NONPARAMETRIC ESTIMATION OF
A BIVARIATE SURVIVAL FUNCTION
IN THE PRESENCE OF CENSORING

Wei-Yann Tsai
Sue Leurgans
John Crowley

DEPARTMENT OF HUMAN ONCOLOGY
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN—MADISON
Nonparametric Estimation of a Bivariate Survival Function in the Presence of Censoring*

by

Wei-Yann Tsai, Brookhaven National Laboratory
Sue Leurgans, University of Wisconsin - Madison
John Crowley, The Fred Hutchinson Cancer Research Center and University of Washington

Abstract

A new family of estimators of a bivariate survival function based on censored vectors is obtained from a decomposition of the bivariate survival function. These estimators are uniformly consistent under bivariate censoring and are self-consistent under univariate censoring. An example is included.

Key Words: Bivariate survival function, self consistency, univariate and bivariate censored data.

*Research was supported in part by National Institute of Health Grants 5 R01 CA 18332 and 2 R01 GM28314 and in part by funds provided by the International Cancer Research Data Bank Program of the National Cancer Institute, NIH, under contract No. NO1-CO-65341 (International Cancer Research Technology Transfer ICETT) and by the International Union Against Cancer. This manuscript has been authored under contract DE-AC02-76CH00016 with the U.S. Department of Energy. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.
I. Introduction

Methods for analyzing univariate censored data have been studied by many people over the last few decades. Relatively little research has been devoted to the analysis of bivariate observations in the presence of censoring. In some studies, two times are observed because experimental units consist of pairs of components, and the lifetime of each component is recorded. Censoring occurs when the experimental unit is removed from the study before both components have been observed to fail. Examples include twin studies and matched pair studies. In other studies, two failure times are recorded for each individual or piece of equipment. Thus in studies of chronic diseases, both recurrence times and death times are recorded, and their joint distribution needs to be estimated.

In this paper, we discuss nonparametric estimation of the bivariate survival function in the presence of censoring. In Section 3, we present a new class of estimators of the bivariate survival function. These estimators are based on a decomposition given in Section 2 of the bivariate survival function in terms of estimable functions. Sections 3 and 4 discuss properties of the estimators and Section 5 gives an example. The rest of this section summarizes previous work on this problem.

We first review estimation of a univariate survival distribution. In this case, let $T_i^0$, $i=1,...,n$ be $n$ independent and identically distributed (iid) lifetimes with continuous survival function $S^0(t) = P(T_i^0 > t)$. Let $C_i$, $i=1,...,n$ be an independent sample of $n$ censoring variables with survival function $H(t)$. It is not possible to observe both $T_i^0$ and $C_i$. Let $X \wedge Y$ denote $\min(X,Y)$ and $[A]$ denote the indicator of the event $A$. We observe $T_i$ and $D_i$,
where $T_i = T_i^0 \land C_i$ and $D_i = [T_i^0 < C_i]$. Kaplan and Meier (1958) suggested

$$\hat{S}^0(t) = \prod_{T_i < t} \left(1 - \frac{D_i}{\sum_j [T_j > T_i]} \right)$$

as an estimator of $S^0$. They also showed that $\hat{S}^0$ is a nonparametric maximum likelihood estimator in the sense that $\hat{S}^0$ formally maximizes an expression that would be a likelihood function if a parameter were being estimated. Johansen (1978) showed that $\hat{S}^0$ is a generalized maximum likelihood estimation (GMLE) as defined by Kiefer and Wolfowitz (1956). Efron (1967) showed that $\hat{S}^0$ is the unique solution of the self-consistency equation (1.1):

$$(1.1) \quad \hat{S}^0(t) = \sum_i [T_i > t]/n + \sum_i (1-D_i) [T_i < t] \hat{S}^0(t)/\hat{S}^0(T_i)$$

Extensions of the product-limit estimator to bivariate times have been studied by Hanley and Parnes (1983), Muñoz (1980 a,b), Campbell (1981, 1982), and Campbell and Foldes (1980). Hanley and Parnes and Campbell studied maximum likelihood estimators for discrete data and extended self-consistency to this situation. Muñoz (1980 a) defined self-consistency of estimators for continuous bivariate distributions when censoring occurs simultaneously in both coordinates. He also showed that all GMLE's are self-consistent estimators. Unfortunately, there can be many GMLE's in this situation. Leurgans, Tsai and Crowley (1982) point out that some sequences of self-consistent estimators of bivariate survival function are asymptotically inconsistent.
Campbell and Foldes (1980) proposed two other estimators. Before describing these, we extend the notation to bivariate times. The true pair of survival times will be denoted by \((T_1^0, T_2^0)\). The bivariate survival function of this vector is \(S^0(t_1, t_2) = P(T_1^0 > t_1, T_2^0 > t_2)\). The pair of censoring times is \((C_1, C_2)\) and has bivariate survival function \(H\). The observed vector is \((T_1, T_2, D_1, D_2)\), where \(T_j = T_j^0 \wedge C_j, D_j = [T_j^0 < C_j], j = 1, 2\). When \(P(C_1 = C_2) = 1\), the censoring will be referred to as univariate censoring. Otherwise, the censoring will be called bivariate censoring. (For discussion of these censoring mechanisms, see Leurgans, Tsai and Crowley (1982).)

The first estimator Campbell and Foldes propose is based on paths from \((0,0)\) to \((s,t)\) and the observation \(S^0(t_1, t_2) = S^0(t_1, 0) P(T_2^0 > t_2 | T_1^0 > t_1)\). Each term in this product is estimated separately.

Their second estimator is based on an estimator \(\hat{R}(t_1, t_2)\) of hazard function \(R(t_1, t_2) = -\ln S^0(t_1, t_2)\) and the following equation:

\[
(1.2) \quad R(t_1, t_2) = \int_{(0,0)}^{(t_1, t_2)} \left( \frac{\partial \ln S^0(x,y)}{\partial x}, \frac{\partial \ln S^0(x,y)}{\partial y} \right) (dx, dy),
\]

where the integral is a line integral from \((0,0)\) to \((t_1, t_2)\). The survival function is then estimated by \(\exp - \hat{R}(t_1, t_2)\). They show the estimator \(\hat{R}\) is path dependent and point out another serious weakness: both of their estimators of \(S^0(t_1, t_2)\) may fail to be survival functions. Campbell (1982) showed the weak convergence of these estimators.
2. Decomposition of Bivariate Survival Functions.

Peterson (1977) introduced a decomposition of a univariate function in terms of identifiable survival and subsurvival functions. Theorem 2.1 below is a bivariate analogue of Peterson's decomposition. In section 3 we present estimators of a bivariate survival function based on this decomposition.

Throughout the rest of this paper, we give formulas for \( t_1 > t_2 \). Definitions for \( t_1 < t_2 \) are obtained by reversing the coordinates.

We use two assumptions (A1) and (A2) to derive the decomposition.

(A1) The vectors \((T_1^0, T_2^0)\) and \((C_1, C_2)\) are mutually independent.

(A2) The functions \(S^0\) and \(H\) are absolutely continuous with respect to Lebesgue measure on \(\mathbb{R}^2\).

Without some assumptions about the relationship between \((T_1^0, T_2^0)\) and \((C_1, C_2)\), \(S^0\) is not identifiable. The assumption (A1) is realistic in some situations. The assumption (A2) can be weakened, but is convenient for exposition.

The decomposition is expressed in terms of the following quantities and sets:

\[ S(x,y) = P(T_1 > x, T_2 > y), \]

\[ S_3^0(y) = S^0(y,y), \]

\[ S_2(x,y) = P(T_1 > x_1, T_2 > y, D_2 = 1), \]
(2.1) \[ S_{12}(x,y) = P(T_1 > x, T_2 > y), \quad D_1 = 0, \quad D_2 = 1, \]

\[ S^0(x|y) = P(T_1^0 > x| T_2^0 = y), \]

\[ R(s,t) = \{(x,y) \mid x > s \land y > t\}, \]

\[ \Delta(s,t) = \{(x,y) \mid s > x > y > t\}. \]

Lemma 2.1 is a preliminary decomposition of \( S^0(s,t) \). It is not a complete decomposition in terms of identifiable functions, because \( S^0_3(y) \) and \( S^0(x|y) \) cannot be estimated directly. However, since \( S^0_3 \) and \( S^0(.|y) \) are univariate survival functions, the univariate decomposition given in Lemma 2.2 below applies.

Lemma 2.1. Assume conditions (A1) and (A2) hold and let \( s > t > 0 \) be such that \( S(s,s) > 0 \). Then

(2.2) \[ S^0(s,t) = S^0_3(s) + \int \int_{R(s,t)} \frac{S^0_3(y)}{S(y,y)} D S_2(x,y) \]

\[ + \int \int_{\Delta(s,t)} \frac{S^0_3(y)}{S(y,y)} \frac{S^0(s|y)}{S^0(y|y)} D S_{12}(x,y), \]

where \( \frac{0}{0} = 0 \).

Proof: We prove this lemma by rewriting each of the double integrals as integrals with respect to \( D S^0(x,y) \). For the first integral, we interchange limits to show that this integral is equal to
\[- \int_t^s \frac{S_3^O(y)}{S(y,y)} D_y S_2^O(s,y). \]

Since the definitions and (A1) imply that \(S(y,y) = S_3^O(y)H(y,y)\) and that \(D_y S_2^O(s,y) = H(s,y) D_y S^O(s,y)\), this integral reduces to

\[(2.3) \quad - \int_t^s \frac{H(s,y)}{H(y,y)} D_y S^O(s,y) = \int \int_{R(s,t)} \frac{H(s,y)}{H(y,y)} DS^O(x,y). \]

To evaluate the second integral, let

\[f^O(x,y) = \frac{\partial S^O(x,y)}{\partial x \partial y} \quad \text{and} \quad h(x,y) = \frac{\partial^2 H(x,y)}{\partial x^2}.\]

Assumption (A1) implies that

\[(2.4) \quad \frac{S^O(x|y)}{S^O(x|y)} D S_2^O(x,y) \]

\[= \int_s^\infty f^O(u,y) \, du \int_y^\infty h(x,n) \, dn \, dx \, dy. \]

Substituting (2.4) in the second integral of (2.2) and using \(S(y,y) = S_3^O(y)H(y,y)\) again gives

\[\int_t^s \int_s^\infty \frac{f^O(u,y)du}{H(y,y)} \left( \int_x^y \int_n^y h(x,n) \, dn \, dx \right) \, dy \]

\[= \int_t^s \int_s^\infty \frac{f^O(u,y)}{H(y,y)} \, du \left[ H(y,y) - H(s,y) \right] \, dy \]
\[ (2.5) \quad \int_{\mathcal{R}(s,t)} \int \left[ 1 - \frac{H(s,y)}{H(y,y)} \right] f^0(u,y) du \, dy = \int_{\mathcal{R}(s,t)} \left[ 1 - \frac{H(s,y)}{H(y,y)} \right] ds^0(x,y). \]

Therefore the sum of the two integrals is the sum of (2.3) and (2.5) or

\[ \int_{\mathcal{R}(s,t)} DS^0(x,y) = S^0(s,t) - S^0(s,s) = S^0(s,t) - S^0_3(s). \]

The lemma follows. Q.E.D.

Lemma 2.2 is Beran's (1981) extension of Peterson's (1977) decomposition to allow simultaneous discontinuities in death and censoring times. The statement of the decomposition in terms of the product integral clarifies the similarity of the theoretical decomposition and the estimators of Section 3.

The product integral of a function \( g \) is defined by

\[ \gamma(g)(t) = \lim_{\max \{ u_k - u_{k-1} \} + 0} \prod_{i=1}^{r} \{ 1 - (g(u_i) - g(u_{i-1})) \}, \]

where \( 0 = u_0 < u_1 < \ldots < u_r = t. \)

For continuous functions \( g \), \( \gamma(g)(t) = \exp(-g(t)) \). If \( g \) is an empirical cumulative hazard, \( \gamma(g) \) is the corresponding product-limit or Kaplan-Meier estimator.
Lemma 2.2.

If \( S_0(t) = P(T^0 > t), S(t) = P(T^0 \wedge C > t), \) and \( S_u(t) = P(C > T^0 > t), \)
where \( T^0 \) and \( C \) are independent random variables, then

\[
S_0(t) = \gamma \left( -\int_0^{t^+} \frac{DS_u(x)}{S(x^-)} \right) (t).
\]

The function \( S^0_3 \) is the survival function of \( T^0_3 = T^0_1 \wedge T^0_2 \).
Define the corresponding censoring time \( C_3 = T^0_1 \wedge C_2 \), the observed time \( T_3 = T^0_3 \wedge C_3 \), and the indicator \( D_3 = [T_3^0 \leq C_3] \). The survival function \( S^0_3 \) can be decomposed in terms of \( S_3(t) = P(T_3 > t) \) and \( S_{3u}(t) = P(T_3 > t, D=1) \). Since \( T_3 = T_1 \wedge T_2 \) and \( D_3 = [T_1 > T_2] D_2 + [T_1 < T_2] D_1 \), \( (T_3, D_3) \) is a function of \( (T_1, T_2, D_1, D_2) \) and \( S_3 \) and \( S_{3u} \)
can be estimated empirically.

We now state the decomposition.

Theorem 2.1. If the conditions (A1) and (A2) are met, then

\[
S^0(s,t) = \gamma \left( \int_0^{t^+} \frac{DS_{3u}(z)}{S_3(z^-)} \right) (s) + \int \int \frac{\gamma (-\int_0^{t^+} \frac{DS_{3u}(z)}{S_3(z^-)} (z^-) ) (y^-) }{S(y^-, y^-)} \frac{DS_2(x,y)}{R(s,t)} + \int \int \frac{\gamma (-\int_0^{t^+} \frac{DS_3(z)}{S_3(z^-)} ) (y^-) }{S(y^-, y^-)} \frac{DS_{12}(x,y)}{\Delta(s,t)} + \int \int \frac{\gamma (-\int_0^{t^+} \frac{DS_u(z | y)}{S(z | y)} (s) ) (x^-) }{\gamma (-\int_0^{t^+} \frac{DS_u(z | y)}{S(z | y)} ) (x^-)} \frac{DS_{12}(x,y)}{\Delta(s,t)},
\]

where \( S_u(t | y) = P(T_1 > t, D_1 = 1 | T_2 = y, D_2 = 1) \).
and \[ S(t|y) = P(T_1 > t | T_2 = y, D_2 = 1). \]

Proof: From the assumption (A1), \( C_2 \) is independent of \( T_1^0 \), we observe

\[
S^0(t|y) = P(T_1^0 > t | T_2^0 = y) = P(T_1^0 > t | T_2^0 = y, C_2 > y) = P(T_1^0 > t | T_2 = y, D_2 = 1).
\]

The theorem follows from applying Lemma 2.2 to \( S_3 \) and \( P(T_1^0 > t | T_2 = y, D_2 = 1) \) and substituting the resulting representation in Lemma 2.1.

Q.E.D.

Remark. The conclusions of Lemma 2.1 and Theorem 2.1 also hold in the presence of univariate censoring.

3. Estimators of \( S^0 \)

Suppose the iid random vectors \( \{ T_{1i}, T_{2i}, D_{1i}, D_{2i} \}, i = 1, \ldots, n \) have the same distribution as the random vector \( (T_1, T_2, D_1, D_2) \). In this section we develop an estimator of \( S^0 \) and establish its consistency.

Natural unbiased estimators of the (sub)survival functions in (2.6) are defined below in terms of \( T_{1i}, T_{2i}, D_{1i}, D_{2i}, T_{3i} = T_{1i} \wedge T_{2i} \) and \( D_{3i} = [T_{1i} > T_{2i}] D_{2i} + [T_{1i} < T_{2i}] D_{2i} \):
\[ S^e(x, y) = \frac{1}{n} \sum_i \mathbb{1}[T_{1i} > x, T_{2i} > y], \]
\[ S^e_{12}(x, y) = \frac{1}{n} \sum_i \mathbb{1}[T_{1i} > x, T_{2i} > y, D_{1i} = 1, D_{2i} = 1], \]
\[ S^e_2(x, y) = \frac{1}{n} \sum_i \mathbb{1}[T_{1i} > x, T_{2i} > y, D_{2i} = 1], \]
\[ S^e_3(x) = \frac{1}{n} \sum_i \mathbb{1}[T_{3i} > x], \]
\[ \text{and } S^e_{3u}(x) = \frac{1}{n} \sum_i \mathbb{1}[T_{3i} > x, D_{3i} = 1]. \]

Substituting \( S^e_3 \) and \( S^e_{3u} \) into the equation (2.4), we have the Kaplan-Meier estimator for \( S^O_3(t) \):

\[ (3.1) \quad \hat{S}^O_3(t) = \gamma(- \int_0^t \frac{D S^e_{3u}(x)}{S^e_3(x-)} (t) \]

The functions \( S(x|y) \) and \( S_u(x|y) \) are the conditional probabilities given \( T_{2i} = y \) and \( D_{2i} = 1 \). Since the assumption of absolute continuity implies that at most one \( T_{2i} = y \) with \( D_{2i} = 1 \), \( S(x|y) \) and \( S_u(x|y) \) cannot be estimated stably without smoothing. To estimate a conditional survival function given \( y \), we apply the non-negative weight \( W_{ni}(y) \) to \((T_{1i}, D_{1i})\). The weights depend on the data through the second components, \{\((T_{2j}, D_{2j})\), \( j = 1(1)n \)\}. With the assumption that \( \sum_{i=1}^{n} W_{ni}(y) D_{2i} = 1 \), the following estimators are discrete survival functions:
\[ \hat{S}(x|y) = \sum_{i=1}^{n_i} W_i (y) \mathbb{I}_{T_i > x, D_{2i} = 1} \]

(3.2)

\[ \hat{S}_u (x|y) = \sum_{i=1}^{n_i} W_i (y) \mathbb{I}_{T_i > x, D_{1i} = D_{2i} = 1}. \]

This class of estimators has several attractive features. One feature is that the use of Theorem 3.1 to construct the estimators makes it a natural extension of the product-limit estimator to two dimensions. In Section 4 we show that these estimators are self-consistent. Another advantage is that, once the weights have been computed, no iteration is required.

Substituting the estimators \( \hat{S}(x|y) \) and \( \hat{S}_u (x|y) \) into the equation yields the following natural estimator for \( \hat{S}^0(x|y) \):

\[ \hat{S}^0(x|y) = \gamma \left( - \int_0^u \frac{D_t \hat{S}_u (t|y)}{\hat{S}(t|y)} \right) (x). \]

If \( \hat{S}(x|y) > 0 \), and if the jump points of \( \hat{S}_u (.|y) \) are \( x_1, \ldots, x_m \), then

\[ \hat{S}^0(x|y) = \prod_{x_i < x} \left[ 1 - \frac{\hat{S}_u (x_i^- |y) - \hat{S}_u (x_i |y)}{\hat{S}(x_i^- |y)} \right]. \]

Substituting \( \hat{S}^0_3 \), \( S^e_2 \), \( S^e_{12} \), \( S^e \) and \( \hat{S}^0(x|y) \) into (2.2), we obtain the following estimator of \( S^0 \):

\[ \hat{S}^0(s,t) = \hat{S}^0_3(s) + \int \int_{R(s,t)} \hat{S}^0_3(y^-) / s^e(y^-, y^-) D s^e_2(x,y) \]

\[ + \int \int_{A(s,t)} \hat{S}^0_3(y^-) \hat{S}^0(s|y^-) / (s^e(y^-, y^-) \hat{S}^0(x^--y^-)) D s^e_{12} (x,y). \]
We devote the rest of this section to a discussion of the consistency of this class of estimators. First we discuss the conditions on the weight functions which imply consistent estimation of the conditional survival functions. Lemma 3.1 asserts that these conditions imply consistent estimation of $\hat{S}^0 (\cdot | y)$. Theorem 3.1 gives conditions for the consistency of $\hat{S}^0$.

Choice of an estimator within this class requires the specification of the weight functions $W_n_i(y)$. We confine our investigation to a demonstration that the weight functions can be chosen to provide a consistent sequence of estimators. Of course, important issues remain in the selection of weight functions for use with specific sample sizes.

Here we focus on kernel weights. These weights are constructed by selecting a non-negative function $k(\cdot)$ of bounded variation on the real line and a sequence of positive bandwidths $\{h(n), n \geq 1\}$ converging to zero. The probability weights conditional on $y$ are then

$$W_{n,i}(y) = k((T_{2i} - y)/h(n)) D_{2i}/(\sum_j k((T_{2j} - y)/h(n)) D_{2j}).$$

Theorem A.1 of the Appendix (also see Theorem 4.3 of Tsai (1982)) shows that if the true conditional survival functions $S(x | y)$ are uniformly continuous in $x$ and in $y$ and if bandwidths converge slowly enough that $\sum_{i=1}^\infty \exp(-rh^2(n)) < \infty$ for every positive $r$, then

$$\sup_{0 < x, y < M} |\hat{S}(x | y) - S(x | y)| \rightarrow 0, \text{ a.s. and}$$

$$\sup_{0 < x, y < M} |\hat{S}_u(x | y) - S_u(x | y)| \rightarrow 0, \text{ a.s. as } n \rightarrow \infty,$$

where $M$ is such that $\int_0^\infty \int_t^\infty \int_y^\infty f^0(t, y) h(u, v) dv du dt > 0$ for every $y < M$. 

-13-
In our proofs below, we do not assume that the kernel weights have been used. Nearest neighbor weights can also be used. Beran (1981), in related work on nonparametric regression in the presence of censoring, discusses these possibilities.

(A3) The probability weights have been chosen so that (3.4) holds.

Lemma 2.1. Assume (A.1) and (A.3) hold. If the constant M satisfies \( S(x,y) > 0 \) for \( x,y < M \),

then \( \sup_{0 < x,y < M} | \tilde{S}^0(x|y) - S^0(x|y) | \to a.s. \text{ as } n \to \infty. \)

Proof: Consider

\[
T(G_1, G_2, G_3)(x, y) = -G_1(x, y) + G_2(x, y) - \int_0^x \frac{G_1(z, y)}{G_2(z, y)} D_z [G_2(z, y) - G_3(z, y)],
\]

where for any fixed \( y \), \( G_i(x, y) \) \( i=1, 2 \) are survival functions and \( G_3(x, y) \) is subsurvival function such \( G_3(x, y) < G_2(x, y) \) for \( x, y > 0 \). Theorem 2.4 of Tsai (1983) shows that the unique solution of \( T(\psi(G_2, G_3), G_2, G_3)(x, y) = 0 \) for any fixed \( y = y_0 \) is

\[
\psi(G_2, G_3)(x, y_0) = \gamma \left( \frac{u^+}{\alpha} G_3(t, y_0) \right) (x) \quad \text{if } G_2(x, y_0) > 0.
\]
Therefore, for $x, y < M$,

$$
\psi(\hat{S}(\cdot | \cdot), \hat{S}_u(\cdot | \cdot))(x, y) = \\
\gamma \left( \int_0^u + \frac{D_t \hat{S}_u(t | y)}{\hat{S}(t^- | y)} \right)(x) = \hat{S}^0(x | y)
$$

is the unique solution of $T(\hat{G}_1, \hat{S}(\cdot | \cdot), \hat{S}_u(\cdot | \cdot)) = 0$. Similarly, $S^0(x | y)$ is the unique solution of $T(G_1, S(\cdot | \cdot), S_u(\cdot | \cdot)) = 0$.

Let $g(x, y)$ denote $\sup_{0 < x, y < M} |g(x, y)|$. For any conditional survival function $S^*(x | y)$ satisfying $S^*(x | y) - S^0(x | y) > \epsilon$, we have $T(S^*(\cdot | \cdot), S(\cdot | \cdot), S_u(\cdot | \cdot)) \not\to 0$. The assumption (A.3) implies that

$$
T(S^*(\cdot | \cdot), \hat{S}(\cdot | \cdot), \hat{S}_u(\cdot | \cdot)) - T(S^*(\cdot | \cdot), S(\cdot | \cdot), S_u(\cdot | \cdot))
$$

$\to 0$, a.s. as $n \to \infty$.

Therefore, for almost all realizations, there exists an $m$ such that for all $n > m$, $T(S^*(\cdot | \cdot), \hat{S}(\cdot | \cdot), \hat{S}_u(\cdot | \cdot))(x, y) \neq 0$.

Thus the unique solution $\hat{S}^0(x | y)$ to $T(\hat{S}^0(\cdot | \cdot), \hat{S}(\cdot | \cdot), \hat{S}_u(\cdot | \cdot))(x, y) = 0$ must satisfy $\hat{S}(x | y) - S^0(x | y) < \epsilon$. As $\epsilon$ is arbitrary, we have that $\hat{S}^0(x | y) - S^0(x | y) \to 0$, a.s. as $n \to \infty$. Q.E.D.

The uniform consistency of $\hat{S}^0_3$, which is needed to derive the consistency of $\hat{S}^0$, follows from the one-dimensional result for the Kaplan-Meier estimator from Foldes and Rejtő (1981).
Proposition 3.1

If (A.1) holds and the constant M satisfies $S_3(M) > 0$, then

$$\sup_{0 \leq t \leq M} |\hat{S}_3^0(t) - S_3^0(t)| + 0, \text{ a.e. } n \to \infty.$$ 

Theorem 3.1. Assume (A.1) through (A.3) hold. If the constant M is such that $S(x|y) > 0$, for $x, y < M$ and $S_3(M) > 0$, then

$$\sup_{0 < x < y < M} |\hat{S}_1^0(x, y) - S_1^0(x, y)| + 0, \text{ a.s. as } n \to \infty$$

Proof: Proposition 3.1, the uniform consistency of the empirical survival function, and Lemma 3.1 imply that $\hat{S}_3^0(y) / S(y), \hat{S}_2^0(x, y), \hat{S}_3^0(y) \hat{S}_1^0(x|y) (S(y) \hat{S}_0^0(x|y))$ and $S_1^0(x, y)$ converge uniformly to $S_3^0(y)/S(y), S_2(x, y), S_3^0(y)S_1^0(s|y)/(S(y)S(x|y))$ and $S_1^0(x, y)$ respectively. The Theorem follows from the bivariate generalization of Lemma 6 of Aalen (1976). (See Appendix for a statement of this Lemma.)

4. Self-consistency of $\hat{S}_0^0$ under univariate censoring

Self-consistency was defined for estimators of survival function for univariate observations by Efron (1967) and extended to bivariate observations by Hanley and Parnes (1983), Campbell and Foldes (1981) and Muñoz (1980 a, b). An estimator of $\hat{S}_0^0$ of $S_0^0$ is said to be self-consistent if

$$n \hat{S}_0^0(t_1, t_2) = \sum_{i=1}^{n} \left[ [T_{1i} > t_1, T_{2i} > t_2] + [D_{1i} = 0, D_{2i} = 1, T_{1i} < t, T_{2i} > t_2] \right] \left( \frac{\hat{S}_0^0(t_1, T_{2i}^-) - \hat{S}_0^0(t_1, T_{2i})}{\hat{S}_0^0(T_{1i}, T_{2i}^-) - \hat{S}_0^0(T_{1i}, T_{2i})} \right)$$
\[ + [D_{1i} = 1, D_{2i} = 0, T_{1i} > t, T_{2i} < t] \left( \frac{\hat{S}_0(T_{1i}, t) - \hat{S}_0(T_{2i}, t)}{\hat{S}_0(T_{1i}, T_{2i}) - \hat{S}_0(T_{1i}, T_{2i})} \right) \]
\[ + [D_{1i} = D_{2i} = 0, T_{1i} < t_1 \text{ or } T_{2i} < t_2] \left( \frac{\hat{S}_0(\max(t_1, T_{1i}), \max(t_2, T_{2i}))}{\hat{S}_0(T_{1i}, T_{2i})} \right) \]

Proposition 4.1 from Muñoz (1980a) characterizes self-consistency under univariate censoring in terms of \( m \), the probability mass function of the product-limit estimator of the univariate censoring distribution:

\[
m(t) = \begin{cases} 
1/n \sum_{j=1}^{n} \frac{1}{\sum_{j=1}^{n} \mathbb{I}[T_{3j} > T_{3i}]}(1-D_{3j}) & \text{if } t = T_{3k}, D_{3k} = 1, k=1, \ldots, n. \\
0 & \text{otherwise.}
\end{cases}
\]

We state Muñoz's characterization and use it to show that \( \hat{S}_0 \) is self-consistent.

**Proposition 4.1.**
Under univariate censoring, if the \( 2n \) observed times \( \{T_{ki}, 1 \leq i \leq n, k=1,2\} \) are distinct, then \( \hat{S}_0 \) is a self-consistent estimator of \( S_0 \) if and only if condition (iv) holds and, for each \( i \), the appropriate condition of (i) through (iii) holds:

(i) \( \hat{S}_0(T_{1i}, T_{2i}) - \hat{S}_0(T_{1i}, T_{2i}) - \hat{S}_0(T_{1i}, T_{2i}) + \hat{S}_0(T_{1i}, T_{2i}) = m(T_{1i} \wedge T_{2i}), \text{ if } D_{1i} = D_{2i} = 1, \)

(ii) \( \hat{S}_0(T_{1i}, T_{2i}) - \hat{S}_0(T_{1i}, T_{2i}) = m(T_{2i}), \text{ if } D_{1i} = 0, D_{2i} = 1 \)
(iii) \( \hat{S}^0(T_{1i}^{\sim}, T_{2i}^{\sim}) = \hat{S}^0(T_{1i}, T_{2i}) = m(T_{1i}), \) if \( D_{1i} = 1, \) \( D_{2i} = 0 \)

(iv) \( \hat{S}^0(t, t) = \sum_{j=1}^{n} \left[ D_{3j} = 1, T_{1j}^T, T_{2j}^T \right] m(T_{1j}^T, T_{2j}). \)

Theorem 4.1

Under univariate censoring, if the 2n observed times \( \{T_{ki}, 1 \leq i \leq n, k = 1, 2\} \) are distinct, then \( \hat{S}^0 \) is a self-consistent estimator of \( S^0. \)

Proof: The theorem will be proved by showing that \( \hat{S}^0 \) satisfies the four characterizing equations of Proposition 4.1.

(i) If this equation applies, \( D_{3i} = 1. \) Since \( \hat{S}^0(\cdot, \cdot) \) is the product limit estimator of \( S^0 \), Efron's (1967) equation (7.9) asserts that \( \hat{S}^0(T_{3i}^{\sim}) - \hat{S}^0(T_{3i}) = m(T_{3i}). \) Therefore we must show that the probability mass \( \hat{S}^0 \) places at \( T_{1i} > T_{2i} \) is \( m(T_{1i}^T, T_{2i}). \) If \( D_{3i} = 1 \) and \( T_{1i} > T_{2i} \), the definition of \( \hat{S}^0 \) implies that the mass \( \hat{S}^0 \) places at \( (T_{1i}, T_{2i}) \) is \( n^{-1} \hat{S}^0(T_{2i}^T)/ S^e(T_{2i}^T, T_{2i}^T) = \hat{S}^0(T_{2i}^T) - \hat{S}^0(T_{2i}), \) as required. The proof for \( T_{2i} > T_{1i} \) is similar.

(ii) If this equation applies under univariate censoring, \( T_{1i} > T_{2i}. \) The definition of \( \hat{S}^0 \) implies that

\[
\hat{S}^0(T_{1i}^{\sim}, T_{2i}^{\sim}) - \hat{S}^0(T_{1i}, T_{2i})
= \hat{S}^0(T_{2i}^{\sim}, T_{2i}^{\sim}) \hat{S}^0(T_{1i}^{\sim} | T_{2i}^{\sim}) / (n S^e(T_{2i}^{\sim}, T_{2i}^{\sim}) \hat{S}^0(T_{1i}^{\sim} | T_{2i}^{\sim})).
= \hat{S}^0(T_{2i}^{\sim})/ n S^e(T_{2i}^{\sim}, T_{2i}^{\sim}) = S_3(T_{3i}^{\sim}) - S(T_{2i}) = m(T_{2i}) \text{ for } D_{1i} = 0, D_{2i} = 1.
\]
(iii) This case follows from (ii) by symmetry.

(iv) Since \( \hat{S}^0(t, t) = \hat{S}_3^0(t) \), we have

\[
\hat{S}^0(T_{3i}, T_{3i}) - \hat{S}_3^0(T_{3i}, T_{3i}) = m(T_{3i}) D_{3i}.
\]

Therefore, \( \hat{S}^0(t, t) = \sum_{i=1}^{n} [T_{3i} > t, D_{3i} = 1] m(T_{3i}) \).

Q.E.D.

5. Example

In this section, we apply our estimator to data from Muñoz (1980a). Table 5.1 gives the survival times and recurrence times from treatment (surgery) for 25 patients with stage III malignant melanoma.
Table 5.1 Days to recurrence and days to death of stage III malignant melanoma patient after treatment

<table>
<thead>
<tr>
<th>i</th>
<th>$X_{1i}$</th>
<th>$X_{2i}$</th>
<th>$D_{1i}$</th>
<th>$D_{2i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>220</td>
<td>253</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>262</td>
<td>1225</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1569</td>
<td>1569</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2233</td>
<td>2233</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1305</td>
<td>1305</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>348</td>
<td>658</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>816</td>
<td>894</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>249</td>
<td>380</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>182</td>
<td>189</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>313</td>
<td>1088</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>326</td>
<td>719</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>100</td>
<td>1326</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1336</td>
<td>1336</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>473</td>
<td>568</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>32</td>
<td>66</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>602</td>
<td>602</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>180</td>
<td>782</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>121</td>
<td>773</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>333</td>
<td>453</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>140</td>
<td>144</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>424</td>
<td>743</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>303</td>
<td>303</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>88</td>
<td>249</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>88</td>
<td>207</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>116</td>
<td>116</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$X_{1i}$: Days to recurrence after treatment

$X_{2i}$: Days to death after treatment

$D_{1i}$: Recurrence indicator of $X_{1i}$; 1 = recurrence, 0 = censored

$D_{2i}$: Death indicator of $X_{2i}$; 1 = death, 0 censored.

Source: Muñoz (1980a, p. 47)
In order to obtain an estimate \( \hat{S}^0 \) as defined in the last section, we have to choose the kernel function \( k \) and window width \( h(n) \). In this example, the kernel function is the standard normal density and window width \( h(n) = 100 \). The estimator \( \hat{S}^0 \) can be computing by first calculating the conditional probability \( \hat{S}^0(x|y) \) and K-M estimate \( \hat{S}^0 \), and then substituting \( \hat{S}^0(x|y) \), \( \hat{S}^0 \) and the empirical (sub)survival functions into (3.2).

We describe \( \hat{S}^0 \) by its step sizes at each jump point. Since \( \hat{S}^0(2233, 2233) = .2184 \) and \( \hat{S}^0(t_1, t_2) \) is undefined for \( t_1 > 2233 \) or \( t_2 > 2233 \), for convenience we take \( \hat{S}^0 \) to have a jump at \( (2233, 2233) \) of size .2184. Table 5.2 contains the jump sizes and jump points of estimate \( \hat{S} \) and an estimate proposed by Muñoz (1980a).
<table>
<thead>
<tr>
<th>$\hat{S}^0$</th>
<th>Muñoz</th>
<th>Jump Points $(a,b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.04</td>
<td>.04</td>
<td>$(32,66); (88,207); (100,1326)$</td>
</tr>
<tr>
<td>.042</td>
<td>.042</td>
<td>$(140,144); (182,189); (220,253); (249,280); (262,1225)$</td>
</tr>
<tr>
<td>.0455</td>
<td>.0455</td>
<td>$(326,719); (348,658); (473,568)$</td>
</tr>
<tr>
<td>.0053</td>
<td>.0059</td>
<td>$(121,1225)$</td>
</tr>
<tr>
<td>.0367</td>
<td>.0056</td>
<td>$(121,1326)$</td>
</tr>
<tr>
<td>.00207</td>
<td>.0455</td>
<td>$(180,1225)$</td>
</tr>
<tr>
<td>.0213</td>
<td>.0059</td>
<td>$(180,1326)$</td>
</tr>
<tr>
<td>.0053</td>
<td>.0305</td>
<td>$(121,2233+)$</td>
</tr>
<tr>
<td>.0449</td>
<td>.0064</td>
<td>$(180,2233+)$</td>
</tr>
<tr>
<td>.0006</td>
<td>.0061</td>
<td>$(313,1225)$</td>
</tr>
<tr>
<td>.0034</td>
<td>.0031</td>
<td>$(313,1326)$</td>
</tr>
<tr>
<td>.0100</td>
<td>.0031</td>
<td>$(333,568)$</td>
</tr>
<tr>
<td>.0102</td>
<td>.0031</td>
<td>$(333,658)$</td>
</tr>
<tr>
<td>.0216</td>
<td>.0051</td>
<td>$(333,719)$</td>
</tr>
<tr>
<td>.0003</td>
<td>.0048</td>
<td>$(333,1225)$</td>
</tr>
<tr>
<td>.0000</td>
<td>.0253</td>
<td>$(333,1326)$</td>
</tr>
<tr>
<td>.0455</td>
<td>.0064</td>
<td>$(424,1225)$</td>
</tr>
<tr>
<td>.0000</td>
<td>.0061</td>
<td>$(424,1326)$</td>
</tr>
<tr>
<td>.0030</td>
<td>.0253</td>
<td>$(424,2233+)$</td>
</tr>
<tr>
<td>.0546</td>
<td>.0076</td>
<td>$(816,1225)$</td>
</tr>
<tr>
<td>.0000</td>
<td>.0073</td>
<td>$(816,1326)$</td>
</tr>
<tr>
<td>.0000</td>
<td>.0397</td>
<td>$(816,2233+)$</td>
</tr>
<tr>
<td>.2184</td>
<td>.2184</td>
<td>$(2233,2233)$</td>
</tr>
</tbody>
</table>

* Muñoz put the mass at line from $(1,b)$ to $(a,\infty)$
Lemma A.1 Let \( \{ h(n) = n > 1 \} \) be a sequence of positive real numbers such that \( h(n) \to 0 \) and \( \sum_{n=1}^{\infty} \exp(\gamma nh^2(n)) \to \infty \) for every \( \gamma > 0 \). Let \( k(u) \) be a density function of bounded variation on the real line. Set

\[
G_1(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^0(t,y) h(u,v) \, dv \, dv \, dt,
\]

\[
G_2(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^0(t,y) h(u,v) \, dr \, du \, dt,
\]

\[
\hat{G}_1(x,y) = \sum_i k((T_{2i} - y)/h(n))/h(n) [T_{1i} \geq x, D_{2i} = 1]
\]

and

\[
\hat{G}_2(x,y) = \sum_i k((T_{2i} - y)/h(n))/h(n) [T_{1i} \geq x, D_{1i} = D_{2i} = 1].
\]

If \( G_i(x,y), i=1,2 \), are uniformly continuous for \( 0 < x, y < \infty \), then

\[
\sup_{0<x,y<\infty} |\hat{G}_i(x,y) - G_i(x,y)| \to 0, \text{ a.s., as } n \to \infty.
\]

for \( i = 1,2 \).

Proof: The proof is similar to that of Theorem 1 of Nadaraya (1965) where uniform consistency is established for univariate kernel density estimator.

We abbreviate \( \sup_{0<x,y<\infty} \) by \( \sup \). Integrating the double integral by parts (Young (1917)) we find
\[ A_n = \sup |\hat{G}_1(x,y) - E(\hat{G}_1(x,y))| \]

\[ = \sup \left| \frac{1}{h(n)} \int \int k \left( \frac{t-y}{h(n)} \right) [s>x] DS^e_2(s,t) - \frac{1}{h(n)} \int \int \left( \frac{t-y}{h(n)} \right) [s>x] DS_2(s,t) \right| \]

\[ = \sup \left| \frac{1}{h(n)} \int \int S^e_2(s,t) - S_2(s,t) d_t \left( \frac{t-y}{h(n)} \right) d_s [s>x] \right| \]

\[ < \sup \left| \frac{1}{h(n)} \int \int S^e_2(s,t) - S_2(s,t) d_t \left( \frac{t-y}{h(n)} \right) d_s [s>x] \right| \]

where \( \nu = \int |dk| \) is the variation of \( k \). Theorem 1 of Kiefer and Wolfowitz (1956) implies that for any \( \varepsilon \),

\[ P(A_n > \varepsilon) < P\{\sup |S^e_2(x,y) - S_2(x,y)| > \varepsilon h(n)/\nu\} \]

\[ < C \exp - \varepsilon_1 nh^2(n), \]

where \( \varepsilon_1 = \varepsilon/\nu \)^2 and \( 0 < C < \infty \). The Borel-Cantelli Lemma and the assumption that \( k(\cdot) \) has bounded variation imply that \( A_n \to 0 \) a.s. as \( n \to \infty \). It remains to show that

\[ \sup |E(\hat{G}_1(x,y)) - G_1(x,y)| \to 0, \text{ a.s. as } n \to \infty. \]

Let \( g(x,y) = \frac{a^2 S_2(x,y)}{a_x a_y} \). Since \( G_1(x,y) = \frac{a}{a_y} S_2(x,y) \), and \( G_1(0,y) = 0 \), we have

\[ G_1(x,y) = \int [s>x] y[s,y] ds = \int \frac{1}{h(n)} h\left( \frac{t}{h(n)} \right) \int [s>x] g(0,y) ds dt. \]
Therefore, for \( \delta > 0 \),

\[
\sup |\hat{E}G_1(x, y) - G_1(x, y)|
\]

\[
< \sup \{ \int |t| < \delta + \int |t| > \delta \} \left( \frac{1}{h(n)} h\left( \frac{t}{h(n)} \right) \right) \int [s > x] |g(s, y - t) - g(s, y)| \, ds \} \, dy
\]

\[
< \sup \sup |\int [s > x] |g(s, y - t) - g(s, y)| \, ds | \left| t \right| < \delta
\]

\[
+ 2M \int |t| > \delta / h(n) k(t) \, dt.
\]

\[
= \sup \sup |G(x, y - t) - G(x, y)| \left| t \right| < \delta
\]

\[
+ 2M \int |t| > \delta / h(n) k(t) \, dt
\]

(A.1)

where \( M = G_1(x, y) < \infty \). Let \( \varepsilon \) be an arbitrarily small positive number. Because \( G_1 \) is uniformly continuous, we can make the first term of the right-hand side of (A.1) less than \( \varepsilon / 2 \) by choosing \( \delta \) sufficiently small. Having so chosen \( \delta \), we can then choose \( n \) so large that \( \delta / h(n) \) is large enough so that \( \int |t| > \delta / h(n) k(t) \, dt < \varepsilon / 2M \).

Thus (A.1) implies that \( \sup |\hat{E}G_1(x, y) - G_1(x, y)| < \varepsilon \). Therefore \( |\hat{G}(x, y) - G_1(x, y)| \to 0 \), a.s. as \( n \to \infty \).
The uniform consistency of $\hat{G}_2$ can be established similarly. Q.E.D.

Theorem A.1. Assume the conditions of Lemma A.1 hold. If the constant $M$ satisfies that $G_m(y) > 0$ for $y < m$, then

(a) $\sup_{0 < x, y < M} |\hat{S}(x|y) - S(x|y)| \to 0$, a.s. as $n \to \infty$

(b) $\sup_{0 < x, y < M} |\hat{S}_u(x|y) - S_u(x|y)| \to 0$, a.s. as $n \to \infty$.

Where $\hat{S}(x|y)$ and $\hat{S}_u(x|y)$ are defined in (3.2) with

$$W_{n,i}(g) = k(T_{2i} - y)/h(n) \sum_{j} k((T_{2j} - y)/h(n)) D_{2i}.$$ 

Proof: The proof is based on Lemma A.1 and the fact

$$\hat{S}(x|y) = \frac{\hat{G}_1(x,y)}{\hat{G}_1(0,y)} \text{ and } \hat{S}_M(x|y) = \frac{\hat{G}_2(x,y)}{\hat{G}_2(0,y).}$$

Q.E.D.

Theorem A.2. Let $S_0(x,y), S_1(x,y), S_2(x,y), \ldots$, be a sequence of (sub)survival functions such that $0 = S_i(0,0) < S_i(1,1) < 1, i = 0, 1, \ldots$, $g(x,y)$ be a bounded continuous function. Let $g_1(x,y), g_2(x,y), \ldots$, be functions that for each $n$, $g_n$ is integrable with respect to the function $S_n$. Assume that $\sup_{0 < x, y \leq 1} |S_n(x,y) - S_0(x,y)| \to 0$ a.s. and $\sup_{0 < x, y \leq 1} |g_n(x,y) - g_0(x,y)| \to 0$, a.s. as $n \to \infty$. Then
\[ \sup_{0 \leq x, y \leq 1} \left| \int_0^x \int_0^y g_n(s,t) \, DS_n(s,t) \right| \rightarrow 0, \text{ a.s., as } n \rightarrow \infty. \]

Proof: This theorem is a bivariate extension of Lemma 6.1 of Aalen (1976).

Q.E.D.
References


Campbell, G. and Foldes, B. (1980). Large-sample properties of nonparametric bivariate estimators with censored data. Mimeo Series No. 80-10, Department of Statistics, Purdue University, West Lafayette, IN.


