Identifiability of Discrete Bivariate Survival Curves from Censored Data

By

Ronald C. Pruitt

Technical Report No. 535

School of Statistics

University of Minnesota

8 January 1990

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

Abstract

We show that the survival curve is identifiable in discrete bivariate censored data problems under weaker independence assumptions than have been commonly been made. The common assumtion has been mutual independence of (T_1, T_2) and (Z_1, Z_2) where (T_1, T_2) is the true survival vector and (Z_1, Z_2) is a nuisance censoring vector. We show that the distribution of (T_1, T_2) is identifiable under weaker independence assumptions if the distribution of (T_1, T_2, Z_1, Z_2) has rich enough support.

1 Introduction.

We wish to infer about a bivariate distribution $\vec{T} = (T_1, T_2)$ subject to censoring. Assume $\vec{T}$ and the censoring variable $\vec{Z} = (Z_1, Z_2)$ are defined on a common probability space $(\Omega, \mathcal{F}, P)$ and have survival functions $F(s,t) = \Pr(T_1 > s, T_2 > t)$ and $G(s,t) = \Pr(Z_1 > s, Z_2 > t)$. The observable random variables are given by $\vec{Y} = (Y_1, Y_2)$ and $\vec{\delta} = (\delta_1, \delta_2)$, where $Y_i = T_i \wedge Z_i$ and $\delta_i = 1[T_i = Y_i]$, for $i = 1, 2$. To estimate F, suppose we have a sample $(\vec{Y}_i, \vec{\delta}_i)$, $i = 1, \ldots, n$ which consists of independent, identically distributed copies of $(\vec{Y}, \vec{\delta})$.

The usual assumption used in this problem in order to make the distribution of $\vec{T}$ identifiable has been to assume $\vec{T}$ and $\vec{Z}$ to be independent. This approach has been pursued by Muñoz (1980, Campbell (1981), Langberg and Shaked (1982), Leurgans, Tsai, and Crowley (1982), Campbell and Földes (1982), Hanley and Parnes (1983), Tsai, Leurgans, and Crowley (1986), and Dabrowska (1988), among others. Here we show that for discrete distributions with full support, the distribution of $\vec{T}$ is identifiable under the much weaker independence conditions

$$(1) \qquad \begin{aligned} &T_1 \text{ independent of } Z_1 | T_2, Z_2 \\ \text{and} \quad &T_2 \text{ independent of } Z_2 | T_1, Z_1. \end{aligned}$$

The assumption of full support is stronger than that needed when assuming mutual independence. There is a trade-off between the two assumptions. In particular situations one or the other may be more reasonable to make.

The outline of the paper is as follows. First, we explore some general aspects of censored data problems, and try to explain what the goals are when showing identifiability. We define three notions related to coherence and explore some aspects of these for the univariate right censoring case. For the bivariate case the usual assumption of mutual independence seems quite different than the assumption in the univariate case. This leads us to a closer look at the univariate case for discrete data, which allows us to extend some of the ideas to the bivariate case. An estimator and its consistency are obtained for the discrete bivariate censored data problem with full support.

2 Internal consistency.

In this section we wish to discuss a concept which we call internal consistency of an estimator for an incomplete data problem, which helps to illustrate the difference between the univariate and bivariate censored data problems. We describe an incomplete data problem in the following fashion. Suppose $\vec{X}_1, \ldots, \vec{X}_n$ are iid random vectors on $(\mathcal{X}, \sigma(\mathcal{X}))$ with distribution function F_x. Instead of observing the $\vec{X}_i$ we observe $\vec{W}_1, \ldots, \vec{W}_n$, which are iid random vectors on $(\mathcal{W}, \sigma(\mathcal{W}))$ where $\vec{W}_i = H(\vec{X}_i)$ and H is a mapping from $\mathcal{X}$ into $\mathcal{W}$. We wish to estimate the unknown distribution function F_x or some functional of it.

Associated with the mapping H is a mapping $\tilde{H}$ from $\mathcal{P}_x$ to $\mathcal{P}_w$ where $\mathcal{P}_x$ is a family of distributions on $(\mathcal{X}, \sigma(\mathcal{X}))$, and $\mathcal{P}_w$ is a family of distributions on $(\mathcal{W}, \sigma(\mathcal{W}))$. The mapping is defined as follows. Suppose $\vec{X}$ has distribution function $F_x, \vec{W} = H(\vec{X})$, and $\vec{W}$ has distribution function F_w. Then $\tilde{H}(F_x) = F_w$. The mapping $\tilde{H}$ is generally many-to-one in a censored data problem. The general approach to a censored data problem is to find a restricted domain for the function $\tilde{H}$, say $\mathcal{A}$, which maps into $\mathcal{B}$, denote this restricted version of $\tilde{H}$ by H^*. The mapping H^* is chosen to have the following properties:

1. ${H^*}^{-1}(F_w)$ is a set of probability measures which coincide on $\mathcal{F}(F_w) \subset \sigma(\mathcal{X})$. Here $F_w \in \mathcal{B}$, and $\mathcal{F}(F_w)$ should be a suitably large σ-field which may depend on F_w.

2. The set $\mathcal{B}$ is large in some suitable sense.

We then say that F_x is identifiable from censored data for any distribution in $\mathcal{A}$ on $\mathcal{F}(F_w)$. The goal is to find good estimators of F_x based on the incomplete data $\vec{W}_1, \ldots, \vec{W}_n$. Let F_w^n be the empirical distribution function of $\vec{W}_1, \ldots, \vec{W}_n$.

Definition 1 *An estimator* $\hat{F}_x^n$ *of* F_x *is* internally consistent *if* $H^*(\hat{F}_x^n) = F_w^n$.

Internal consistency seems a very natural property to desire in an estimator. In particular, properties of the resulting estimator can be studied based on the well-known behavior of F_w^n and properties of the mapping $(H^*)^{-1}$. This concept is related to the notion of self-consistency originally formulated by Efron (1967). The notion of self-consistency involves a decomposition of the complete data into $(\vec{T}, \vec{Z})$ where $\vec{T}$ is of interest and $\vec{Z}$ is not.

Definition 2 *An estimator $\hat{F}_x^n$ of F_x based on $\vec{W}_1, \ldots, \vec{W}_n$ is* self-consistent *if*

$$\hat{F}_T^n(\vec{t}) = \int \mathrm{E}_{\hat{F}_x^n}[\vec{T} \le \vec{t} | \vec{W} = \vec{w}] dF_w^n(\vec{w})$$

whenever $\{\vec{T} \le \vec{t}\} \in \mathcal{F}(F_w^n)$, where $\hat{F}_T^n$ is the marginal distribution for $\vec{T}$ of $\hat{F}_x^n$.

Note that the definition requires specification of an estimator for F_x and not just F_T. The definition may be simplified and the expectation taken with respect to $\hat{F}_T^n$ only if $\vec{T}$ is independent of $\vec{Z}$ in $\mathcal{A}$. This definition coincides with the definition of internal consistency whenever $F_w^n \in \mathcal{B}$. The definition of self-consistency admits self-consistent estimators even when $F_w^n \notin \mathcal{B}$ which is not allowed by the definition of internal consistency. A related weaker notion is that of redistribution of mass.

Definition 3 *An estimator $\hat{F}_T^n$ of F_T is* redistributive *if there exists $\hat{F}_x \in \mathcal{P}_x$ such that $\tilde{H}(\hat{F}_x) = F_w^n$, where $\hat{F}_T^n$ is the marginal distribution for $\vec{T}$ of $\hat{F}_x$.*

An estimator which is not redistributive is not anywhere in the pre-image of F_w^n, and hence is guaranteed not to be the estimator that someone seeing the complete data would use if they were using the empirical distribution function. It is readily apparent that any internally consistent estimator is redistributive.

We now wish to use the notion of internal consistency to explore differences between the univariate and bivariate incomplete data problem. We present results for the univariate case and then some examples for the bivariate case indicating some of the differences.

2.1 The univariate case.

Let $\mathcal{X} = \Re_+$. Here we have $\vec{X} = (T, Z), \vec{W} = (Y, \delta)$, and $H(T, Z) = (Y, \delta) = (T \wedge Z, 1(T \le Z))$. The usual restricted domain is given by

$$\mathcal{A} = \{F_x(\cdot,\cdot) : F_x(t,z) = F_T(t)F_Z(z) \text{ for all } t, z\}.$$

In this case $\mathcal{B}$ is all probability measures. The σ-field $\mathcal{F}(F_w)$ is $\sigma(\mathcal{X})$ restricted to $[0, M) \times [0, M)$ where $M = \sup_t\{F_Y(t) < 1\}$.

We then have the following theorem regarding the Kaplan-Meier estimator.

Theorem 1 *Suppose $F_x \in \mathcal{A}$. Let $\hat{F}_x^n$ be the usual Kaplan-Meier estimator for T and Z with independence between T and Z. Then $\hat{F}_x^n$ is internally consistent.*

Proof: Note that $F_w^n \in \mathcal{B}$. Since the Kaplan-Meier estimator is self-consistent it is internally consistent by the remark following Definition 2. $\square$

2.2 The bivariate case.

Let $\mathcal{X} = \Re_+^2$. Take $\vec{X} = (T_1, T_2, Z_1, Z_2), \vec{W} = (Y_1, Y_2, \delta_1, \delta_2)$, and $H(T_1, T_2, Z_1, Z_2) = (Y_1, Y_2, \delta_1, \delta_2) = (T_1 \wedge Z_1, T_2 \wedge Z_2, 1(T_1 \le Z_1), 1(T_2 \le Z_2))$. The usual restricted domain is given by

$$\mathcal{A} = \{F_x(\cdot,\cdot,\cdot,\cdot) : F_x(t_1, t_2, z_1, z_2) = F_{T_1,T_2}(t_1, t_2) F_{Z_1,Z_2}(z_1, z_2) \text{ for all } t_1, t_2, z_1, z_2\}.$$

We assert $\mathcal{B}$ is a small subset of all probability measures. To justify this statement, we need to specify our measure of size. Certainly one measure we have a practical interest in is the probability that F_w^n is in $\mathcal{B}$ given that F_x is in $\mathcal{A}$. We have the following theorem.

Theorem 2 *Suppose $F_x \in \mathcal{A}$. Also assume the distributions F_{T_1,T_2}, F_{Z_1,Z_2}, and F_{Y_1,Y_2} are absolutely continuous. Then*

$$\lim_{n \to \infty} P[F_w^n \in \mathcal{B}] = 0.$$

Proof: This is a simple consequence of the almost sure convergence of Dabrowska's (1988) bivariate survival curve estimator, and Theorem 9 of Pruitt (1989). Note that Dabrowska's estimator does not depend on the sample size except through empirical subsurvival functions, so that a sample of size $2n$ which consists of two copies of each observation in a sample of size n gives rise to the same estimator. Let D_n be the event that Dabrowska's estimator for a sample of size n from F_w does not assign negative mass. Then the almost sure consistency gives

$$P[F_w^n \in \mathcal{B}] \le P[D_n]$$

and $P[D_n]$ converges to zero by Theorem 9 of Pruitt (1989). $\square$

This is quite different from the univariate case. If we let $\mathcal{A}$ be as above there are no internally consistent estimators, and we have a very large class of possible observed distributions which correspond to no possible distribution for the uncensored data. The assumption of mutual independence of $\vec{T}$ and $\vec{Z}$ has overly restricted the domain of the function $\tilde{H}$. It would be nice to have a weaker independence restriction which widened the set $\mathcal{B}$. This is the direction we try and pursue in the rest of the paper.

As a final note on the usual independence assumptions, note that the assertion that $\vec{T}$ is independent of $\vec{Z}$ may be assessed from the censored data. For instance, one possible method of assessment is as follows. Use a technique to find estimators for $F_{\vec{T}}$ and for $F_{\vec{Z}}$ (by considering $\vec{Z}$ to be of interest rather than $\vec{T}$). Using the independence this gives a joint estimated distribution for $(\vec{T}, \vec{Z})$. This generates a distribution for $(\vec{Y}, \vec{\delta})$ via H. This induced distribution and the observed distribution for $(\vec{Y}, \vec{\delta})$ can then be compared, for instance by discretizing and computing a chi-squared statistic. If mutual independence is satisfied, we would expect estimators making use of this to perform better than estimators which are applicable over a wider range of probability distributions.

3 A look at the Kaplan-Meier estimator.

Here we provide a convenient way to look at the Kaplan-Meier estimator which shows many of its properties. First suppose the distributions of T and Z are discrete, and for convenience suppose they are both concentrated on $\{1,2,\ldots,r\}$. Also assume that $P[Y = r] > 0$, in general $P[T > k]$ is not identifiable for k with $P[Y > k, \delta = 1] = 0$ and $P[Y > k] > 0$. Then there are $2r - 1$ possible observations since $Y = r, \delta = 0$ is impossible. These are illustrated in the (T, Z) plane in Figure 1. Each censored value corresponds to one or more uncensored values. For example the censored value $Y = r - 2, \delta = 1$ corresponds to the uncensored values $T = r - 2, Z \geq r - 2$.

Now in order for the distribution of T to be identifiable, it suffices to have T independent of Z. We have the following theorem.

Theorem 3

$$P[T \geq k+1, Y \geq k-j] = P[Y \geq k+1] \prod_{i=0}^{j} \left(1 + \frac{P[Y = k-i, \delta = 0]}{P[Y > k-i]}\right)$$

for $j = -1, 0, 1, \ldots, k-1$, *and* $k = 0, 1, \ldots, r$. *The product is unity if* $j = -1$.

Before proceeding to the proof, we give the following corollary which shows the distribution of T is identifiable.

Corollary 4

$$P[T \geq k+1] = P[Y \geq k+1] \prod_{i=0}^{k-1} (1 + \frac{P[Y = k-i, \delta = 0]}{P[Y > k-i]}).$$

Note the right hand side is estimable, and if we replace the probabilities with sample frequencies we obtain the Kaplan-Meier estimator.

Proof: [of Theorem 3] Fix k. We use induction on j. The case $j = -1$ is obvious. Now assume the theorem is true for $-1, 0, \ldots, j$. Then

$$P[T \geq k+1, Y \geq k-j-1] = P[T \geq k+1, Y \geq k-j](1 + \frac{P[T \geq k+1, Y = k-j-1]}{P[T \geq k+1, Y \geq k-j]})$$

and

$$\begin{aligned} \frac{P[T \geq k+1, Y = k-j-1]}{P[T \geq k+1, Y \geq k-j]} &= \frac{P[Y = k-j-1 | T \geq k+1]}{P[Y \geq k-j | T \geq k+1]} \\ & \frac{P[Y = k-j-1 | T \geq k-j]}{P[Y \geq k-j | T \geq k-j]} \\ & \frac{P[Y = k-j-1, \delta = 0]}{P[Y \geq k-j]}. \end{aligned}$$

Note the denominators are all greater than zero since $P[Y = r] > 0$ by hypothesis. □

Let the survival function for T be denoted by $S_T(t) = P[T > t]$, and let $S_Y(t) = P[Y > t], S_c(t) = P[Y > t, \delta = 0]$, and $S_u(t) = P[Y > t, \delta = 1]$. Then note we can write the result of Corollary 4 as

$$S_T(k) = S_Y(k) \prod_{i=0}^{k-1} (1 + G(k-i-1) - G(k-i)),$$

where

$$G(t) = \int_0^{t+} \frac{DS_c(u)}{S_Y(u)},$$

where k is an integer and the operator "D" and integral are Riemann-Stieltjes differential operator and integral. In fact,

$$\frac{S_T(k)}{S_Y(k)} = \lim_{n \to \infty} \prod_{i=1}^{n} (1 + G(u_{i-1}) - G(u_i)) \tag{2}$$

where $0 = u_0 < \cdots < u_n = t$ and the limit is taken as $u_i - u_{i-1} \to 0$. In fact, (2) can be shown to hold for general distributions by a discretization argument, so that

$$S_T(t) = S_Y(t) exp(-\oint_0^t \frac{DS_c(u)}{S(u)}) \prod_{u \leq t} (1 - \frac{\Delta S_c(u)}{S(u)}), \tag{3}$$

where the integral $\oint_0^t DF/G$ means integration over the intervals of points less than t for which F is continuous, and $\Delta F(u) = F(u-) - F(u+)$.

In two dimensions we will not have any explicit formulas, so it is useful to look at what is going on in the univariate case a little more closely. The point masses $P[T = t, Z = z]$ can be figured out sequentially by considering the observed values $Y = r; Y = r - 1, \delta = 0; Y = r - 1, \delta = 1; \ldots ; Y = 1, \delta = 1$. This is summarized in the following theorem.

Theorem 5 *The probabilities*

$$P[Y = y, \delta = 1] \qquad y = k \ldots r$$
$$P[Y = y, \delta = 0] \qquad y = k - a \ldots r - 1$$

determine the probabilities

$$P[T = t, Z = z] \quad t = k \ldots r \quad z = k - a \ldots r$$

for $a = 0, 1$, *and* $k = 1, \ldots, r$ *except* $k = 1, a = 1$.

Proof: The proof is by induction on k. The theorem is true for $k = r$ and $a = 0, 1$. We show the inductive step for $a = 0$, the case $a = 1$ is similar. We wish to show that $P[T = t, Z = z]$ for $t = k \ldots r, z = k - 1 \ldots r$, and $P[Y = k - 1, \delta = 1]$ determine $P[T = k - 1, Z = l]$ for $l = k - 1 \ldots r$. Note that $P[Y = k - 1, \delta = 1] = P[T = k - 1, Z \geq k - 1]$. If $P[Y = k - 1, \delta = 1] = 0$, then $P[T = k - 1, Z = l] = 0$ for all $l = k - 1 \ldots r$. If $P[Y = k - 1, \delta = 1] > 0$,

$$\begin{aligned} P[T = k - 1, Z = l] &= P[Z = l | T = k - 1] P[T = k - 1] \\ &= \frac{P[Z = l | T \geq k] P[T \geq k]}{P[Z \geq k - 1 | T \geq k] P[T \geq k]} P[Z \geq k - 1 | T = k - 1] P[T = k - 1] \\ &= \frac{P[T \geq k, Z = l] P[T = k - 1, Z \geq k - 1]}{P[T \geq k, Z \geq k - 1]} \\ &= \frac{P[T \geq k, Z = l] P[Y = k - 1, \delta = 1]}{P[T \geq k, Z \geq k - 1]} \end{aligned}$$

which are all known. □

Corollary 6 *The probabilities*

$$P[Y=y,\delta=1] \qquad y=1\ldots r$$
$$P[Y=y,\delta=0] \qquad y=1\ldots r-1$$

determine the probabilities

$$P[T=t,Z=z] \quad t=1\ldots r \quad z=1\ldots r.$$

4 Generalization to discrete bivariate data.

If we are willing to assume $\vec{T}$ independent of $\vec{Z}$, the distribution of $\vec{T}$ is identifiable for $\vec{t}$ with $S_t(\vec{t})>0$. Here we consider the weaker independence assumption (2). This weaker condition does not allow identification of the distribution of $\vec{T}$ over as wide a range as the mutual independence condition. For convenience, assume the distributions of T_1, T_2, Z_1, and Z_2 are all concentrated on $1,2,\ldots,r$. To guarantee identifiability, we assume that

$$(4) \qquad P[Y_i=y_i,\delta_i=a]>0 \text{ implies } P[Y_i=y_i,\delta_i=a,Y_{3-i}=r,\delta_{3-i}=1]>0$$
$$\text{for } i=1,2, a=0,1, \text{ and } y_i=1,\ldots,r.$$

This is a much stronger condition than that needed in the mutual independence scenario. The condition is satisfied in the important special case when (T_1,T_2,Z_1,Z_2) has full support. One consequence of (5) which we will use repeatedly is

$$(5) \qquad P[Y_1=r,\delta_1=1,Y_2=r,\delta_2=1]=P[T_1=r,Z_1=r,T_2=r,Z_2=r]>0.$$

We can now state the theorem, the identifiability is obtained with $k_2=1, b=0$. It may be useful to make reference to Figure 2.

Theorem 7 *The probabilities*

$$P[Y_1=y_1,\delta_1=1,Y_2=y_2,\delta_2=1] \qquad y_1=1\ldots r \qquad y_2=k_2\ldots r$$
$$P[Y_1=y_1,\delta_1=0,Y_2=y_2,\delta_2=1] \qquad y_1=1\ldots r-1 \qquad y_2=k_2\ldots r$$
$$P[Y_1=y_1,\delta_1=1,Y_2=y_2,\delta_2=0] \qquad y_1=1\ldots r \qquad y_2=k_2-b\ldots r-1$$
$$P[Y_1=y_1,\delta_1=0,Y_2=y_2,\delta_2=0] \qquad y_1=1\ldots r-1 \qquad y_2=k_2-b\ldots r-1$$

determine the probabilities

$$P[T_1 = t_1, Z_1 = z_1, T_2 = t_2, Z_2 = z_2] \quad t_1 = 1 \ldots r \quad t_2 = k_2 \ldots r$$
$$z_1 = 1 \ldots r \quad z_2 = k_2 - b \ldots r$$

for $b = 0, 1$, *and* $k_2 = 1 \ldots r$ *except* $k_2 = 1, b = 1$.

Proof: The proof is by induction on k_2. The case with $k_2 = r$ and $b = 0$ then follows from the univariate result. With $b = 1$ either $P[Y_2 = r - 1, \delta_2 = 0] = 0$ which implies $P[T_1 = t_1, Z_1 = z_1, T_2 = r, Z_2 = r - 1] = 0$ for $t_1 = 1 \ldots r, z_1 = 1 \ldots r$; or else $P[Y_2 = r - 1, \delta_2 = 0] > 0$ in which case $P[T_1 = r, Z_1 = r, T_2 = r, Z_2 = r - 1] > 0$ so the univariate result applies. We show the inductive step for $b = 0$, the other case is similar. The inductive step follows as a special case of Lemma 8 given below. □

Lemma 8 *The probabilities*

$$P[T_1 = t_1, Z_1 = z_1, T_2 = t_2, Z_2 = z_2] \quad t_1 = 1 \ldots r \quad t_2 = k_2 \ldots r$$
$$z_1 = 1 \ldots r \quad z_2 = k_2 - 1 \ldots r$$

and

$$P[Y_1 = y_1, \delta_1 = 1, Y_2 = k_2 - 1, \delta_2 = 1] \qquad y_1 = k_1 \ldots r$$
$$P[Y_1 = y_1, \delta_1 = 0, Y_2 = k_2 - 1, \delta_2 = 1] \quad y_1 = k_1 - a \ldots r - 1$$

determine the probabilities

$$P[T_1 = t_1, Z_1 = z_1, T_2 = t_2, Z_2 = z_2] \qquad t_1 = k_1 \ldots r \qquad t_2 = k_2 - 1$$
$$z_1 = k_1 - a \ldots r \quad z_2 = k_2 - 1 \ldots r$$

for $a = 0, 1$, *and* $k_1 = 1 \ldots r$, *except* $k_1 = 1, a = 1$.

Proof: Note the lemma is trivially true if $P[Y_2 = k_2 - 1, \delta_2 = 1] = 0$. Otherwise, the proof is by induction on k_1. For $k_1 = r$ we check the case when $a = 0$. The case $a = 1$ is similar.

$$\begin{aligned} & P[T_1 = r, Z_1 = r, T_2 = k_2 - 1, Z_2 = l] \\ &= P[Z_2 = l | T_2 = k_2 - 1, T_1 = r, Z_1 = r] P[T_2 = k_2 - 1, T_1 = r, Z_1 = r] \end{aligned}$$

$$
\begin{aligned}
&= \frac{P[Z_2 = l | T_2 \geq k_2, T_1 = r, Z_1 = r] P[T_2 \geq k_2, T_1 = r, Z_1 = r]}{P[Z_2 \geq k_2 - 1 | T_2 \geq k_2, T_1 = r, Z_1 = r] P[T_2 \geq k_2, T_1 = r, Z_1 = r]} \\
&\quad \times P[Z_2 \geq k_2 - 1 | T_2 = k_2 - 1, T_1 = r, Z_1 = r] P[T_2 = k_2 - 1, T_1 = r, Z_1 = r] \\
&= \frac{P[T_2 \geq k_2, Z_2 = l, T_1 = r, Z_1 = r] P[Y_2 = k_2 - 1, \delta_2 = 1, Y_1 = r, \delta_1 = 1]}{P[T_2 \geq k_2, Z_2 \geq k_2 - 1, T_1 = r, Z_1 = r]}
\end{aligned}
$$

Note the denominators are all positive by (5). We now assume the lemma is true for $k_1, k_1 + 1, \ldots, r$ and $a = 0, 1$, and show it is true for $k_1 - 1$ and $a = 0$. The case $a = 1$ is similar. We need to determine $P[T_1 = k_1 - 1, Z_1 = l, T_2 = k_2 - 1, Z_2 = m]$ for $l = k_1 - 1 \ldots r, m = k_2 - 1 \ldots r$. First note that if $P[Y_1 = k_1 - 1, \delta_1 = 1, Y_2 = k_2 - 1, \delta_2 = 1] = 0$ then all these probabilities are zero. Otherwise, $P[Y_1 = k_1 - 1, \delta_1 = 1, T_2 = r, Z_2 = r] > 0$ which implies $P[T_1 = k_1 - 1, Z_1 = r, T_2 = r, Z_2 = r] > 0$ since

$$
P[T_1 = k_1 - 1, Z_1 = r, T_2 = r, Z_2 = r] = P[T_1 = r, Z_1 = r, T_2 = r, Z_2 = r] \times \frac{P[T_1 = k_1 - 1, Z_1 \geq k_1 - 1, T_2 = r, Z_2 = r]}{P[T_1 = r, Z_1 \geq k_1 - 1, T_2 = r, Z_2 = r]}.
$$

Similarly $P[T_1 = r, Z_1 = r, T_2 = k_2 - 1, Z_2 = r] > 0$. Define $p_{lm} = P[T_1 = k_1 - 1, Z_1 = l, T_2 = k_2 - 1, Z_2 = m]$. Then

$$
\frac{p_{lr}}{p_{rr}} = \frac{P[T_1 \geq k_1, Z_1 = l, T_2 = k_2 - 1, Z_2 = r]}{P[T_1 \geq k_1, Z_1 = r, T_2 = k_2 - 1, Z_2 = r]}
$$

Note the denominator is positive which implies $p_{rr} > 0$. We would now wish to say that $p_{lm}/p_{rr} = p_{lm}/p_{lr} \times p_{lr}/p_{rr}$, but we need to be careful if $p_{lr} = 0$. So we will first show that $p_{lr} = 0$ implies $p_{lm} = 0$ for all m. The proof of this is by contradiction. Assume $p_{lm} > 0$ for some m, and note

$$
\begin{aligned}
p_{lm} > 0 &\Rightarrow P[T_1 \geq k_1, Z_1 = l, T_2 = k_2 - 1, Z_2 \geq k_2 - 1] > 0 \\
&\Rightarrow P[Y_1 = l, \delta_1 = 0, Y_2 = k_2 - 1, \delta_2 = 1] > 0 \\
&\Rightarrow P[Y_1 = l, \delta_1 = 0, T_2 = r, Z_2 = r] > 0 \\
&\Rightarrow P[T_1 = r, Z_1 = l, T_2 = r, Z_2 = r] > 0 \\
&\Rightarrow P[T_1 = k_1 - 1, Z_1 = l, T_2 = r, Z_2 = r] > 0.
\end{aligned}
$$

This is a contradiction since

$$
\frac{p_{lr}}{P[T_1 = k_1 - 1, Z_1 = l, T_2 = r, Z_2 = r]} = \frac{p_{lm}}{P[T_1 = k_1 - 1, Z_1 = l, T_2 = r, Z_2 = m]}
$$

and the left hand side is zero, but the right hand side is positive. So we now have a formula which expresses p_{lm} in terms of p_{rr} for all l and m so we can figure out each of the p_{lm} since their sum is $P[Y_1 = k_1 - 1, \delta_1 = 1, Y_2 = k_2 - 1, \delta_2 = 1]$. □

This shows that if we are willing to restrict attention to discrete probability measures which satisfy (5), the independence condition (2) is sufficient to ensure identifiability. The assumption of full support, or (5), is a strong one, but one which in practice would generally be much more acceptable than assuming mutual independence of $\vec{T}$ and $\vec{Z}$. A natural estimator to use in the setting of this section is one that is internally consistent. This estimator is not uniquely defined until F_w^n satisfies the assumption (5), but is uniquely defined for all n larger than $N =$ the smallest n for which (5) is satisfied. There is no closed form for the estimator, but it may be obtained by the formulas in this section. Note that consistency of the estimator for all discrete distributions F_x that have $\tilde{H}(F_x)$ satisfy (5) is automatic from the 1-1 nature of $H^{*^{-1}}$ for $n \geq N$, and the consistency of F_w^n.

References

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.

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. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics, IV.* New York, Prentice-Hall 831-853.

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

KAPLAN, E.L., and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457-481.

LANGBERG, N., 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. (1989). On negative mass assigned by the bivariate Kaplan-Meier estimator. *Preprint.* University of Minnesota.

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.

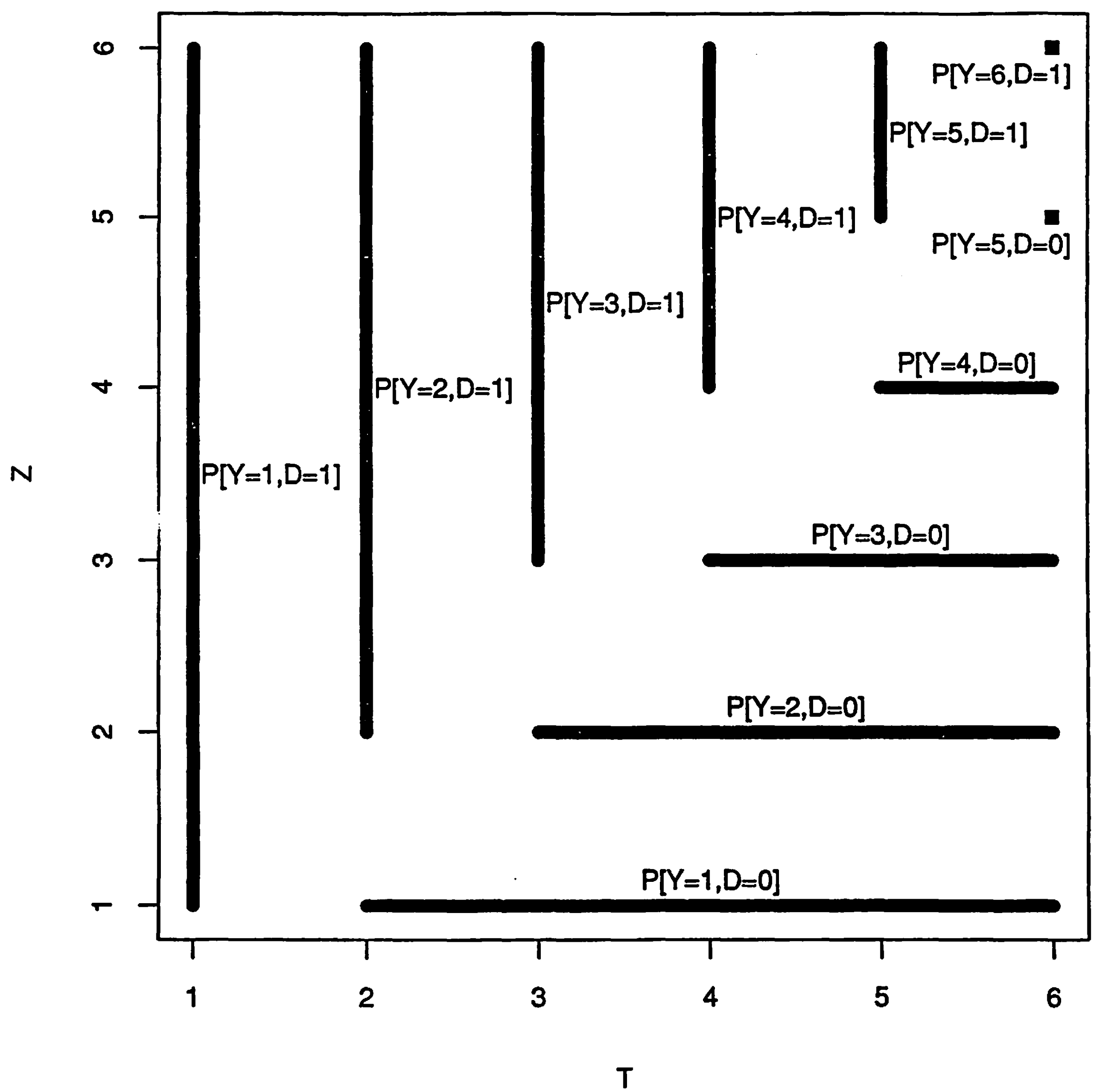

Figure 1: Possible incomplete observations are indicated as sets in the (T, Z)-plane for discrete right censored data.

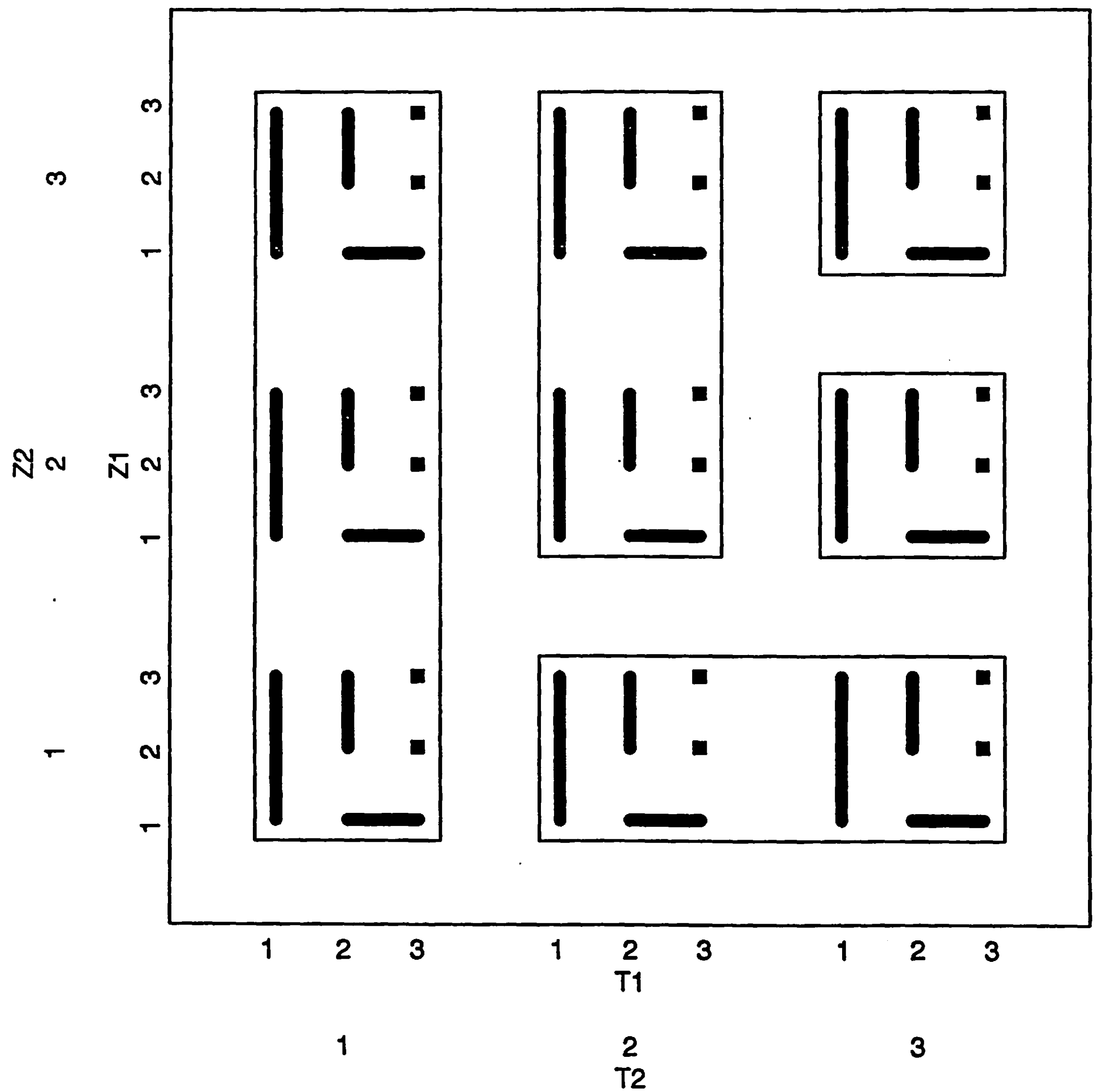

Figure 2: Possible incomplete observations are indicated as sets in (T_1, T_2, Z_1, Z_2) space for discrete bivariate right censored data.