

An inverse probability weighted estimator for the bivariate distribution function under right censoring

Hongsheng Dai¹, University of Lancaster, UK, and

Yanchun Bao, University of Manchester, UK

Abstract

An inverse probability weighted estimator is proposed for the joint distribution function of bivariate random vectors under right censoring. The new estimator is based on the idea of transformation of bivariate survival functions and bivariate random vectors to univariate survival functions and univariate random variables. The estimator converges weakly to a zero-mean Gaussian process with an easily estimated covariance function. Numerical studies show that the new estimator is more efficient than some existing inverse probability weighted estimators.

Keywords: Bivariate survival function; bivariate censored observations; consistency; correlated failure times; inverse probability weighted estimator; matched pairs.

1 Introduction

In some experiments each unit consists of a pair of components and life times for each component are recorded. We use $(T^{(1)}, T^{(2)})$ to denote the pair of life times. Both times are subject to random right censoring at the observed censoring time $(C^{(1)}, C^{(2)})$. Examples include twin studies, eye studies and matched pair studies, where censoring is due to units are removed from the study before failure has been observed. In such studies, the joint distribution of bivariate times need to be estimated.

Nonparametric estimators of bivariate distributions under right censoring have been proposed by Campbell (1981); Tsai *et al.* (1986); Burke (1988); Dabrowska (1988); Prentice and Cai (1992); Lin and Ying (1993); van der Laan (1996); Wang and Wells (1997); Akritas and Keilegom (2003) and Prentice *et al.* (2004). The Non-parametric MLE (MPMLE) in Campbell (1981) is not unique and is computationally intractable. The repaired NPMLE in van der Laan (1996) deals with the nonuniqueness of NPMLE based on *reduced* data, but it is sensitive to the choice of bandwidth. The repaired NPMLE can be viewed as a special case of a class of estimators in Prentice *et al.* (2004). Moodie and Prentice (2005) improve the repaired NPMLE to make it more robust to the choice of bandwidth.

¹Department of Mathematics and Statistics, Fylde College, University of Lancaster, Lancaster, LA1 4YF, UK; E-mail: h.dai@lancaster.ac.uk;

Burke (1988) introduces two inverse probability weighted (IPW) estimators which do not depend on any smoothing parameter or bandwidth, required by the repaired NPMLE and kernel estimates in Tsai *et al.* (1986); Akritas and Keilegom (2003). IPW estimators induce nonnegative probability mass and satisfy the monotonicity requirements of a distribution function. Monotonicity of a bivariate distribution estimator is very important, since the non-monotone estimators (Dabrowska, 1988; Prentice and Cai, 1992) may result in negative conditional probabilities in application. Thus this paper focuses on IPW estimators and we propose a new IPW estimator of the bivariate distribution function in the presence of right censoring. The novelty of our method is the variable transformation, which enables us to transfer the bivariate estimation problem to a univariate estimation problem and prove in a simple way that the estimator converges weakly to a zero-mean Gaussian process with an easily estimated covariance function. Numerical studies show that the new estimator performs more efficiently than the estimators in Burke (1988) and the covariance function estimator also performs remarkably well.

2 Preliminaries

Let $(T^{(1)}, T^{(2)})$ be a pair of nonnegative random variables defined on a probability space (Ω, \mathcal{F}, P) . The bivariate cumulative distribution function and bivariate survival function of this random vector are $F(t_1, t_2) = P(T^{(1)} \leq t_1, T^{(2)} \leq t_2)$ and $S(t_1, t_2) = P(T^{(1)} > t_1, T^{(2)} > t_2)$ respectively. The pair of censoring times is $(C^{(1)}, C^{(2)})$ which has survival function $G(t_1, t_2) = P(C^{(1)} > t_1, C^{(2)} > t_2)$. For simplicity, throughout this paper we assume that S and G are continuous functions. When S and G are discrete our results also hold.

The observable random variables are given by $(X^{(1)}, X^{(2)})$ and $(\delta^{(1)}, \delta^{(2)})$ where $X^{(k)} = \min(T^{(k)}, C^{(k)})$ and $\delta^{(k)} = I[T^{(k)} \leq C^{(k)}]$, $k = 1, 2$. Let

$$H(t_1, t_2) = P(X^{(1)} > t_1, X^{(2)} > t_2) \quad (1)$$

be the survival function of $(X^{(1)}, X^{(2)})$.

Throughout this paper we assume that the following assumption holds.

(A): $(T^{(1)}, T^{(2)})$ and $(C^{(1)}, C^{(2)})$ are independent.

Under assumption (A) we have that $H(t_1, t_2) = S(t_1, t_2)G(t_1, t_2)$.

Let $F^*(t_1, t_2) = P(X^{(1)} \leq t_1, \delta^{(1)} = 1, X^{(2)} \leq t_2, \delta^{(2)} = 1)$. Then it can be derived that $F^*(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} G(v_1-, v_2-) F(dv_1, dv_2)$. It follows immediately that

$$F(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} \frac{1}{G(v_1-, v_2-)} F^*(dv_1, dv_2). \quad (2)$$

Suppose that $\{(T_i^{(1)}, T_i^{(2)}, C_i^{(1)}, C_i^{(2)}), i = 1, \dots, n\}$ are i.i.d. samples of the random vector $(T^{(1)}, T^{(2)}, C^{(1)}, C^{(2)})$. The observed data are $\{(X_i^{(1)}, X_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)}), i = 1, \dots, n\}$ where $X_i^{(k)} = \min\{T_i^{(k)}, C_i^{(k)}\}$ and $\delta_i^{(k)} = I[T_i^{(k)} \leq C_i^{(k)}]$, $k = 0, 1$. Suppose that $\hat{G}(t_1, t_2)$ is an estimator of $G(t_1, t_2)$ based on the observed data. Then an IPW estimator for $F(t_1, t_2)$ is given by

$$\hat{F}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} \frac{F_n^*(dv_1, dv_2)}{\hat{G}(v_1-, v_2-)} \quad (3)$$

where $F_n^*(t_1, t_2) = \sum_{i=1}^n I[X_i^{(1)} \leq t_1, \delta_i^{(1)} = 1, X_i^{(2)} \leq t_2, \delta_i^{(2)} = 1]/n$. Note that (3) provides a class of estimators by choosing different estimators $\hat{G}(t_1, t_2)$ and the estimated distribution function $\hat{F}(t_1, t_2)$ is obviously monotone and induces non-negative probability mass. Based on different $\hat{G}(t_1, t_2)$ given in Campbell and Földes (1982), using (3), Burke (1988) constructs two IPW estimators. However, Burke (1988) does not derive their asymptotic distributions.

3 Variable transformation and methodology

3.1 Transformation of $S(t_1, t_2)$ and $G(t_1, t_2)$ to univariate functions

Define transformation from (t_1, t_2) to (α, z) as $\alpha = t_2/t_1$ and $z = \sqrt{t_1^2 + t_2^2}$. Given $\theta \in [0, \pi/2]$, such that $\cos \theta = 1/\sqrt{1 + \alpha^2}$ or $\sin \theta = 1/\sqrt{1 + \alpha^{-2}}$, we have $t_1 = z \cos \theta$ and $t_2 = z \sin \theta$. Obviously (z, θ) is the polar coordinates of (t_1, t_2) . If (t_1, t_2) is given, then α is fixed and $S(t_1, t_2)$ can be transformed to a univariate function, $S(z|\alpha)$, according to the following formula,

$$\begin{aligned} S(t_1, t_2) &= P(T^{(1)} > t_1, T^{(2)} > t_2) = P\left(\min\left\{\frac{T^{(1)}}{t_1}, \frac{T^{(2)}}{t_2}\right\} > 1\right) \\ &= P\left(\min\left\{T^{(1)}\sqrt{1 + \left(\frac{t_2}{t_1}\right)^2}, T^{(2)}\sqrt{1 + \left(\frac{t_2}{t_1}\right)^{-2}}\right\} > \sqrt{t_1^2 + t_2^2}\right) \\ &= P(Z(\alpha) > z) := S(z|\alpha), \end{aligned} \quad (4)$$

where $Z(\alpha) = \min\{T^{(1)}\sqrt{1 + \alpha^2}, T^{(2)}\sqrt{1 + \alpha^{-2}}\}$.

The above transformation can be explained by Figure 1. Points $\mathbf{p}_1 : (T^{(1)}, T^{(1)}\alpha)$ and $\mathbf{p}_2 : (T^{(2)}\alpha^{-1}, T^{(2)})$ correspond to vertical and horizontal projection of $(T^{(1)}, T^{(2)})$ onto the line $\mathbf{c} : v_2 = \alpha v_1$. The value of $Z(\alpha)$ is the minimum of the distances of these two points from the origin. Therefore $S(z|\alpha) = P(Z(\alpha) > z)$ means the survival function of $Z(\alpha)$ on line \mathbf{c} . Note that if $t_1 = 0, t_2 > 0$ then $Z(\alpha) = T^{(2)}$ and if $t_1 > 0, t_2 = 0$ then $Z(\alpha) = T^{(1)}$. Therefore the above transformation exists for all $(t_1, t_2) \in [0, \infty) \times [0, \infty)$.

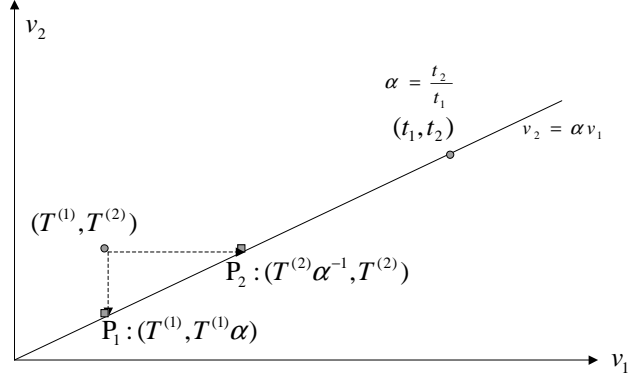


Figure 1: $(T^{(2)}\alpha^{-1}, T^{(2)})$ is obtained by horizontal projection; $(T^{(1)}, T^{(1)}\alpha)$ is obtained by vertical projection.

Similarly we have

$$G(t_1, t_2) = P(C^{(1)} > t_1, C^{(2)} > t_2) = P(Z'(\alpha) > z) := G(z|\alpha), \quad (5)$$

where $Z'(\alpha) = \min\{C^{(1)}\sqrt{1+\alpha^2}, C^{(2)}\sqrt{1+\alpha^{-2}}\}$.

Note that $S(z|\alpha)$ and $G(z|\alpha)$ are not conditional survival functions. They are univariate survival functions if α is fixed and they are bivariate functions if α is not fixed.

3.2 The IPW estimator based on transformation for censored data

In practice, due to the censorship of $(T^{(1)}, T^{(2)})$ and $(C^{(1)}, C^{(2)})$, the values of $Z(\alpha)$ and $Z'(\alpha)$ may not be obtained. We can only obtain

$$\tilde{X}^{(1)} = X^{(1)}\sqrt{1+\alpha^2}, \quad \tilde{X}^{(2)} = X^{(2)}\sqrt{1+\alpha^{-2}}. \quad (6)$$

Let

$$\begin{aligned} \tilde{Z}(\alpha) &= \min\{\tilde{X}^{(1)}, \tilde{X}^{(2)}\}, \\ \delta'(\alpha) &= (1 - \delta^{(1)})I[\tilde{X}^{(1)} < \tilde{X}^{(2)}] + (1 - \delta^{(2)})I[\tilde{X}^{(1)} > \tilde{X}^{(2)}] \\ &\quad + (1 - \min(\delta^{(1)}, \delta^{(2)}))I[\tilde{X}^{(1)} = \tilde{X}^{(2)}]. \end{aligned} \quad (7)$$

Let

$$H(z|\alpha) = P(\tilde{Z}(\alpha) > z). \quad (8)$$

Then from (6), (7) and (1) we have

$$H(z|\alpha) = P(X^{(1)} > t_1, X^{(2)} > t_2) = H(t_1, t_2). \quad (9)$$

We also have $\tilde{Z}(\alpha)$ and $\delta'(\alpha)$ are the censored value and censoring indicator for $Z'(\alpha)$, since if $\delta'(\alpha) = 1$ then $Z'(\alpha) = \tilde{Z}(\alpha)$ is observed and if $\delta'(\alpha) = 0$ then $\tilde{Z}(\alpha) = Z(\alpha) < Z'(\alpha)$ which means that $Z'(\alpha)$ is censored. Thus based on observations $\{\tilde{Z}_i(\alpha), \delta'_i(\alpha), i = 1, \dots, n\}$ obtained from transformation in (7), if we define

$$\begin{aligned} \bar{N}'(s|\alpha) &= \sum_{i=1}^n I[\tilde{Z}_i(\alpha) \leq s, \delta'_i(\alpha) = 1]/n, \\ H_n(s|\alpha) &= \sum_{i=1}^n I[\tilde{Z}_i(\alpha) > s]/n, \end{aligned} \quad (10)$$

then the KM estimator for $G(z|\alpha) = P(Z'(\alpha) > z)$ is $\hat{G}^{KM}(z|\alpha) = \prod_{s \leq z} \left[1 - \frac{\Delta \bar{N}'(s|\alpha)}{H_n(s|\alpha)}\right]$, where $\Delta \bar{N}'(s|\alpha) = \bar{N}'(s|\alpha) - \bar{N}'(s - |\alpha)$. When α is fixed, $\hat{G}^{KM}(z|\alpha)$ is a univariate KM estimator. Since $G(t_1, t_2) = G(z|\alpha)$, $\hat{G}^{KM}(z|\alpha)$ must also be a consistent estimator of $G(t_1, t_2)$.

According to (3) and $\hat{G}^{KM}(z|\alpha)$, we can construct a monotone IPW estimator of $F(t_1, t_2)$,

$$\hat{F}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} \frac{F_n^*(dv_1, dv_2)}{\hat{G}(v_1-, v_2-)} := \int_0^{t_1} \int_0^{t_2} \frac{F_n^*(dv_1, dv_2)}{\hat{G}^{KM}(s - |\alpha_v)}$$

where $s = \sqrt{v_1^2 + v_2^2}$, $\alpha_v = v_2/v_1$.

Note that the proposed IPW estimator can be written as

$$\begin{aligned} \hat{F}(t_1, t_2) &= \frac{1}{n} \sum_{i=1}^n \frac{I[X_i^{(1)} \leq t_1, X_i^{(2)} \leq t_2] \delta_i^{(1)} \delta_i^{(2)}}{\hat{G}^{KM} \left(\sqrt{(X_i^{(1)})^2 + (X_i^{(2)})^2} - |\alpha_i| \right)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{I[X_i^{(1)} \leq t_1, X_i^{(2)} \leq t_2] \delta_i^{(1)} \delta_i^{(2)}}{\hat{G}^{KM}(\tilde{Z}_i(\alpha_i) - |\alpha_i|)}, \end{aligned}$$

where $\alpha_i = X_i^{(2)}/X_i^{(1)}$ and the second equality sign is due to $\sqrt{(X_i^{(1)})^2 + (X_i^{(2)})^2} = \tilde{Z}_i(\alpha_i)$. We can see that the IPW estimator $\hat{F}(t_1, t_2)$ can be calculated by summing up all probability mass on points $(X_i^{(1)}, X_i^{(2)})$ satisfying $X_i^{(1)} \leq t_1, X_i^{(2)} \leq t_2$ and $\delta_i^{(1)} = 1, \delta_i^{(2)} = 1$. In other words,

the estimator has positive probability mass only on doubly-observed points $(X_i^{(1)}, X_i^{(2)}, \delta_i^{(1)} = 1, \delta_i^{(2)} = 1)$. The probability mass on each doubly-observed point is $1/\left[n\hat{G}^{KM}(\tilde{Z}_i(\alpha_i) - |\alpha_i|)\right]$, which can be estimated by the following steps.

- (1) Calculate $\alpha_i = X_i^{(2)}/X_i^{(1)}$.
- (2) Project all $\{X_j^{(1)}, X_j^{(2)}, j = 1, \dots, n\}$ vertically or horizontally on the line $v_2 = \alpha_i v_1$ and calculate the corresponding values $\{\tilde{Z}_j(\alpha_i), \delta'_j(\alpha_i), j = 1, \dots, n\}$.
- (3) Calculate the univariate KM estimator $\hat{G}^{KM}(z - |\alpha_i)$ based on $\{\tilde{Z}_j(\alpha_i), \delta'_j(\alpha_i), j = 1, \dots, n\}$ and $1/\left[n\hat{G}^{KM}(\tilde{Z}_i(\alpha_i) - |\alpha_i|)\right]$.

3.3 The large sample properties

Based on the following Lemma (representing $\hat{G}^{KM}(z|\alpha)$ and $1/\hat{G}^{KM}(z|\alpha)$ as sum of i.i.d. random variables) we can show the large sample properties for $\hat{F}(t_1, t_2)$.

Lemma 3.1. *Let*

$$\begin{aligned} H_0(s|\alpha) &= P(\tilde{Z}(\alpha) \leq s, \delta'(\alpha) = 1), \\ H_{0n}(s|\alpha) &= \sum_{j=1}^n I[\tilde{Z}_j(\alpha) \leq s, \delta'_j(\alpha) = 1]/n, \end{aligned} \quad (11)$$

and $\xi_i(z|\alpha) = \frac{\delta'_i(\alpha)I[\tilde{Z}_i(\alpha) \leq z]}{H(\tilde{Z}_i(\alpha)|\alpha)} - \int_0^z \frac{I[s \leq \tilde{Z}_i(\alpha)]}{H^2(s|\alpha)} dH_0(s|\alpha)$. Put $\varsigma_n = n^{-3/4}(\log n)^{3/4}$. Let τ be such that $(\tau/\sqrt{1+\alpha^2}, \tau/\sqrt{1+\alpha^{-2}}) > 0$ and $S(\tau/\sqrt{1+\alpha^2}, \tau/\sqrt{1+\alpha^{-2}}) > 0$ for any $\alpha \in [0, \infty]$. Then we have

$$\begin{aligned} (1) \quad \hat{G}^{KM}(z|\alpha) &= \frac{G(z|\alpha)}{n} \sum_{i=1}^n [1 - \xi_i(z|\alpha)] + R_n(z, \alpha) \\ (2) \quad \frac{1}{\hat{G}^{KM}(z|\alpha)} &= \frac{1}{nG(z|\alpha)} \sum_{i=1}^n [1 + \xi_i(z|\alpha)] + R_n(z, \alpha), \end{aligned}$$

and $\sup_{z \in [0, \tau], \alpha \in [0, \infty]} |R_n(z, \alpha)| = O(\varsigma_n)$, a.s. \square

Proof. See appendix A. \square

The following theorem provides the asymptotic distribution of $\hat{F}(t_1, t_2)$.

Theorem 3.1. *Let*

$$\eta_i = \frac{I[X_i^{(1)} \leq t_1, X_i^{(2)} \leq t_2] \delta_i^{(1)} \delta_i^{(2)}}{G(\tilde{Z}_i(\alpha_i)|\alpha_i)}, \quad \mu_i = \int_0^{t_1} \int_0^{t_2} [\xi_i(s - |\alpha_v|)] \frac{F^*(dv_1, dv_2)}{G(v_1, v_2)}. \quad (12)$$

For any (t_1, t_2) such that $\sqrt{t_1^2 + t_2^2} < \tau$, where τ is given in Lemma 3.1, we have $\sqrt{n}(\hat{F}(t_1, t_2) -$

$F(t_1, t_2) \Rightarrow N(0, \sigma^2(t_1, t_2))$, where $\sigma^2(t_1, t_2) = \text{Var}(\eta_1) + \text{Var}(\mu_1) + 2\text{Cov}(\eta_1, \mu_1)$. \square

Proof. See Appendix B. \square

Then it follows immediately that $\hat{F}(t_1, t_2)$ converges to $F(t_1, t_2)$ in probability.

Let $\hat{\eta}_i = \frac{I[(X_i^{(1)}, X_i^{(2)}) \leq (t_1, t_2)] \delta_i^{(1)} \delta_i^{(2)}}{\hat{G}^{KM}(\tilde{Z}_i(\alpha_i) | \alpha_i)}$, $\hat{\xi}_i(z | \alpha) = \frac{\delta_i'(\alpha) I[\tilde{Z}_i(\alpha) \leq z]}{H_n(\tilde{Z}_i(\alpha) | \alpha)} - \int_0^z \frac{I[s \leq \tilde{Z}_i(\alpha)]}{H_n^2(s | \alpha)} dH_{0n}(s | \alpha)$, and $\hat{\mu}_i = \int_{(0,0)}^{(t_1, t_2)} \left[\hat{\xi}_i(s - |\alpha_v) \right] \frac{F_n^*(dv_1, dv_2)}{\hat{G}^{KM}(s - |\alpha_v)}$. We then obtain that $\widehat{\text{Cov}}(\eta_1, \mu_1) := \sum_i \hat{\eta}_i \hat{\mu}_i / n$ and $\widehat{\text{Var}}(\mu_1) := \sum_i \hat{\mu}_i^2 / n$ are consistent estimators for $\text{Cov}(\eta_1, \mu_1)$ and $\text{Var}(\mu_1)$. In addition, according to

$$\text{Var}(\eta_1) = \int_0^{t_1} \int_0^{t_2} \frac{F^*(dv_1, dv_2)}{G^2(v_1, v_2)} - F(t_1, t_2)^2,$$

a consistent estimator for $\text{Var}(\eta_1)$ is

$$\widehat{\text{Var}}(\eta_1) = \int_0^{t_1} \int_0^{t_2} \frac{F_n^*(dv_1, dv_2)}{[\hat{G}^{KM}(s - |\alpha_v)]^2} - \left(\int_0^{t_1} \int_0^{t_2} \frac{F_n^*(dv_1, dv_2)}{\hat{G}^{KM}(s - |\alpha_v)} \right)^2.$$

Thus a consistent estimator for $\sigma^2(t_1, t_2)$ is $\hat{\sigma}^2(t_1, t_2) = \widehat{\text{Var}}(\eta_1) + \widehat{\text{Var}}(\mu_1) + 2\widehat{\text{Cov}}(\eta_1, \mu_1)$.

4 Simulation studies

In this section we study the properties of the proposed estimator via sets of 200 simulations under two different scenarios.

Scenario 1: We choose the well-known bivariate parametric model in Clayton (1978). The joint distribution of $(T^{(1)}, T^{(2)})$ is $F(t_1, t_2) = (F_1(t_1)^{-\phi} + F_2(t_2)^{-\phi} - 1)^{-\phi^{-1}}$ with $\phi = 4$. The marginal distributions $F_i(t_i), i = 1, 2$ are specified as unit exponential and $C^{(1)}, C^{(2)}$ are independent and identically distributed as $\exp(\beta)$. Simulation studies are carried out on different censoring percentages.

We compare the mean squared errors (MSE), $\sum_{k=1}^{200} (\hat{F}_k - F)^2 / 200$, of Burke's estimators and our proposed estimator. For simplicity the mean estimates and empirical variances of Burke's estimators are not provided. Comparing the MSEs of the estimators in Table 1, we can see that the proposed estimator is more efficient (has smaller MSE) than Burke's estimators, at the tail of distribution functions and under high censoring. The simulation also show that the variance estimators perform very well.

Scenario 2: Data are generated from, $T^{(1)} = 0.9\tau_1 + 0.1\tau_2$ and $T^{(2)} = 0.2\tau_1 + 0.8\tau_2$, where $\tau_1 \sim \text{Gamma}(3, 0.3)$ and $\tau_2 \sim \text{Gamma}(2, 0.3)$. The distributions of independent censoring variables $C^{(1)}$ and $C^{(2)}$ are both chosen to be $\exp(\beta)$. Such models arise in many reliability problems,

Table 1: Simulation studies: (a): theoretical survival probabilities, (b): empirical means of $\widehat{F}(x, y)$, (c): empirical variances of $\widehat{F}(x, y)$, (d): empirical means of variance estimates for $\widehat{F}(x, y)$, (M): MSEs of our proposed estimator, (M1) and (M2): MSEs of Burke's two estimators (i) $\beta = 4.0$ corresponds to 12% censoring for $T^{(1)}$ and $T^{(2)}$ respectively, (ii) $\beta = 3.0$ corresponds to 25% censoring, (iii) $\beta = 1.0$ corresponds to 50% censoring.

$n = 80$									
(x, y)	(0.5, 0.5)			(1.0, 0.5)			(1.5, 0.5)		
(a)	0.3060			0.3834			0.3926		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(b)	0.2952	0.2964	0.2928	0.3724	0.3745	0.3697	0.3824	0.3845	0.3828
(c)	0.0029	0.0031	0.0047	0.0034	0.0038	0.0064	0.0034	0.0038	0.0077
(d)	0.0029	0.0031	0.0044	0.0035	0.0037	0.0061	0.0035	0.0038	0.0071
(M)	0.0030	0.0032	0.0049	0.0035	0.0039	0.0066	0.0035	0.0039	0.0078
(M1)	0.0046	0.0046	0.0075	0.0052	0.0053	0.0100	0.0051	0.0052	0.0121
(M2)	0.0047	0.0046	0.0077	0.0053	0.0054	0.0104	0.0051	0.0052	0.0159
(x, y)	(1.0, 1.0)			(1.5, 1.0)			(1.5, 1.5)		
(a)	0.5743			0.6253			0.7414		
(b)	0.5644	0.5675	0.5623	0.6200	0.6212	0.6160	0.7363	0.7367	0.7243
(c)	0.0046	0.0053	0.0128	0.0041	0.0047	0.0150	0.0042	0.0054	0.0193
(d)	0.0042	0.0046	0.0117	0.0043	0.0049	0.0148	0.0043	0.0051	0.0251
(M)	0.0047	0.0053	0.0129	0.0041	0.0047	0.0151	0.0042	0.0054	0.0196
(M1)	0.0064	0.0067	0.0184	0.0061	0.0063	0.0261	0.0062	0.0071	0.0309
(M2)	0.0064	0.0068	0.0234	0.0062	0.0064	0.0428	0.0064	0.0075	0.0586

for example in many systems which are composed of components, the system failure time can be expressed as a sum or a linear combination of component life times. Under this scenario the proposed estimator also performs much better than Burke's estimators. For simplicity the simulation results are not shown here.

5 Discussion

This paper proposed a new IPW estimator for bivariate distribution function under right censoring. Its large sample properties are proved and it is more efficient than some existing IPW estimators. Comparing the proposed IPW estimator with other estimators, such as NPMLE and repaired NPMLE, is left as a future research work.

The Kaplan-Meier estimate for univariate data is known to be uniformly consistent over the entire support of the censored data distribution. This guarantees the univariate Kaplan-Meier estimate is reliable over the entire support of the censored data distribution. Another future research work is to study the global consistency for the bivariate estimator.

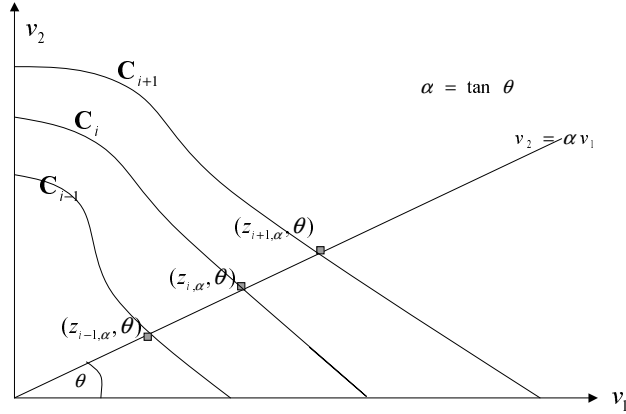


Figure 2: For all points (t_1, t_2) (with polar coordinates $(z_{i,\alpha}, \theta)$) in \mathcal{C}_i , the function $H(t_1, t_2) = H(z_{i,\alpha} | \alpha = \tan \theta) = u_i$.

A Proof of Lemma 3.1

The idea of proving Lemma 3.1 follows from Lo and Singh (1985). We first need the following lemma.

Lemma A.1. *Following the definitions of $H_n(z|\alpha)$ in (10), $H_{0n}(z|\alpha)$ in (11) and $\varsigma_n = n^{-3/4}(\log n)^{3/4}$, we have*

$$\sup_{\alpha \in [0, \infty]} \sup_{z \in [0, \tau]} \left| \int_0^z [H_n^{-1}(s|\alpha) - H^{-1}(s|\alpha)] d [H_{0n}^{-1}(s|\alpha) - H_0^{-1}(s|\alpha)] \right| = O(\varsigma_n) \text{ a.s.}$$

Proof. Let $H_n(t_1, t_2) = \sum_{i=1}^n I[X_i^{(1)} > t_1, X_i^{(2)} > t_2]/n$. Obviously we have $H_n(t_1, t_2) = H_n(z|\alpha)$, where $z = \sqrt{t_1^2 + t_2^2}$, $\alpha = t_2/t_1$.

We divide $[0, 1]$ into subintervals $[u_{i+1}, u_i]$, $i = 0, \dots, k_n$, where $k_n = O(\sqrt{n/\log n})$ and the sequence $1 = u_0 > u_1 > \dots > u_{k_n} = 0$ are such that $|u_i - u_{i-1}| \leq O(\sqrt{\log n/n})$, $i = 1, \dots, k_n$. Let $\mathcal{C}_i = \{(v_1, v_2) : \text{such that } H(v_1, v_2) = u_i\}$. Since $H(v_1, v_2)$ is continuous, \mathcal{C}_i is a continuous curve. Given α , the line $v_2 = \alpha v_1$ can be partitioned by points with polar coordinates $(z_{i,\alpha}, \theta = \arctan \alpha)$, $i = 1, \dots, k_n$, where $z_{i,\alpha}$ is such that $H(z_{i,\alpha} | \alpha) = u_i$. Note that $(z_{i,\alpha}, \theta)$ is the polar coordinates of the intersection point of line $v_2 = \alpha v_1$ and \mathcal{C}_i . See Figure 2 for details.

Following from the proof of Lemma 2 in Lo and Singh (1985), we have that for any $\alpha \in [0, \infty]$,

$$\begin{aligned}
& \left| \int_0^z [H_n^{-1}(s|\alpha) - H^{-1}(s|\alpha)] d[H_{0n}^{-1}(s|\alpha) - H_0^{-1}(s|\alpha)] \right| \\
& \leq k_n \sup_{0 \leq z \leq \tau} |H_n^{-1}(s|\alpha) - H^{-1}(s|\alpha)| \max_{0 \leq i \leq k_n - 1} |H_{0n}(z_{i+1, \alpha}|\alpha) - H_{0n}(z_{i, \alpha}|\alpha) - H_0(z_{i+1, \alpha}|\alpha) + H_0(z_{i, \alpha}|\alpha)| \\
& \quad + 2 \max_{0 \leq i \leq k_n - 1} \sup_{s \in [z_{i, \alpha}, z_{i+1, \alpha}]} |H_n^{-1}(s|\alpha) - H_n^{-1}(z_{i, \alpha}|\alpha) - H^{-1}(s|\alpha) + H^{-1}(z_{i, \alpha}|\alpha)| \\
& := A(\alpha) + B(\alpha).
\end{aligned} \tag{13}$$

Now we prove $\sup_\alpha B(\alpha) = O(\varsigma_n)$. We further partition $[u_{i+1}, u_i]$ into subintervals $[u_{i(j+1)}, u_{ij}]$, $j = 0, \dots, a_n - 1$, such that $|u_{i(j+1)} - u_{ij}| = O(n^{-3/4}(\log n)^{3/4})$ uniformly in i, j and $a_n = O(n^{1/4}(\log n)^{-1/4})$. Define $\mathcal{C}_{ij} = \{(v_1, v_2) : \text{such that } H(v_1, v_2) = u_{ij}\}$ and $z_{ij, \alpha}$ is such that $H(z_{ij, \alpha}|\alpha) = u_{ij}$. Note that $(z_{ij, \alpha}, \theta)$ is the polar coordinates of the intersection point of line $v_2 = \alpha v_1$ and \mathcal{C}_{ij} .

Since $\sup_{s, \alpha} |H_n(s|\alpha) - H(s|\alpha)|^2 = \sup_{v_1, v_2} |H_n(v_1, v_2) - H(v_1, v_2)|^2 = O((\log n)/n)$, similarly as the results in Lo and Singh (1985), we have that for a given value of α ,

$$\begin{aligned}
& \sup_{s \in [z_{i, \alpha}, z_{i+1, \alpha}]} |H_n^{-1}(s|\alpha) - H_n^{-1}(z_{i, \alpha}|\alpha) - H^{-1}(s|\alpha) + H^{-1}(z_{i, \alpha}|\alpha)| \\
& \leq \rho \sup_{s \in [z_{i, \alpha}, z_{i+1, \alpha}]} |H_n(s|\alpha) - H_n(z_{i, \alpha}|\alpha) - H(s|\alpha) + H(z_{i, \alpha}|\alpha)| + O\left(\frac{\log n}{n}\right), \text{ a.s.},
\end{aligned}$$

where ρ is a constant and does not depend on α .

We partition $[0, \pi/2]$ into subintervals $0 = \theta_0 < \theta_1 < \dots < \theta_{b_n} = \pi/2$ and let $\alpha_l = \tan \theta_l$. Let point $\mathbf{p}_{l, ij}$ with coordinates $(t_{1, l, ij}, t_{2, l, ij})$ be the intersection point of line $v_2 = \alpha_l v_1$ and curve \mathcal{C}_{ij} . A rectangle $\mathcal{R}_{l, ij}$ is given by points $\mathbf{p}_{l, ij}$, $\mathbf{p}_{l+1, ij}$ and points $\mathbf{p}_{l, ij}^* := (t_{1, l, ij}, t_{2, l+1, ij})$, $\mathbf{p}_{l+1, ij}^* := (t_{1, l+1, ij}, t_{2, l, ij})$. See Figure 3 for details. If b_n is large enough, we can choose the sequence $\theta_l, l = 0, \dots, b_n$ such that

$$\begin{aligned}
& u_{i(j-1)} > H(t_{1, l+1, ij}, t_{2, l, ij}) > u_{ij} \\
& u_{ij} > H(t_{1, l, ij}, t_{2, l+1, ij}) > u_{i(j+1)}, \quad \forall i, j, l,
\end{aligned} \tag{14}$$

which means that points $\mathbf{p}_{l, ij}^*$ and $\mathbf{p}_{l+1, ij}^*$ are between $\mathcal{C}_{i(j+1)}$ and $\mathcal{C}_{i(j-1)}$. Thus any two points (t_1, t_2) and (t_1^*, t_2^*) within the rectangle $\mathcal{R}_{l, ij}$ are such that $|H(t_1, t_2) - H(t_1^*, t_2^*)| \leq |u_{i(j+1)} - u_{i(j-1)}| = O(\varsigma_n)$, uniformly in i, j, l .

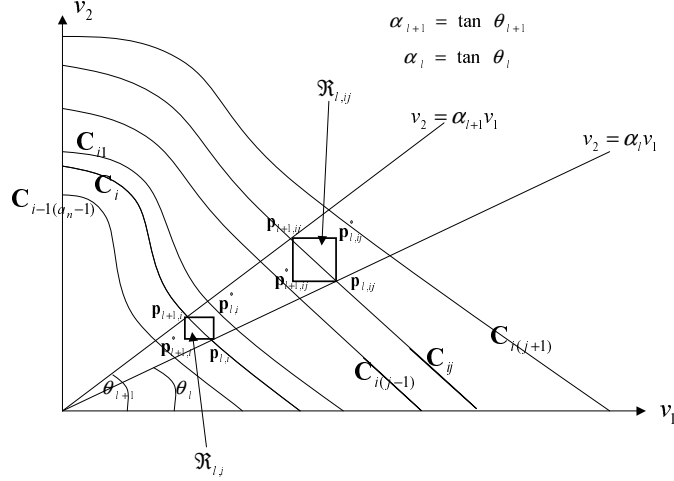


Figure 3: Partition. The lower three curves are $C_{i-1(a_n-1)}, C_i, C_{i1}$ and the upper three curves are $C_{i(j-1)}, C_{ij}, C_{i(j+1)}$. Lines $v_2 = \alpha_{l+1}v_1$ and $v_2 = \alpha_l v_1$ intersect with C_i at points $\mathbf{p}_{l+1,i}$ and $\mathbf{p}_{l,i}$, which gives a rectangle $\mathcal{R}_{l,i}$. The two lines intersect with C_{ij} at points $\mathbf{p}_{l+1,ij}$ and $\mathbf{p}_{l,ij}$, which gives a rectangle $\mathcal{R}_{l,ij}$.

With such partitions and following the results in Lo and Singh (1985), we have

$$\begin{aligned}
& \sup_{\alpha} |B(\alpha)| \tag{15} \\
& \leq \rho \max_{0 \leq l \leq b_n - 1} \sup_{\alpha \in [\alpha_l, \alpha_{l+1}]} \max_i \sup_{s \in [z_{i,\alpha}, z_{i+1,\alpha}]} |H_n(s|\alpha) - H_n(z_{i,\alpha}|\alpha) - H(s|\alpha) + H(z_{i,\alpha}|\alpha)| + O\left(\frac{\log n}{n}\right) \\
& = \rho \max_l \sup_{\alpha \in [\alpha_l, \alpha_{l+1}]} \max_{i,j} |H_n(z_{ij,\alpha}|\alpha) - H_n(z_{i,\alpha}|\alpha) - H(z_{ij,\alpha}|\alpha) + H(z_{i,\alpha}|\alpha)| + O(\varsigma_n), \text{ a.s.}
\end{aligned}$$

For points $\mathbf{p} := (t_1, t_2)$, we denote $H_n(t_1, t_2)$ and $H(t_1, t_2)$ as $H_n(\mathbf{p})$ and $H(\mathbf{p})$ for simplicity. According to the monotonicity of $H_n(t_1, t_2)$ and $H(t_1, t_2)$ we have $H_n(\mathbf{p}_{l,ij}^*) \leq H_n(z_{ij,\alpha}|\alpha) \leq H_n(\mathbf{p}_{l+1,ij}^*)$ and $H(\mathbf{p}_{l,ij}^*) \leq H(z_{ij,\alpha}|\alpha) \leq H(\mathbf{p}_{l+1,ij}^*)$, for all $\alpha \in [\alpha_l, \alpha_{l+1}]$. Then we have

$$\begin{aligned}
& \sup_{\alpha \in [\alpha_l, \alpha_{l+1}]} |H_n(z_{ij,\alpha}|\alpha) - H_n(z_{i,\alpha}|\alpha) - H(z_{ij,\alpha}|\alpha) + H(z_{i,\alpha}|\alpha)| \\
& \leq \min \{ |H_n(\mathbf{p}_{l+1,ij}^*) - H_n(\mathbf{p}_{l,i}^*) - H(\mathbf{p}_{l,ij}^*) + H(\mathbf{p}_{l+1,i}^*)|, \\
& \quad |H_n(\mathbf{p}_{l,ij}^*) - H_n(\mathbf{p}_{l+1,i}^*) - H(\mathbf{p}_{l+1,ij}^*) + H(\mathbf{p}_{l,i}^*)| \} \\
& \leq \min \{ |H_n(\mathbf{p}_{l+1,ij}^*) - H_n(\mathbf{p}_{l,i}^*) - H(\mathbf{p}_{l+1,ij}^*) + H(\mathbf{p}_{l,i}^*)|, \\
& \quad |H_n(\mathbf{p}_{l,ij}^*) - H_n(\mathbf{p}_{l+1,i}^*) - H(\mathbf{p}_{l,ij}^*) + H(\mathbf{p}_{l+1,i}^*)| \} + O(\varsigma_n), \text{ a.s.} \tag{16}
\end{aligned}$$

Using similar methods as that in Lo and Singh (1985), we can also prove

$$\begin{aligned}
& \min \{ |H_n(\mathbf{p}_{l+1,ij}^*) - H_n(\mathbf{p}_{l,i}^*) - H(\mathbf{p}_{l+1,ij}^*) + H(\mathbf{p}_{l,i}^*)|, \\
& \quad |H_n(\mathbf{p}_{l,ij}^*) - H_n(\mathbf{p}_{l+1,i}^*) - H(\mathbf{p}_{l,ij}^*) + H(\mathbf{p}_{l+1,i}^*)| \} = O(\varsigma_n), \text{ a.s.} \tag{17}
\end{aligned}$$

From (15), (16) and (17), we have $\sup_{\alpha} |B(\alpha)| = O(\varsigma_n)$. Similarly we can prove $\sup_{\alpha} A(\alpha) = O(\varsigma_n)$. The lemma then follows from (13). \square

Proof of Lemma 3.1. We have

$$\begin{aligned} & \log \hat{G}^{KM}(z|\alpha) - \log G(z|\alpha) \\ = & \left(\log \hat{G}^{KM}(z|\alpha) - \int_0^z H_n^{-1}(s|\alpha) dH_{0n}(s|\alpha) \right) + \left(\int_0^z H_n^{-1}(s|\alpha) dH_{0n}(s|\alpha) - \log G(z|\alpha) \right) \\ := & I(z, \alpha) + II(z, \alpha). \end{aligned} \quad (18)$$

Following from the proof of Theorem 1 in Lo and Singh (1985) and Lemma A.1, we have

$$\begin{aligned} & II(z, \alpha) \\ = & -\frac{\sum_j \xi_j(z|\alpha)}{n} + \int_0^z [H_n^{-1}(s|\alpha) - H^{-1}(s|\alpha)] d [H_{0n}^{-1}(s|\alpha) - H_0^{-1}(s|\alpha)] + O\left(\frac{\log n}{n}\right) \\ = & -\frac{\sum_j \xi_j(z|\alpha)}{n} + O(\varsigma_n) \quad a.s.. \end{aligned} \quad (19)$$

Now we show that

$$\sup_{z \in [0, \tau], \alpha \in [0, \infty]} |I(z, \alpha)| = \sup_{z \in [0, \tau], \alpha \in [0, \infty]} \left| \log \hat{G}^{KM}(z|\alpha) - \int_0^z H_n^{-1}(s|\alpha) dH_{0n}(s|\alpha) \right| = O(n^{-1}) \quad a.s.. \quad (20)$$

To see this note that for any value of α ,

$$\begin{aligned} & \sup_{z \in [0, \tau], \alpha \in [0, \infty]} \left| \log \hat{G}^{KM}(z|\alpha) - \int_0^z H_n^{-1}(s|\alpha) dH_{0n}(s|\alpha) \right| \\ \leq & \sup_{\alpha} \sum^* \left| \log \left(\frac{n-i}{n-i+1} \right) + \frac{1}{n} \frac{1}{\tilde{H}(\tilde{Z}_{(i)}(\alpha)|\alpha)} \right| \\ \leq & \sup_{\alpha} \sum^* \left| \log \left(1 - \frac{1}{n-i+1} \right) + \frac{1}{n} \frac{n}{n-i} \right|, \end{aligned}$$

where \sum^* extends over all i such that $\tilde{Z}_{(i)}(\alpha) \leq z$ and $\delta'_{(i)} = 1$.

Since $G(\tau/\sqrt{1+\alpha^2}, \tau/\sqrt{1+\alpha^{-2}}) > 0$ and $S(\tau/\sqrt{1+\alpha^2}, \tau/\sqrt{1+\alpha^{-2}}) > 0$ for any value of $\alpha \in [0, \infty]$, we have that for $z \leq \tau$ the integer i appearing in \sum^* is $\leq (1-\epsilon)n$, for a positive ϵ , for all large n a.s.. Then (20) follows using two term Taylor's expansion for $\log(1+w)$ with $w = -\frac{1}{n-i+1}$.

Following from (18), (19) and (20), the first result of the lemma is proved by using two term Taylor's expansion for $\log \hat{G}^{KM} - \log G$ and the second result of the lemma is proved by using two term Taylor's expansion for $\log \hat{w} - \log w$ with $\hat{w} = 1/\hat{G}^{KM}$, $w = 1/G$. \square

B Proof of Theorem 3.1

Proof. According to Lemma 3.1 we can write

$$\begin{aligned}
\hat{F}(t_1, t_2) &= \int_0^{t_1} \int_0^{t_2} \frac{F_n^*(dv_1, dv_2)}{\hat{G}^{KM}(s - |\alpha_v|)} \\
&= \int_0^{t_1} \int_0^{t_2} F_n^*(dv_1, dv_2) \left[\frac{1}{G(s - |\alpha_v|)} + \frac{1}{G(s - |\alpha_v|)} \sum_j \xi_j(s - |\alpha_v|) \right] \\
&\quad + \int_0^{t_1} \int_0^{t_2} R_n(s, \alpha_v) F_n^*(dv_1, dv_2) \\
&= \int_0^{t_1} \int_0^{t_2} F_n^*(dv_1, dv_2) \left[\frac{1}{G(s - |\alpha_v|)} + \frac{1}{G(s - |\alpha_v|)} \sum_j \xi_j(s - |\alpha_v|) \right] + O(\varsigma_n), a.s.,
\end{aligned}$$

where $s = \sqrt{v_1^2 + v_2^2}$, $\alpha_v = v_2/v_1$.

Let $\alpha_i = X_i^{(2)}/X_i^{(1)}$. For simplicity we use \tilde{Z}_i to denote $\tilde{Z}_i(\alpha_i)$. Then

$$\begin{aligned}
\hat{F}(t_1, t_2) &= \frac{1}{n} \sum_{i=1}^n \frac{I[X_i^{(1)} \leq t_1, X_i^{(2)} \leq t_2] \delta_i^{(1)} \delta_i^{(2)}}{G(\tilde{Z}_i | \alpha_i)} \\
&\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{I[X_i^{(1)} \leq t_1, X_i^{(2)} \leq t_2] \delta_i^{(1)} \delta_i^{(2)}}{G(\tilde{Z}_i | \alpha_i)} \xi_j(\tilde{Z}_i - |\alpha_i) + O(\varsigma_n)
\end{aligned}$$

From the definition of η_i in (12), we have

$$\begin{aligned}
\hat{F}(t_1, t_2) - F(t_1, t_2) &= \frac{1}{n} \sum_{i=1}^n [\eta_i - F(t_1, t_2)] + \frac{1}{n^2} \sum_{i \neq j}^n \eta_i \xi_j(\tilde{Z}_i - |\alpha_i) + O(\varsigma_n) \\
&:= \frac{1}{n} \sum_{i=1}^n [\eta_i - F(t_1, t_2)] + U_n + O(\varsigma_n).
\end{aligned}$$

Note that U_n is a U-statistic. From Serfling (1980) and $EU_n = 0$, we have

$$U_n = \sum_{k=1}^n E(U_n | X_k^{(1)}, X_k^{(2)}, \delta_k^{(1)}, \delta_k^{(2)}) + o(n^{-1}(\log n)^\gamma), \text{ for some } \gamma > 0.$$

We also obtain that for $i \neq j$, $E(\eta_i \xi_j(\tilde{Z}_i - |\alpha_i) | X_i^{(1)}, X_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)}) = 0$ and

$$\mu_j = E(\eta_i \xi_j(\tilde{Z}_i - |\alpha_i) | X_j^{(1)}, X_j^{(2)}, \delta_j^{(1)}, \delta_j^{(2)}) = \int_0^{t_1} \int_0^{t_2} [\xi_j(s - |\alpha_v|)] \frac{F^*(dv_1, dv_2)}{G(v_1, v_2)}.$$

Thus $E(U_n | X_k^{(1)}, X_k^{(2)}, \delta_k^{(1)}, \delta_k^{(2)}) = \sum_{i \neq j} E[\eta_i \xi_j(\tilde{Z}_i - |\alpha_i) | X_k^{(1)}, X_k^{(2)}, \delta_k^{(1)}, \delta_k^{(2)}] / n^2 = (n-1)\mu_k / n^2$.

Therefore we have $\hat{F}(t_1, t_2) - F(t_1, t_2) = \frac{1}{n} \sum_{i=1}^n [\eta_i - F(t_1, t_2)] + \frac{1}{n} \sum_{k=1}^n \mu_k + O(\varsigma_n)$. It follows that $\sqrt{n}(\hat{F}(t_1, t_2) - F(t_1, t_2)) \Rightarrow N(0, \sigma^2(t_1, t_2))$ with $\sigma^2(t_1, t_2) = \text{Var}(\eta_1) + \text{Var}(\mu_1) + 2\text{Cov}(\eta_1, \mu_1)$.

□

References

- Akritis M. G. and Keilegom I. V. (2003). Estimation of bivariate and marginal distributions with censored data. *Journal of Royal Statistical Society, B*, **65**: 457–471.
- Burke M.D. (1988). Estimation of a bivariate distribution function under random censorship. *Biometrika*, **75**:379-382.
- 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*, Colloq. Math. Soc. Janos Bolyai, **32**:103-121. Amsterdam: North-Holland.
- Clayton D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**:141-151.
- Dabrowska D. M. (1988). Kaplan-Meier estimate on the plane. *The Annals of Statistics*, **16**:1475-1489.
- Fleming T. R. and Harrington D. P. (1991). *Counting Processes and Survival Analysis*, John Wiley & Sons, Inc.
- Gill R. D. (1980). Censoring and stochastic integrals. *Math. Centre Tract 124, Amsterdam Math. Centrum*.
- Lin D. Y. and Ying Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, **80**(3):573-581.
- Lo S-H. and Singh K. (1985). The product-limit estimator and the bootstrap: some asymptotic representations. *Probability Theory and Related Fields*, **71**:455-465.
- Moodie F. Z. and Prentice R. L. (2005). An adjustment to improve the bivariate survivor function repaired NPMLE. *Lifetime Data Analysis*, **11**:291-307.
- Prentice R. L. and Cai J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, **79**:495–512.
- Prentice R. L. and Moodie F. Z. and Wu J. (2004). Hazard-based nonparametric survivor function estimation. *Journal of Royal Statistical Society, B*, **66**:305–319.
- Serfling R. J. (1980). *Approximation theorems of Mathematical Statistics*. John Wiley & Sons.
- Tsai W.-Y. and Leurgans S. and Crowley J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *The Annals of Statistics*, **14**:1351-1365.
- van der Laan M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Annals of Statistics*, **24**:596-627.
- Wang W. and Wells M. T. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika*, **84**:863-880.