The Annals of Statistics
1991, Vol. 19, No. 1, 443–453

ON NEGATIVE MASS ASSIGNED BY THE BIVARIATE KAPLAN–MEIER ESTIMATOR[1]

BY RONALD C. PRUITT

University of Minnesota

Conditions under which the bivariate Kaplan–Meier estimate of Dabrowska is not a proper survival function are given. All points assigned negative mass are identified under the assumption that the observations follow an absolutely continuous distribution. The number of points assigned negative mass increases as n^2 and the total amount of negative mass does not disappear as $n \to \infty$, where n is the sample size. A simulation study is reported which shows that large amounts of negative mass are assigned by the estimator, amounts ranging from 0.3 to 0.6 for a sample of size 50 over a variety of parameters for bivariate exponential distributions.

1. Introduction. Dabrowska (1988) introduced a multivariate survival curve estimate. In her paper she points out that her estimate may fail to be monotone and hence may not be a survival function. This paper describes when and how Dabrowska's bivariate estimate is not a survival function.

Throughout we follow the notation of Dabrowska (1988). We wish to infer about a bivariate distribution $\mathbf{T} = (T_1, T_2)$ subject to censoring. Assume $\mathbf{T}$ and the censoring variable $\mathbf{Z} = (Z_1, Z_2)$ are defined on a common probability space $(\Omega, \mathscr{F}, \mathscr{P})$ 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 $\mathbf{Y} = (Y_1, Y_2)$ and $\boldsymbol{\delta} = (\delta_1, \delta_2)$, 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 $(\mathbf{Y}_i, \boldsymbol{\delta}_i)$, $i = 1, \dots, n$, which consists of independent, identically distributed copies of $(\mathbf{Y}, \boldsymbol{\delta})$. Let

$$\hat{H}(s,t) = n^{-1} \sum 1[Y_{1i} > s, Y_{2i} > t],$$

$$\hat{K}_1(s,t) = n^{-1} \sum 1[Y_{1i} > s, Y_{2i} > t, \delta_{1i} = 1, \delta_{2i} = 1],$$

$$\hat{K}_2(s,t) = n^{-1} \sum 1[Y_{1i} > s, Y_{2i} > t, \delta_{1i} = 1],$$

$$\hat{K}_3(s,t) = n^{-1} \sum 1[Y_{1i} > s, Y_{2i} > t, \delta_{2i} = 1].$$

These functions can be used to estimate the bivariate cumulative hazard

Received April 1989; revised November 1989.
[1]Research partially supported by National Institutes of General Medical Sciences Grant GM-42921-01.
AMS 1980 *subject classifications*. Primary 62H10; secondary 62G05.
Key words and phrases. Censored data, improper survival function.

function by

$$\hat{\Lambda}_{11}(s,t) = \int_0^s \int_0^t \hat{K}_1(du, dv)/\hat{H}(u-, v-),$$

$$\hat{\Lambda}_{10}(s,t) = -\int_0^s \hat{K}_2(du, t)/\hat{H}(u-, t),$$

$$\hat{\Lambda}_{01}(s,t) = -\int_0^t \hat{K}_3(s, dv)/\hat{H}(s, v-).$$

With $f(\Delta x) = f(x) - f(x-)$, define

$$\hat{L}(\Delta u, \Delta v) = \frac{\hat{\Lambda}_{10}(\Delta u, v-)\hat{\Lambda}_{01}(u-, \Delta v) - \hat{\Lambda}_{11}(\Delta u, \Delta v)}{\{1 - \hat{\Lambda}_{10}(\Delta u, v-)\}\{1 - \hat{\Lambda}_{01}(u-, \Delta v)\}}$$

if the denominator of the right-hand side is nonzero, and otherwise let $\hat{L}(\Delta u, \Delta v) = 0$. Dabrowska's estimate is

$$\hat{F}(s,t) = \hat{F}(s,0)\hat{F}(0,t) \prod_{\substack{0<u\le s\\0<v\le t}} \{1 - \hat{L}(\Delta u, \Delta v)\}, \tag{1}$$

where $\hat{F}(s,0)$ and $\hat{F}(0,t)$ are the marginal Kaplan–Meier estimates.

2. Points assigned negative mass. Since the estimate has Kaplan–Meier marginals, it is a survival function if and only if it assigns positive mass to all rectangles. Further, in light of the fact that it is a discrete measure (see Theorem 1), it is a survival function if and only if it assigns negative mass to no points. We only consider the case where $\mathbf{Y}$ is absolutely continuous, since if it is discrete eventually all points at which the survival curve can be estimated will have uncensored observations.

THEOREM 1. *Assume the distribution of* $\mathbf{Y}$ *is absolutely continuous. With probability* 1, *Dabrowska's estimate is a discrete measure and assigns negative mass in accordance with Lemmas* 3–7.

PROOF. Restrict attention to the case when $Y_{i1}, \dots, Y_{in}$ are all distinct for $i = 1,2$. We will follow the convention that if the largest observation is censored, the univariate Kaplan–Meier estimate places the indeterminate mass at the censored value of the largest observation. Define M_i by $Y_{iM_i} = \max\{Y_{i1}, \dots, Y_{in}\}$ for $i = 1,2$. With this convention it is easy to show Dabrowska's estimate does not depend on the value of δ_{iM_i} for $i = 1,2$. We will assume $\delta_{iM_i} = 1$ for $i = 1,2$. First note that mass is concentrated on the set of points $S = \{(y_1, y_2)\colon y_1 = Y_{1i},\ y_2 = Y_{2j},\ \delta_{1i} = 1$ and $\delta_{2j} = 1$ for some $1 \le i, j \le n\}$. The mass assigned to a point (s,t) may be written

$$\hat{F}(\Delta s, \Delta t) = R_0(s,t)\{R_1(s,t)R_2(s,t) - R_3(s,t)\}, \tag{2}$$

where

$$R_0(s,t) = \prod_{\substack{0<u<s \\ 0<v<t}} \{1 - \hat{L}(\Delta u, \Delta v)\},$$

$$R_1(s,t) = \hat{F}(s-,0) - \hat{F}(s,0) \prod_{0<v<t} \{1 - \hat{L}(\Delta s, \Delta v)\},$$

$$R_2(s,t) = \hat{F}(0,t-) - \hat{F}(0,t) \prod_{0<u<s} \{1 - \hat{L}(\Delta u, \Delta t)\},$$

and

$$R_3(s,t) = \hat{L}(\Delta s, \Delta t)\hat{F}(s,0)\hat{F}(0,t) \prod_{0<u<s} \{1 - \hat{L}(\Delta u, \Delta t)\} \times \prod_{0<v<t} \{1 - \hat{L}(\Delta s, \Delta v)\}.$$

Note that the set of points where $\hat{L}(\Delta u, \Delta v)$ is nonzero is contained in S, and the marginal Kaplan–Meier estimates only both change value on points in S. There are seventeen possible cases, the case when $i = j$ and the sixteen cases indicated in Figure 1. By symmetry only plots in Figure 1 on or beneath the diagonal need to be considered; for example, plots 4 and 13 only differ in the labeling of the variables. These fall into five cases which are covered by Lemmas 3–7. □

We split the types of points assigned mass into five types for convenience of exposition: The groupings are arbitrary and based on similarities which simplify the proofs. Points of types I–III which receive negative mass share the characteristic that once they appear in a sample, the addition of more sample observations can never make them disappear. This can be seen from Lemmas 3–5. The amount of negative mass assigned will become smaller, but the fact that negative mass is assigned to the point will never change. These points are also very prevalent (see Theorem 8), constituting a fraction of all points receiving mass. Type IV points are relatively rare; they are only assigned negative mass if the coordinate values of the points involved are among the largest in the sample. Type V points are never assigned negative mass, and these include uncensored observation points.

Before stating and proving Lemmas 3–7 we give an auxiliary lemma which will be used heavily. This lemma delineates precisely what determines the sign of $R_2(s,t)$ [or $R_1(s,t)$]. The centrality of these conditions is apparent from (2). Let $n_{s,t} = n\hat{H}(s-,t-)$ be the number of observations in $[s,\infty) \times [t,\infty)$.

LEMMA 2. *Fix k, with $1 \le k \le n$ and $\delta_{2k} = 1$. Let $s = Y_{1k}$ and $t = Y_{2k}$. Let $C = \{(\mathbf{Y}_m, \boldsymbol{\delta}_m)\colon Y_{1m} < s,\ Y_{2m} > t,\ \delta_{1m} = 0\}$. Then $R_2(x,t) \le 0$ if and only if $\delta_{1k} = 1$, $n_{s,t} > 1$ and $x > s$, with equality if and only if C is empty.*

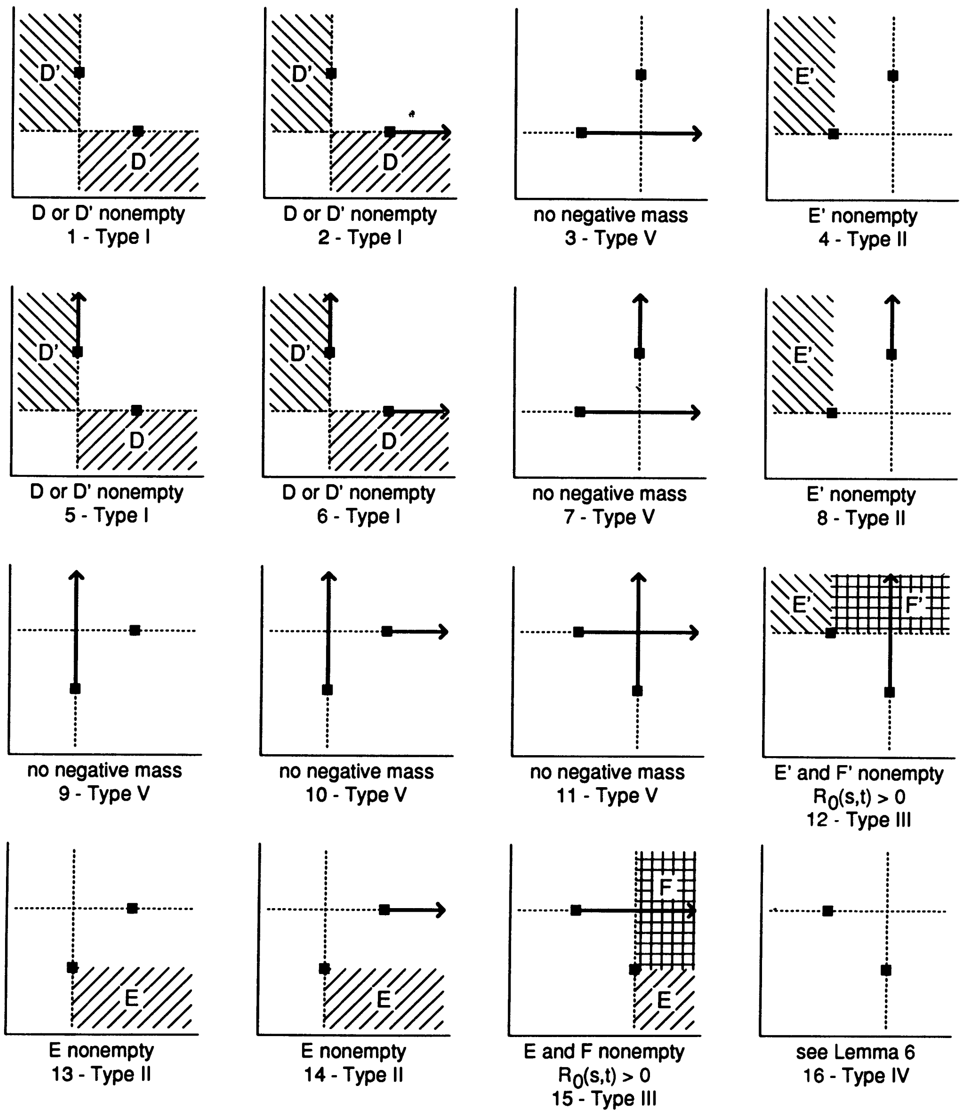

FIG. 1. *Graphical representation of configurations of observations described in Lemmas 3–7. Two observation points are indicated in bold on each of the above plots; singly censored observations are indicated by an arrow. Points of possible negative mass are at the intersection of the dashed lines, say (s, t), and conditions for negative mass assignment are given. The sets D and E only contain points with $\delta_2 = 0$, and the sets D' and E' only contain points with $\delta_1 = 0$. The sets F and F' have no censoring restrictions.*

PROOF. Since the observed values are all distinct,

$$\hat{\Lambda}_{01}(x-, \Delta t) = \begin{cases} (n_{x,t})^{-1}, & \text{for } x \le s, \\ 0, & \text{for } x > s, \end{cases}$$

and

$$\hat{\Lambda}_{11}(\Delta x, \Delta t) = \begin{cases} (n_{x,t})^{-1}, & \text{for } x = s, \delta_{1k} = 1, \\ 0, & \text{otherwise}. \end{cases}$$

Combining this with a similar equation for $\hat{\Lambda}_{10}$,

$$(3)\quad 1 - \hat{L}(\Delta x, \Delta t) = \begin{cases} \dfrac{n_{x,t}(n_{x,t} - 2)}{(n_{x,t} - 1)^2}, & \text{if } x < s \text{ and } \hat{K}_2(\Delta x, t) < 0, \\ \dfrac{n_{x,t}}{(n_{x,t} - 1)}, & \text{if } x = s, \delta_{1k} = 1 \text{ and } n_{x,t} > 1, \\ 1, & \text{if } x = s, \delta_{1k} = 1 \text{ and } n_{x,t} = 1, \\ 1, & \text{if } x > s \text{ or } \hat{K}_2(\Delta x, t) = 0. \end{cases}$$

Note that $\hat{K}_2(\Delta x, t) < 0$ precisely if there exists an observation with first coordinate uncensored at x and second coordinate greater than or equal to t, so that $\hat{K}_2(\Delta s, t) < 0$ if and only if $\delta_{1k} = 1$. Now

$$(4)\quad \begin{aligned} \prod_{0<u<s} \{1 - \hat{L}(\Delta u, \Delta t)\} &= \prod_{\substack{0<u<s \\ \hat{K}_2(\Delta u, t)<0}} \left\{ \frac{n_{u,t}(n_{u,t} - 2)}{(n_{u,t} - 1)^2} \right\} \\ &\ge \frac{n_{0+,t}}{(n_{0+,t} - 1)} \frac{(n_{s,t} - 1)}{n_{s,t}}, \end{aligned}$$

with equality if and only if C is empty. If C is empty, the product is a telescoping product and if C is nonempty terms less than one are left out of the product. If $x > s$, $\delta_{1k} = 1$ and $n_{s,t} > 1$,

$$\prod_{0<u<x} \{1 - \hat{L}(\Delta u, \Delta t)\} = \prod_{0<u<s} \{1 - \hat{L}(\Delta u, \Delta t)\} \frac{n_{s,t}}{n_{s,t} - 1} \ge \frac{n_{0+,t}}{(n_{0+,t} - 1)}.$$

This shows the sufficiency after noting that $\hat{F}(0, t-)/\hat{F}(0, t) = n_{0+,t}/(n_{0+,t} - 1)$. For the necessity, $R_2(x, t) \ge \hat{F}(0, t-) - \hat{F}(0, t) > 0$ whenever $x \le s$, $\delta_{1k} = 0$ or $n_{s,t} \le 1$. □

For any set $A \in \mathbb{R}_+^2 \times \{0, 1\}^2$, let $A' = \{(y_1, y_2, \delta_1, \delta_2): (y_2, y_1, \delta_2, \delta_1) \in A\}$.

LEMMA 3 (Type I). *Assume* $Y_{1i} < Y_{1j}$, $Y_{2i} > Y_{2j}$, $\delta_{1i} = 1$ *and* $\delta_{2j} = 1$. *Then negative mass is assigned to the point* (Y_{1i}, Y_{2j}) *if and only if* D *or* D' *is nonempty, where* $D = \{(\mathbf{Y}_k, \boldsymbol{\delta}_k): Y_{1k} > Y_{1i}, Y_{2k} < Y_{2j}$ *and* $\delta_{2k} = 0$ *for some* $k = 1, \dots, n\}$.

PROOF. Let $s = Y_{1i}$, $t = Y_{2j}$, and note $R_0(s,t) > 0$. From (4),

$$\prod_{0<u<s} \{1 - \hat{L}(\Delta u, \Delta t)\} \geq \frac{n_{0+,t}}{(n_{0+,t} - 1)} \frac{(n_{s,t} - 1)}{n_{s,t}},$$

with equality if and only if D' is empty. Applying this to (2) gives

$$\frac{\hat{F}(\Delta s, \Delta t)}{R_0(s,t)} \leq \hat{F}(s,0) \frac{n_{s,0+}}{(n_{s,0+} - 1)} \frac{1}{n_{s,t}} \hat{F}(0,t) \frac{n_{0+,t}}{(n_{0+,t} - 1)} \frac{1}{n_{s,t}}$$

$$- \frac{1}{(n_{s,t} - 1)^2} \hat{F}(s,0) \hat{F}(0,t) \frac{n_{s,0+}}{(n_{s,0+} - 1)} \frac{(n_{s,t} - 1)}{n_{s,t}}$$

$$\times \frac{n_{0+,t}}{(n_{0+,t} - 1)} \frac{(n_{s,t} - 1)}{n_{s,t}} = 0,$$

with equality if and only if D and D' are both empty. □

LEMMA 4 (Type II). *Assume $Y_{1i} < Y_{1j}$, $Y_{2i} < Y_{2j}$, $\delta_{1i} = 1$, $\delta_{2j} = 1$ and $\delta_{2i} = 1$. Then negative mass is assigned to the point (Y_{1i}, Y_{2j}) if and only if the set $E = \{(\mathbf{Y}_k, \boldsymbol{\delta}_k)\colon Y_{1k} > Y_{1i},\ Y_{2k} < Y_{2i}$ and $\delta_{2k} = 0$ for some $k = 1, \dots, n\}$ is nonempty.*

PROOF. Note that $R_3(Y_{1i}, Y_{2j}) = 0$, and by Lemma 2, $R_2(Y_{1i}, Y_{2j}) > 0$ and $R_1(Y_{1i}, Y_{2j}) < 0$ if and only if E is nonempty. Also note from (3) that $1 - \hat{L}(\Delta u, \Delta v) = 0$ only if $n_{u+,v+} = 0$ and hence $R_0(Y_{1i}, Y_{2j}) > 0$ since $n_{u+,v+} \geq 1$ for $0 \leq u < Y_{1i}$ and $0 \leq v < Y_{2j}$. □

LEMMA 5 (Type III). *Assume $Y_{1i} > Y_{1j}$, $Y_{2i} < Y_{2j}$, $\delta_{1i} = 1$, $\delta_{2j} = 1$, $\delta_{1j} = 0$ and $\delta_{2i} = 1$. Then negative mass is assigned to the point (Y_{1i}, Y_{2j}) if and only if the set E of Lemma 4 is nonempty, the set $F = \{(\mathbf{Y}_k, \boldsymbol{\delta}_k)\colon Y_{1k} > Y_{1i}$ and $Y_{2k} > Y_{2i}$ for some $k = 1, \dots, n\}$ is nonempty, and $R_0(Y_{1i}, Y_{2j}) > 0$.*

PROOF. Note that $R_3(Y_{1i}, Y_{2j}) = 0$, and by Lemma 2, $R_2(Y_{1i}, Y_{2j}) > 0$ and $R_1(Y_{1i}, Y_{2j}) < 0$ if and only if E and F are each nonempty. □

LEMMA 6 (Type IV). *Assume $Y_{1i} > Y_{1j}$, $Y_{2i} < Y_{2j}$ and $\delta_{1i} = \delta_{1j} = \delta_{2i} = \delta_{2j} = 1$. Then (Y_{1i}, Y_{2j}) is assigned negative mass if and only if [the set $G = \{(\mathbf{Y}_k, \boldsymbol{\delta}_k)\colon Y_{1k} > Y_{1j}$ and $Y_{2k} > Y_{2j}$ for some $k = 1, \dots, n\}$ is empty, the sets E and F of Lemma 5 are each nonempty and $R_0(Y_{1i}, Y_{2j}) > 0$] or [the set G' is empty, the sets E' and F' are each nonempty and $R_0(Y_{1i}, Y_{2j}) > 0$].*

PROOF. Note $R_3(Y_{1i}, Y_{2j}) = 0$, so that negative mass is assigned if and only if [$R_0(Y_{1i}, Y_{2j}) > 0$, $R_1(Y_{1i}, Y_{2j}) < 0$ and $R_2(Y_{1i}, Y_{2j}) > 0$] or [$R_0(Y_{1i}, Y_{2j}) > 0$, $R_1(Y_{1i}, Y_{2j}) > 0$ and $R_2(Y_{1i}, Y_{2j}) < 0$]. By Lemma 2, $R_1(Y_{1i}, Y_{2j}) < 0$ and $R_2(Y_{1i}, Y_{2j}) > 0$ if and only if $G = \varnothing$, $E \neq \varnothing$ and $F \neq \varnothing$. □

LEMMA 7 (Type V). *Negative mass is not assigned to any points not covered by Lemmas* 3–6.

PROOF. For the case of mass assigned to an uncensored point, we have $R_1(Y_{1i}, Y_{2j}) > 0$, $R_2(Y_{1i}, Y_{2j}) > 0$ and $R_3(Y_{1i}, Y_{2j}) \le 0$. For the other cases, $R_3(Y_{1i}, Y_{2j}) = 0$, $R_2(Y_{1i}, Y_{2j}) > 0$, $R_1(Y_{1i}, Y_{2j}) > 0$ and $R_0(Y_{1i}, Y_{2j}) \ge 0$. □

3. Number and magnitude of negative mass points. In this section, we provide a rough lower bound on the expected number of points assigned negative mass which grows as n^2, and then show that the total amount of negative mass remains bounded away from 0. This is followed by discussion of a simulation study.

3.1. *Number of negative mass points.* Let T_n be the total number of points assigned negative mass from the sample $\mathbf{Y}_1, \dots, \mathbf{Y}_n$.

THEOREM 8. *Assume the distributions of* $\mathbf{Y}$, $\mathbf{T}$ *and* $\mathbf{Z}$ *are absolutely continuous. Also assume that the supports of* $\mathbf{T}$ *and* $\mathbf{Z}$ *have nonempty intersection which contains an open set. Then* $E(T_n) = O(n^2)$.

PROOF. Define the point $p_{ij} = (Y_{1i}, Y_{2j})$ and let $T_n(i,j)$ be the indicator that p_{ij} is assigned negative mass. Let B be the event that points 1 and 2 form a type I pair; that is, $B = \{Y_{11} < Y_{12}, Y_{21} > Y_{22}, \delta_{11} = \delta_{22} = 1\}$. Recall the events D and D' from Lemma 3. Then

$$\begin{aligned} E(T_n) &= \sum \sum E(T_n(i,j)) \\ &= n(n-1)P(T_n(1,2) = 1) \\ &\ge n(n-1)P(B)P(T_n(1,2) = 1|B) \\ &= n(n-1)P(B)[1 - P(D = \varnothing, D' = \varnothing|B)]. \end{aligned}$$

The conditions of the theorem suffice for $P(D = \varnothing, D' = \varnothing|B)$ to be less than 1, and $P(B)$ to be greater than 0. □

Note that this same method of attack works for points of type II or III as well to show there are $O(n^2)$ of these type of points. We can also evaluate these bounds for special cases. The sets D and D' depend on n, denote these events as D_n and D_n' if the sample size is n. Then

$$\begin{aligned} &P\big(D_n = \varnothing, D_n' = \varnothing|B\big) \\ &\quad = \int \big\{P\big(D_3 = \varnothing, D_3' = \varnothing|B, m_1, m_2\big)\big\}^{n-2} \, dF(m_1, m_2), \end{aligned}$$

where $W_i = (Y_{i1} \wedge Y_{i2})$ for $i = 1, 2$ and F is the distribution function of $(W_1, W_2)|B$.

In particular if the supports of $\mathbf{T}$ and $\mathbf{Z}$ are both all of $\mathbb{R}_+^2$, this probability decreases to 0. Of all the points of type I (II or III) in such a sample, a vanishing fraction does not receive negative mass as the sample size increases. In such cases, the percentage of points receiving negative mass in large samples will be determined by the relative frequency of type I–III points.

3.2. *Magnitude of negative mass points.* This assignment of negative mass might be acceptable in practice, although philosophically troubling, if the amount of negative mass assigned was small. This is not the case, however. The amount of negative mass assigned to each point receiving negative mass does go to 0 which can be seen since the estimator is consistent. Here we develop a bound on the rate at which this occurs. Let U_n be the total amount of negative mass assigned by Dabrowska's estimator.

THEOREM 9. *Assume the distributions of* **Y**, **T** *and* **Z** *are absolutely continuous. Also assume that the supports of* **T** *and* **Z** *have nonempty intersection which contains an open set. Then* $\liminf_{n\to\infty} U_n > 0$ *almost surely.*

PROOF. Suppose (by renumbering if necessary) that $\mathbf{Y}_1$ and $\mathbf{Y}_2$ occur in the configuration given in plot 13, and let $s = Y_{1i}$ and $t = Y_{2j}$. Let $K_n = \#E$, where E is given in Lemma 4, and let $M_n = n_{s,t} - 1$. Note that (K_n, M_n) has a multinomial distribution with parameters $n-2$, p_1 and p_2, where $p_1 = P[K_3 = 1]$ and $p_2 = P[M_3 = 1]$. Assume $p_1 > 0$ and $p_2 > 0$, and note that the probability of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ being in this arrangement is nonzero by hypothesis. The expected number of pairs of this type in a sample of size n is $O(n^2)$. Consider n so large that K_n and M_n are both nonzero. Note $R_3(s,t) = 0$. We now develop bounds on R_0, R_1 and R_2. Note

$$\prod_{0<v<t} \{1 - \hat{L}(\Delta s, \Delta v)\} \geq \frac{n_{s,0+} - K_n}{n_{s,0+} - K_n - 1},$$

which may be seen by noting that if K_n is 0 the product is a telescoping product, and for every element in E one term gets left out, so to minimize the product we leave out the largest terms. Also note that $\prod_{0<u<s}\{1 - \hat{L}(\Delta u, \Delta t)\} \leq 1$, so that

$$\frac{R_2(s,t)}{\hat{F}(0,t)} \geq \frac{n_{0+,t}}{n_{0+,t} - 1} - 1 > 0$$

and

$$\frac{R_1(s,t)}{\hat{F}(s,0)} \leq \frac{-K_n}{(n_{s,0+} - 1)(n_{s,0+} - K_n - 1)} < 0.$$

Since $1 - \hat{L}(\Delta u, \Delta v) \geq 1 - M_n^{-2}$ for any u, v in $[0,s) \times [0,t)$, we have $R_0(s,t) \geq (1 - M_n^{-2})^{n^2}$. This yields

$$\begin{aligned} n^2 \hat{F}(\Delta s, \Delta t) &\leq \frac{-K_n n^2 \hat{F}(s,0)\hat{F}(0,t)\left(1 - M_n^{-2}\right)^{n^2}}{(n_{s,0+} - 1)(n_{0+,t} - 1)(n_{s,0+} - K_n - 1)} \\ &\leq \frac{-K_n}{n}\left(1 - M_n^{-2}\right)^{n^2}. \end{aligned}$$

The right-hand side of this last inequality converges almost surely to $-p_1 \exp(-p_2^{-2}) < 0$. □

Similar bounds can be obtained for the other points of type I and II, but the algebra is messier.

The bound for the magnitude of negative mass depends on the number of censored observations in a certain region, rather than there just existing such points as was the case for the number of negative mass points. For this reason, we might expect the amount of negative mass to be more variable and to depend on different factors in the distribution. These general observations are supported by the simulation results, but further analysis seems to be difficult.

3.3. *Simulation results.* Data for **T** and **Z** were generated from the bivariate exponential model of Marshall and Olkin (1967): $F(s,t) = \exp\{-\lambda_t s - \lambda_t t - \lambda_T(s \vee t)\}$ for $s, t \geq 0$ and $\lambda_t > 0$, $\lambda_T \geq 0$. The distribution for **Z** has parameters λ_z and λ_Z. Only models symmetric in s and t were considered. Simulations were performed for all combinations of $\lambda_t = 1$; $\lambda_T = 0, 1$; $\lambda_z = 0.5, 1, 2$; $\lambda_Z = 0, 1$. The independent, identically distributed case and the cases which gave extreme readings in terms of numbers and mass of negative mass points are summarized in Table 1. The percentage of grid points assigned negative mass refers to the percentage of the possible points receiving mass

TABLE 1
Simulation results for Dabrowska's estimator for the bivariate exponential model $F(s,t) = \exp\{-\lambda_t s - \lambda_t t - \lambda_T(s \vee t)\}$ *and* $G(s,t) = \exp\{-\lambda_z s - \lambda_z t - \lambda_Z(s \vee t)\}$*

Parameters				Sample size	Percentage of grid points assigned negative mass			Total amount of negative mass assigned		
λ_t	λ_T	λ_z	λ_Z	n	Average	Minimum	Maximum	Average	Minimum	Maximum
1	0	1	0	10	0.29	0	0.73	0.22	0	1.21
				25	0.46	0.14	0.71	0.44	0.04	2.09
				100	0.57	0.41	0.71	0.74	0.34	2.46
				400	0.61	0.52	0.67	0.97	0.52	1.67
1	0	2	1	10	0.19	0	0.62	0.11	0	1.03
				25	0.35	0	0.69	0.26	0	1.62
				100	0.47	0.20	0.68	0.52	0.03	2.04
				400	0.50	0.38	0.62	0.81	0.24	2.27
1	1	0.5	0	10	0.24	0	0.75	0.16	0	0.91
				25	0.45	0.07	0.82	0.25	0.01	0.91
				100	0.69	0.43	0.83	0.36	0.14	0.84
				400	0.76	0.68	0.81	0.41	0.25	0.69
1	1	2	0	10	0.30	0	0.76	0.22	0	1.07
				25	0.50	0.06	0.81	0.41	0.02	1.48
				100	0.64	0.39	0.79	0.69	0.21	1.77
				400	0.67	0.59	0.74	0.91	0.51	1.78

*The trial size is 1000. The estimated standard errors for the averages are all under 0.01.

and not to the percentage of n^2. This tends to remove some of the noise of having a large or small number of censored observations.

In general, observation pairs consisting of a small censored variable and a large other variable cause negative mass points as can be seen in Lemmas 3–5. For this model such conditions tend to occur if the T variables are correlated and the Z variables are uncorrelated, for example, $\lambda_t = \lambda_T = 1$ and $\lambda_Z = 0$. A smaller percentage of points are assigned negative mass if the T variables are uncorrelated and the Z variables are correlated. The total amount of negative mass assigned and the percentage of points receiving negative mass are not strongly positively correlated; in fact, the parameters which gave the highest percentage of negative mass points gave the lowest average total negative mass. The independent case gave the most negative mass.

4. Conclusions. Why does Dabrowska's estimator assign negative mass, and should we be worried about it? The same questions can be asked of the estimator proposed by Langberg and Shaked (1982), which also assigns negative mass. We provide an answer to the first question by analogy with the univariate problem. The assignment of negative mass is not a problem if the estimate at a single point is all that is desired, since it can be checked that the estimator assigns positive mass to all upper orthants. It may also be feasible to get an estimator which is a probability measure by smoothing, either into fixed bins making the problem discrete, or by choosing a smoothing window around each point. Our main concern here is why such a sensible generalization of the univariate problem can give an estimator which essentially always assigns negative mass.

In one dimension, assuming T and Z independent in the censored data problem makes the distribution of T identifiable. The assumption that T and Z are independent can be slightly weakened [Tsai (1986)], but not much. There is nearly a one-to-one correspondence between the (Y, δ) and (T, Z) distributions with T and Z independent. One way to view the Kaplan–Meier estimators for T and Z is as distributions which are consistent with the observed data being the entire population.

The situation for bivariate data is quite different. Consider a simple example. Suppose $Y_1 = (2,2)$ and $Y_2 = (3,1)$, and both points are uncensored ($\delta_{ij} = 1,\ 1 \le i, j \le 2$). Suppose $Y_3 = (1, t)$ with $\delta_{13} = \delta_{23} = 0$. First assume $t < 1$. The given data are consistent with $\mathbf{T}$ and $\mathbf{Z}$ being independent in the sense of there existing distributions for $\mathbf{T}$ and $\mathbf{Z}$ which combine to give the empirical distribution for the observed data. If $\mathbf{Z}$ has mass $1/3$ at $(1, t)$ and $2/3$ at $(3,3)$, and $\mathbf{T}$ has mass $1/2$ at $(2,2)$ and $1/2$ at $(3,1)$; then the observed data are the population values for $\mathbf{Y}$ and $\boldsymbol{\delta}$.

But if $t > 1$ this is no longer true. Now the observation at $(1, t)$ makes it appear that $\mathbf{T}$ and $\mathbf{Z}$ are not independent. If $\mathbf{T}$ and $\mathbf{Z}$ are independent, $\mathbf{T}$ has mass at $(3,1)$, and $\mathbf{Z}$ has mass at $(1, t)$, so that eventually there will be observations at $(1,1)$ with the first coordinate censored. Since no such observations exist in this sample, a tension is created between the independence assumption and the treatment of the sample data via empirical survival and subsurvival curves. Since the estimation procedure is making use of both the independence and the empirical mass assignment this tension causes difficulty. In the particular treatments of Dabrowska or Langberg and Shaked, it causes negative mass assignment.

There are two avenues of approach to deal with the problem just outlined. The first, which apparently has not been attempted, is to relax the assumption that **T** is independent of **Z** and find a weaker assumption which still allows the distribution of **T** to be identified. This is pursued in Pruitt (1989) for discrete distributions with full support where it is shown that assuming T_i independent of Z_i conditional on T_{3-i} and Z_{3-i} for $i = 1, 2$ is enough to ensure identifiability. In this setting, estimators may be found which are consistent with the observed data being the entire population. The other option is to find a distribution for **T** which is not consistent with the observed data being the entire population. There is nothing wrong with this approach, but properties such as being a proper survival function which are guaranteed in the former approach become theorems to be proved.

Acknowledgment. The referee's comments were useful in improving the presentation.

REFERENCES

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

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

MARSHALL, A. W. and OLKIN, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62** 30–44.

PRUITT, R. (1989). Identifiability of discrete bivariate survival curves from censored data. Technical Report 535, Univ. 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.

DEPARTMENT OF THEORETICAL STATISTICS
UNIVERSITY OF MINNESOTA
206 CHURCH STREET S.E.
MINNEAPOLIS, MINNESOTA 55455