifiable from F_y. For instance it applies in the double censoring case where Chang and Yang (1987) have shown that F_x is identifiable under somewhat stronger assumptions than $P[D = 3] = 0, P[D = 1, X \in F] > 0$ for every open set F, and $P[Y \leq y, D = i]$ continuous for $i = 0, 1, 2$.

It is of interest what conditions are necessary to allow identifiability in this general interval censoring framework. A conjecture is that $P[D = 1, X \in F] > 0$ for every open F suffices, but is not necessary.

Let $S_u(t) = P[Y > t, D = 1]; S_r(t) = P[Y > t, D = 0]; S_l(t) = P[Y > t, D = 2]$; and $S_b(s, t) = P[Y_1 > s, Y_2 > t, D = 3]$. For A0) we let

$$\begin{aligned} H(F_y, G_x)(t) &= \bar{G}_x(t) - S_u(t) + \int f_1(l; t, G_x) DS_r(l^-) \\ &\quad + \int f_2(u; t, G_x) DS_l(u^-) + \int f_3(l, u; t, G_x) DS_b(l^-, u^-), \end{aligned}$$

where

$$\begin{aligned} f_1(l; t, G_x) &= \frac{\bar{G}_x(t \vee l)}{\bar{G}_x(l)}, \\ f_2(u; t, G_x) &= 1 - \frac{G_x(t \wedge u)}{G_x(u)}, \\ \text{and } f_3(l, u; t, G_x) &= \frac{G_x(u) - G_x(u \wedge t \vee l)}{G_x(u) - G_x(l)}. \end{aligned}$$

For A1) we use the empirical distribution function. We assume A3) is true. We prove the remaining assumptions as lemmas for completeness.

Lemma 3.4 *For the interval censoring problem,*

$$\sup_{G_x} \| H(F_y^n, G_x) - H(F_y, G_x) \|_\infty \rightarrow 0$$

almost surely, for any sequence F_y^n with $\| F_y^n - F_y \|_\infty \rightarrow 0$ a.s.

Proof: Note that the required convergence holds for $G_x \notin \mathcal{G}$. To see this let $Y(Y_n)$ have distribution function $F_y(F_y^n)$. For $G_x \notin \mathcal{G}$, there exists t_0 such that one of the following is true: 1) $\bar{G}_x(t_0) = 0$ but $P[Y \geq t_0, D = 0] > 0$, 2) $G_x(t_0) = 0$ but $P[Y \leq t_0, D = 2] > 0$, 3) $G_x(t_0) = 0$ but $P[Y_1 \leq t_0, Y_2 \leq t_0, D = 3] > 0$, or 4) $\bar{G}_x(t_0) = 0$ but $P[Y_1 \geq t_0, Y_2 > t_0, D = 3] > 0$. Note that G_x having discrete components causes no difficulty in the denominator of f_3 since S_b places no mass along the diagonal $Y_1 = Y_2$. In any of these cases for large enough n the empirical measure assigns mass to the indicated sets and $\| H(F_y^n, G_x) \|_\infty = \infty$. For $G_x \in \mathcal{G}$, let Ω be the set where the convergence of F_y^n holds. Note that on Ω, all the empirical subsurvival functions converge uniformly and the same argument as in the right censoring case applies. □

Lemma 3.5 *For any sequence of distributions G_n with $\| G_n - G_x \|_\infty \to 0$ almost surely and $G_x \in \mathcal{G}$, we have $\| H(F_y, G_n) - H(F_y, G_x) \|_\infty \to 0$ almost surely.*

Proof: Let Ω be the set where $\| G_n - G_x \|_\infty \to 0$. On Ω

$$\begin{aligned}
\| H(F_y, G_n) - H(F_y, G_x) \|_\infty \leq\ & \| G_x - G_n \|_\infty + \int \sup_t |f_1(l; t, G_x) - f_1(l; t, G_n)| DS_r(l^-) \\
& + \int \sup_t |f_2(u; t, G_x) - f_2(u; t, G_n)| DS_l(u^-) \\
& + \int \sup_t |f_3(l, u; t, G_x) - f_3(l, u; t, G_n)| DS_b(l^-, u^-)
\end{aligned}$$

By bounded convergence it suffices to show that the integrand converges to zero pointwise on the supports of the subsurvival functions. On these either G_x or $\bar{G}_x$ is bounded away from zero since $G_x \in \mathcal{G}$. We illustrate the argument for the integrand involving S_b. Now

$$\begin{aligned}
& \sup_t |f_3(l, u; t, G_n) - f_3(l, u; t, G_x)| \leq \\
& \leq\ 1(G_x(u) > 0, G_n(u) = 0) + 1(G_x(l) < 1, G_n(l) = 1) \\
& \quad + \sup_t \left| \frac{G_x(u) - G_x(u \wedge t \vee l)}{G_x(u) - G_x(l)} - \frac{G_n(u) - G_n(u \wedge t \vee l)}{G_n(u) - G_n(l)} \right| \\
& \quad \times (G_x(u) - G_x(l) > 0, G_n(u) - G_n(l) > 0, l < t < u).
\end{aligned}$$

The first two terms converge to zero. For the last term, fix $0 < \epsilon < 0.5$. Take N_1 so large that $|(F_n(u) - F_n(l)) - (F_x(u) - F_x(l))| < 2^{-1}\epsilon(F_x(u) - F_x(l))$ for $n > N_1$. Take $N_2 > N_1$ so large that $\sup_t |(F_x(u) - F_x(u \wedge t \vee l)) - (F_n(u) - F_n(u \wedge t \vee l))| < 2^{-1}\epsilon(1 - \epsilon)$ whenever

$n > N_2$. Then by checking the maximum positive and negative values for the last term it can be shown to be smaller than ϵ on Ω for $n > N_2$. □

Lemma 3.6 *There exist δ and ϵ such that for $F_x^* \notin Q_\epsilon \equiv \{G_x : \| G_x - F_x \|_\infty < \epsilon\}$ we have $\| H(F_y, F_x^*) \|_\infty > \delta > 0$.*

Proof: Let $\gamma = S_u(0) = P[D = 1] > 0$, and let $\delta = \gamma/4$. Let $\epsilon = 1 - 2\delta$. Assume we can find t_0 such that $F_x^*(t_0) > \epsilon - \delta + F_x(t_0)$. Then

$$\begin{aligned}
&\int \mathrm{E}_{F_x^*}[1(X \leq t_0)|\vec{Y} = \vec{y}]dF_y(\vec{y}) - F_x(t_0) = \\
&= \int \mathrm{E}_{F_x^*}[1(X \leq t_0)|\vec{Y} = \vec{y}]dF_y(\vec{y}) - \int \mathrm{E}_{F_x}[1(X \leq t_0)|\vec{Y} = \vec{y}]dF_y(\vec{y}) \\
&\leq 1 - \gamma
\end{aligned}$$

so that $H(F_y, F_x^*)(t_0) < -\delta$. The case when $F_x^*(t_0) < F_x(t_0) - \epsilon + \delta$ is similar. □

Finally we need to check that $H(F_y, F_x) = 0$. This is straightforward using the decomposition

$$P[X > t] = \sum_{i=0}^{3} P[X > t, D = i].$$

4 A smoothing method for bivariate censored data.

In the bivariate problem, there are generally many self-consistent estimators for sample data which need not converge to any limit. The function H defined at (2.1) is not well defined for all pairs of distributions, and even after this is fixed, A2) and A4) both fail. If $F_n(x)$ is a discrete distribution then

$$\mathrm{E}_{F_n}[1(\vec{X} \leq \vec{t})|\vec{Y} = \vec{y}]$$

is not well defined if $\vec{Y}$ is a singly censored observation, say $(s+, t)$, unless $F_n\{(s+, t)\} > 0$. In the univariate case when the expected value was not defined, it did not particularly matter. In the bivariate case, the definition of this expected value for singly censored observations is problematic. A4) will fail if F_x has an absolutely continuous distribution since

$$\mathrm{E}_{F_x^n}[1(\vec{X} \leq \vec{t})|\vec{Y} = \vec{y}] \not\to \mathrm{E}_{F_x}[1(\vec{X} \leq \vec{t})|\vec{Y} = \vec{y}].$$

The problem only happens for singly censored observations, so a natural method is to treat singly censored observations separately and then use self consistency for the doubly censored observations.

To do this, we first consider the special case where all censoring occurs in both variables or none. Throughout this section, let A be such that $\vec{a} \in A$ implies $P[\vec{Y} > \vec{a}] > 0$.

4.1 Exclusively double censoring.

We first consider the case when $P[D_1 = D_2] = 1$. This means that observations are either uncensored or doubly censored. In this case, Theorem 2.1 applies. We need to check that assumptions A0) - A5) are satisfied.

For A0), the expectation (2.1) is well defined unless $D_1 = D_2 = 0$ and the support of G_x is disjoint from the region $(Y_1, \infty) \times (Y_2, \infty)$. Let

$$\begin{aligned} S_{uu}(\vec{t}) &= P[\vec{Y} > \vec{t}, D_1 = D_2 = 1] \\ \text{and } S_{cc}(\vec{t}) &= P[\vec{Y} > \vec{t}, D_1 = D_2 = 0]. \end{aligned} \tag{4.1}$$

Then we can generally define H by

$$H(F_y, G_x)(\vec{t}) = \bar{G}_x(\vec{t}) - S_{uu}(\vec{t}) + \int \frac{\bar{G}_x(\vec{t} \vee \vec{y})}{\bar{G}_x(\vec{y})} DS_{cc}(\vec{y}^{\,-}), \tag{4.2}$$

where the maximum is taken componentwise, and $\vec{y}^{\,-} = (y_1^-, y_2^-)$. For A1), we will use any sequence of distributions with $\| F_y^n - F_y \|_\infty \to 0$ almost surely, for example the empirical distribution function of the incomplete data. The uniqueness of the estimator follows from the identifiability of F_x from F_y according to the remarks in Section 1. Dabrowska (1988) has shown this identifiability on the set A. As in the univariate case we can check A2), A4), and A5) hold on A. The arguments parallel the univariate case.

Finally we need to check $\| H(F_y, F_x) \|_\infty = 0$. This follows readily from

$$\begin{aligned} P[\vec{X} > \vec{t}, D_1 = D_2 = 0] &= -\int P[\vec{X} > \vec{t} | \vec{Y} = \vec{y}, D_1 = D_2 = 0] DS_{cc}(\vec{y}^{\,-}) \\ &= -\int \frac{\bar{G}_x(\vec{t} \vee \vec{y})}{\bar{G}_x(\vec{y})} DS_{cc}(\vec{y}^{\,-}). \end{aligned}$$

Since the self-consistent estimator based on F_y is unique, it is the GMLE (see Theorem 2.2 of Tsai and Crowley, 1985). This shows uniform consistency of any sequence of self-consistent estimators for this problem, in particular the generalized maximum likelihood estimator. The estimator may be easily computed using the EM algorithm.

4.2 General bivariate right censoring.

The general case is handled in two steps. First the singly censored observations are handled via nonparametric smoothing techniques and then the results of Subsection 4.1 are applied. To this end, recall (4.1) and additionally define

$$S_{uc}(\vec{t}) = P[\vec{Y} > \vec{t}, D_1 = 1, D_2 = 0]$$
$$\text{and } S_{cu}(\vec{t}) = P[\vec{Y} > \vec{t}, D_1 = 0, D_2 = 1].$$

Now let F_z be a distribution on $\mathcal{Y}$ defined by the subsurvival functions

$$\tilde{S}_{uu}(\vec{t}) = S_{uu}(\vec{t}) - \int_{y_2>t_2} \frac{D_2\bar{F}_x(\vec{t} \vee \vec{y})}{D_2\bar{F}_x(\vec{y})} DS_{cu}(\vec{y}^-) - \int_{y_1>t_1} \frac{D_1\bar{F}_x(\vec{t} \vee \vec{y})}{D_1\bar{F}_x(\vec{y})} DS_{uc}(\vec{y}^-)$$

and $\tilde{S}_{cc}(\vec{t}) = S_{cc}(\vec{t})$. Here D_i is the Riemann-Stieltjes partial differential operator for the i^{th} variable. Suppose we are given a sequence of estimators of F_z, say F_z^n. If we define $\hat{F}_x^n$ as a solution of $H(F_z^n, G_x) = 0$ where H is given at (4.2), then by the results in Subsection 4.1, $\hat{F}_x^n$ is uniformly consistent for F_x if $H(F_z, F_x) = 0$ and $\| F_z^n - F_z \|_\infty \to 0$ almost surely. We first check that $H(F_z, F_x) = 0$ and then find an appropriate sequence F_z^n.

Now

$$\begin{aligned} P[\vec{X} > \vec{t}, D_1 = 1, D_2 = 0] &= -\int P[\vec{X} > \vec{t} | \vec{Y} = \vec{y}, D_1 = 1, D_2 = 0] DS_{uc}(\vec{y}^-) \\ &= -\int_{y_1>t_1} P[X_2 > t_2 | \vec{Y} = \vec{y}, D_1 = 1, D_2 = 0] DS_{uc}(\vec{y}^-) \\ &= -\int_{y_1>t_1} \frac{D_1\bar{F}_x(\vec{t} \vee \vec{y})}{D_1\bar{F}_x(\vec{y})} DS_{uc}(\vec{y}^-) \end{aligned}$$

which suffices to show $H(F_z, F_x) = 0$.

We find a sequence F_z^n which converges to F_z. Let $T_1(w_1, y_1, y_2) = P[X_1 > w_1 | X_1 > y_1, X_2 = y_2]$ and $T_2(w_2, y_2, y_1) = P[X_2 > w_2 | X_2 > y_2, X_1 = y_1]$. On A we can write

$$\tilde{S}_{uu}(\vec{t}) = S_{uu}(\vec{t}) - \int_{y_2>t_2} T_1(t_1, y_1, y_2) DS_{cu}(\vec{y}^-) - \int_{y_1>t_1} T_2(t_2, y_2, y_1) DS_{uc}(\vec{y}^-).$$

Let F_z^n on A be given by the subsurvival functions

$$\tilde{S}_{uu}^n(\vec{t}) = S_{uu}^n(\vec{t}) - \int_{y_2>t_2} \hat{T}_1(t_1, y_1, y_2) DS_{cu}^n(\vec{y}^-) - \int_{y_1>t_1} \hat{T}_2(t_2, y_2, y_1) DS_{uc}^n(\vec{y}^-)$$

and $\tilde{S}^n_{cc}(\vec{t}) = S^n_{cc}(\vec{t})$, where a superscript n denotes an empirical subsurvival function and $\hat{T}_1$ is determined in the following manner: $\hat{T}_2$ is determined similarly.

For $\hat{T}_1$, we use a weighted product limit estimator, the weights determined by the distance between X_{2i} and y_2. Let

$$\begin{aligned} Q_u(w_1, y_1, y_2) &= \sum_i W_{ni}(y_2) 1[Y_{1i} > (w_1 \vee y_1), D_{1i} = D_{2i} = 1], \\ \text{and } Q(w_1, y_1, y_2) &= \sum_i W_{ni}(y_2) 1[Y_{1i} > (w_1 \vee y_1), D_{2i} = 1], \end{aligned}$$

where W_{ni} is a sequence of weight functions to be specified. Then $\hat{T}_1$ is the product limit estimator of Q_u with respect to Q, that is

$$\hat{T}_1(w_1, y_1, y_2) = \left[\gamma\left(\Lambda^0_Q\right)\right](w_1) = \lim_{\max_{1 \le k \le r}(u_k - u_{k-1}) \to 0} \prod_{i=1}^{r} \{1 - [\Lambda^0_Q(u_i) - \Lambda^0_Q(u_{i-1})]\},$$

where $0 = u_0 < u_1 < \cdots < u_r = w_1$, and

$$\Lambda^0_Q(t) = -\int_0^{t+} \frac{D_1 Q_u(w_1, y_1, y_2)}{Q(w_1^-, y_1, y_2)}.$$

These estimates are closely related to the conditional survival function estimators proposed by Tsai, Leurgans, and Crowley (1986). Let $S_1(y_1|y_2) = P[X_1 > y_1|X_2 = y_2]$ and note $T_1(w_1, y_1, y_2) = S_1((w_1 \vee y_1)|y_2)/S_1(y_1|y_2)$. Define

$$\begin{aligned} P_u(y_1, y_2) &= \sum_i W_{ni}(y_2) 1[Y_{1i} > y_1, D_{1i} = D_{2i} = 1], \\ \text{and } P(y_1, y_2) &= \sum_i W_{ni}(y_2) 1[Y_{1i} > y_1, D_{2i} = 1], \end{aligned}$$

and let

$$\hat{S}_1(y_1, y_2) = \left[\gamma\left(\Lambda^0_P\right)\right](y_1),$$

where

$$\Lambda^0_P(t) = -\int_0^{t+} \frac{D_1 P_u(y_1, y_2)}{P(y_1^-, y_2)}.$$

Note that

$$\hat{T}_1(w_1, y_1, y_2) = \frac{\hat{S}_1((w_1 \vee y_1), y_2)}{\hat{S}_1(y_1, y_2)}.$$

The conditional subsurvival function estimators proposed by Tsai, Leurgans, and Crowley are $\hat{S}_1(\cdot, \cdot)$. Tsai, Leurgans, and Crowley (1986) give examples of weight functions and give conditions under which $\hat{S}_1$ (and using the same proofs, $\hat{T}_1$) is uniformly consistent. In the following theorem we use a scale estimator in conjunction with the window width. This makes the estimator equivariant under changes of scale. In practice, this is unnecessary as it will be subsumed by choice of the smoothing constant $h(n)$.

Theorem 4.1 *Let*

$$W_{ni}(y_2) = \frac{D_{2i}\, k[(X_{2i} - y_2)/(s_2 h(n))]}{\sum_j D_{2j}\, k[(X_{2j} - y_2)/(s_2 h(n))]},$$

where k is a nonnegative function of bounded variation, $h(n) > 0$, and s_2 is a scale estimate for the distribution of X_2. The weights for the singly censored variables with X_2 censored are defined similarly. Assume

A0) *A is such that $\vec{a} \in A$ implies $P[\vec{Y} > \vec{a}] \geq \delta > 0$ for some $\delta > 0$.*

A1) *$\vec{X}$ and $\vec{C}$ are independent.*

A2) *The distributions of $\vec{X}$ and $\vec{C}$ are absolutely continuous with respect to Lebesgue measure on the plane, and the density for $\vec{X}$ is continuous in each variable.*

A3) *The bandwidths $h(n)$ satisfy $h(n) \to 0$ and $\sum_{n=1}^{\infty} exp(-rnh^2(n)) < \infty$ for every positive r.*

Then $\| F_z^n - F_z \|_\infty \to 0$ a.s.

Proof: It suffices to show that $\tilde{S}^n_{uu}$ is uniformly consistent. Lemma 3.2 of Tsai, Leurgans, and Crowley (1986) shows that $\hat{T}_1$ is uniformly consistent under these assumptions since A2) implies $P[X_1 > y | X_2 = w]$ is uniformly continuous on A. The result then follows from standard methods (e.g. the bivariate extension of Lemma 6 of Aalen, 1976). □

We suspect the theorem is true under weaker conditions than A0) and A2). These assumptions are imposed to ensure convergence of the conditional survival curve estimator, and any improvements in the conditions under which such estimation is possible will provide improvements in this theorem. Although the weak convergence results are not obtained here, it may be worth noting that the estimates will likely converge at rate $n^{-1/2}$. Although nonparametric smoothing techniques are used, they determine the convergence rate of the hazard, not the cumulative hazard.

5 Comments.

In this section, we discuss self-consistency properties of the estimator, indicate some modifications to the estimator for tied data, indicate methods to generalize the estimator to higher dimensions, and discuss features of the estimator.

The estimator $\hat{F}_x^n$ may be thought of in the following way. Begin by assigning mass $1/n$ to each observation. For the singly censored observations, redistribute this mass to possible complete observations which could have generated the incomplete observation according to the weighted conditional product limit estimator $\hat{T}_1$ (or $\hat{T}_2$ as appropriate). Then redistribute the mass associated with the doubly censored observations in the unique self-consistent manner given mutual independence of $\vec{X}$ and $\vec{C}$.

Although the estimator is not self-consistent under bivariate right censoring, the estimator is self-consistent under generalized univariate censoring for absolutely continuous distributions, that is, censoring when $P[C_1 = kC_2] = 1$ for some $k > 0$. The case $k = 1$ gives univariate censoring.

Theorem 5.1 *The estimator $\hat{F}_x^n$ is self-consistent under generalized univariate censoring with probability one if the distribution of $\vec{X}$ is absolutely continuous.*

Proof: From the construction of $\hat{F}_x^n$, the mass of the uncensored and doubly censored observations is assigned in a self-consistent manner. We need only check the singly censored observations have mass assigned self-consistently, that is for an observation $(Y_1 = y_1, Y_2 = y_2, D_1 = 0, D_2 = 1)$,

$$\frac{\hat{F}_x^n(B(w_1))}{\hat{F}_x^n(B(y_1))} = \hat{T}_1(w_1, y_1, y_2)$$

for all $w_1 \geq y_1$, where $B(w) = (w, \infty) \times y_2$. This is true if there are no other uncensored or singly censored observations in $B(w_1)$ and no doubly censored observations $\vec{X} \in E$ such that $EB(w_1)$ is not the empty set or $B(w_1)$. There are no uncensored or singly censored observations in $B(w_1)$ with probability one since $\vec{X}$ has an absolutely continuous distribution, and for any doubly censored observation $\vec{X} \in E$, $EB(w_1)$ is either the empty set or $B(w_1)$ because of the generalized univariate censoring scheme. □

The estimator is not self-consistent in general: doubly censored observations which assign mass to only a portion of the set associated with a singly censored observation will cause the mass associated with the singly censored observation to be assigned self-inconsistently.

Example 5.1 *If we have observations* $(1+,3),(4,4),(5+,1+)$, *and* $(6,2)$ *and choose a rectangular kernel with* $h > 1$, *then the estimator* $\hat{F}_x^n$ *will assign mass 1/8 to (4,3), 1/4 to (4,4), 5/24 to (6,3), and 5/12 to (6,2). This is not self-consistent: the mass associated with (1+,3) is not assigned in a self-consistent manner because part of the mass of (5+,1+) gets assigned to the point (6,2).*

We next consider tied data. Smoothing was necessary since information about the distribution of $\vec{X}$ within the region specified by the singly censored observations could not be obtained directly for absolutely continuous distributions. But if singly censored observations arise, say $(s+,t)$ and and there exist other observations in the set $(s,\infty) \times \{t\}$ then smoothing is unnecessary. This leads to a modified estimator which is uniformly consistent for distributions which are either entirely discrete or entirely absolutely continuous and for which Theorem 5.1 holds under the same conditions. The estimator may be described as follows. Assign mass $1/n$ to each of the uncensored points. Next, assign mass $1/n$ to each of the singly censored observations by the following method. Each of these corresponds to a set E where the singly censored observation conveys the same information as $\vec{X} \in E$. Consider all the sets E which contain mass either from uncensored points or because they are supersets of the sets associated with other singly censored points. Assign the mass associated with these singly censored observations in the unique self-consistent manner (unique by the arguments in Subsection 4.1). For the remainder of the mass unspecified for the singly censored observations use the nonparametric smoothing technique. Finally assign mass $1/n$ to each doubly censored observation and redistribute it in the unique self-consistent manner. This method essentially amounts to using a smoothing window width of zero if smoothing is unnecessary. For discrete data, this means for large enough sample sizes no smoothing will take place and the estimator will be the GMLE. For absolutely continuous data, the estimator will be the estimator described in Section 4.

The application of these techniques to higher dimensional data is not entirely straightforward. We briefly describe the problems that will be encountered. For simplicity we assume all distributions involved are absolutely continuous. For uncensored observations, assign mass $1/n$ to each point. For singly censored observations, follow the techniques described using only observations with all uncensored values near those of the singly censored observation being considered. With doubly censored observations, there is an additional problem to be considered. Suppose we have observations (3+,1+,2) and (2+,2+,2.1). Then these observations need to be considered together since the second suggests mass may be more prevalent in the (3+,2+,2) than just considering the previous mass and the first observation might suggest. The singly censored observations could be considered individually but this is not true for more highly censored observations. Similar techniques using self-consistency and smoothing ideas are probably possible in higher dimensions, but they will be more difficult than those considered here.

The problems presented by higher dimensional data may also be reduced by consideration of other models than $\vec{X}$ and $\vec{C}$ mutually independent. For instance using the model X_i independent of C_i for fixed levels of the other X and C variables, as considered in Pruitt (1990) for the bivariate case, makes the independence conditions of a local rather than global nature. In this situation, smoothing ideas are more natural, since they only use local data.

We finally say some words about why we prefer our estimator to those that have been proposed. We consider both our estimator and the TLC estimator to be acceptable estimators and so see little reason to use an obviously incoherent estimator which assigns negative mass. The efficiency properties of the TLC and this estimator remain to be addressed however. The GMLE is attractive for discrete problems, but its nonuniqueness in general means some modification of it is necessary to makes its use practical. This paper provides one possible modification which retains some of its features.

We finally come to the TLC estimator to which ours is related by the use of nonparametric smoothing techniques and a decomposition of a bivariate survival curve. We prefer our estimator for the following reasons: it is redistributive, it is affine equivariant, and it is more smoothly behaved than the TLC estimator. We illustrate the differences by examining Example 5.1. The TLC estimator for Example 5.1 assigns mass 1/2 to (4,4) and 1/2 to (6,2). The behavior of the estimator for various transformations is listed in Table 1. Clearly the estimator is not redistributive or affine equivariant. Non affine equivariance means the same data can lead to different results, e.g. if X_1 is time since 1980 and X_2 is a time since X_1, someone else doing the analysis with X_1^* being time since 1970 might get different results. It is interesting to examine the behavior as the observation (1+,3) is shifted toward (4+,3). As can be seen from Table 1, there are sharp jumps in the behavior of the estimator. The first jump moves mass outside of the set (5+,1+) which seems counterintuitive. The estimator only assigns mass to uncensored points and singly censored observations with censored values larger than uncensored values which tends to make the estimator very rough (having large jumps) for small samples. The estimator proposed here has the same value for all transformations listed in Table 1.

The decomposition on which the TLC estimator is based is not very intuitive. Contrast this with the decomposition used here $P[\vec{X} > \vec{t}] = \sum\sum P[\vec{X} > \vec{t}, D_1 = i, D_2 = j]$, which is a direct generalization of that used in the univariate case. Only one of the summands is directly estimable, but the others can be estimated using univariate product limit methods and self-consistency. To compute the estimator for a rectangular kernel, it is only necessary to assign mass $1/n$ to each of the observed points. For the singly censored points this mass is spread over the possible values using a univariate product limit estimator on the nearby values, and for the doubly censored points, the mass is spread in the unique self-consistent manner. This is easy to understand based on the concepts of univariate product limit estimation and self-consistency, and it seems a natural estimator to use in this problem.

References

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

CAMPBELL, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika* **68** 417-422.

CAMPBELL, G., and FÖLDES, A. (1982). Large sample properties of nonparametric bivariate estimators with censored data. In *Nonparametric Statistical Inference, Colloquia Mathematica-Societatis János Bolyai* (B.V. Gnedenko, M.L. Puri, and I. Vincze, eds.). North Holland, Amsterdam.

CHANG, M.N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18** to appear.

CHANG, M.N. and YANG, G.L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15** 1536-1547.

DABROWSKA, D.M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.* **16** 1475-1489.

EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* 4 831-853. Univ. of Calif. Press.

GILL, R.D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. Part I, *Scand. J. Statist.* **16** 97-128.

HANLEY, J.A., and PARNES, M.N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics* **39** 129-139.

LANGBERG, N.A. and SHAKED, M. (1982). On the identifiability of multivariate life distribution functions. *Ann. Probab.* **10** 773-779.

LEURGANS, S., TSAI, W.-Y., and CROWLEY, J. (1982). Freund's bivariate exponential distribution and censoring. In *Survival Analysis* (J. Crowley and R. A. Johnson, eds.) 230-242. IMS, Hayward, Calif.

MUÑOZ, A. (1980). Nonparametric estimation from censored bivariate observations. *Technical Report* **60**, Stanford University.

PRUITT, R. (1990). Identifiability of discrete bivariate survival curves from censored data. *Technical Report* **535**, University of Minnesota.

TSAI, W.-Y. (1986). Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian estimation application. *Ann. Statist.* 14 238-249.

TSAI, W.-Y. and CROWLEY, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *Ann. Statist.* **13** 1317-1334; Correction **18**, to appear.

TSAI, W.-Y., LEURGANS, S., and CROWLEY, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* 14 1351-1365.

TURNBULL, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169-173.

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

transformation	mass assigned to A - (4,4)	B - (6,2)	C - (4,3)	D - (6,3)
$(1+,3) \rightarrow (k+,3), 1 \leq k < 2$	1/2	1/2	-	-
$(1+,3) \rightarrow (k+,3), 2 \leq k < 3$	2/3	1/3	-	-
$(1+,3) \rightarrow (k+,3), 3 \leq k < 4$	1/3	1/3	1/6	1/6
$(x,y) \rightarrow (x, y+5)$	1/3	2/3	-	-
$(x,y) \rightarrow (x+3, y)$	1/3	1/3	1/6	1/6
$(x,y) \rightarrow (x, 5y)$	1/3	2/3	-	-
$(x,y) \rightarrow (4x, y)$	1/3	1/3	1/6	1/6

Table 1: The Tsai, Leurgans, and Crowley estimator for various transformations of Example 1. The mass assigned in column A is assigned to the point (4,4) under the given transformation.