Biometrika (1997), **84**, 4, *pp.* 863–880
Printed in Great Britain

Nonparametric estimators of the bivariate survival function under simplified censoring conditions

BY WEIJING WANG
Institute of Statistics, Academia Sinica, Taiwan
e-mail: wjwang@stat.sinica.edu.tw

AND MARTIN T. WELLS
Department of Social Statistics & Statistics Center, Cornell University, New York 14853, U.S.A.
e-mail: mtw1@cornell.edu

SUMMARY

New bivariate survival function estimators are proposed in the case where the dependence relationship between the censoring variables are modelled. Specific examples include the cases when censoring variables are univariate, mutually independent or specified by a marginal model. Large sample properties of the proposed estimators are discussed. The finite sample performance of the proposed estimators compared with other fully nonparametric estimators is studied via simulations. A real data example is given.

Some key words: Archimedean copula; Bivariate failure time data; Independent censoring; Marginal modelling; Univariate censoring.

1. INTRODUCTION

Unlike the univariate Kaplan & Meier (1958) estimator, which has the usual optimal properties, estimators of the bivariate survival function proposed in literature have some unsatisfactory features and are in general quite complex (Gill, 1992). Roughly speaking, the bivariate censoring complicates the analysis. This paper considers situations when the relationship between the censoring variables can be simplified so that estimation of the joint survival function is more direct.

Let (X_i, Y_i) $(i = 1, \ldots, n)$ be n independent and identically distributed pairs of bivariate failure times with a common joint survival function $F(x, y) = \text{pr}(X \geqslant x, Y \geqslant y)$ and let (C_{1i}, C_{2i}) $(i = 1, \ldots, n)$ be n independent and identically distributed pairs of censoring variables with a common joint survival function $G(x, y) = \text{pr}(C_1 \geqslant x, C_2 \geqslant y)$. Let $F_i(.)$ and $G_i(.)$ $(i = 1, 2)$ denote the marginal survival functions of X, Y, C_1 and C_2, respectively. If we assume right censoring, the observed variables become

$$(\tilde{X}_i, \tilde{Y}_i) = \{(X_i \wedge C_{1i}), (Y_i \wedge C_{2i})\}, \quad (\delta_i^x, \delta_i^y) = \{I(X_i < C_{1i}), I(Y_i < C_{2i})\} \quad (i = 1, \ldots, n),$$

where $\wedge$ denotes the minimum and $I(.)$ denotes the indicator function. Denote by $H(x, y) = \text{pr}(\tilde{X} \geqslant x, \tilde{Y} \geqslant y)$ the joint survival function of observables $(\tilde{X}_i, \tilde{Y}_i)$ $(i = 1, \ldots, n)$. It is usually assumed that (X_i, Y_i) are independent of (C_{1i}, C_{2i}) for all i in order to ensure identifiability (Pruitt, 1993) of the survival function. Several nonparametric estimators of $F(x, y)$ have been proposed such as those by Hanley & Parnes (1983), Tsai, Leurgans &

Crowley (1986), Dabrowska (1988), Prentice & Cai (1992) and van der Laan (1996). Without imposing further structure on the model the expressions of these estimators, and their limiting variance, are quite complex.

Consider the following decomposition of the survival function, whose validity relies on the assumption of independence between the failure times and the censoring times:

$$F(x, y) = \frac{H(x, y)}{G(x, y)}. \tag{1·1}$$

The survival function of the observables, $H(x, y)$, can be estimated by the empirical survival function,

$$\hat{H}(x, y) = \sum_{i=1}^{n} I(\tilde{X}_i \geq x, \tilde{Y}_i \geq y)/n.$$

Estimation of $G(x, y)$, in general, is dual to the estimation of $F(x, y)$ since (C_1, C_2) are also censored by (X, Y). However, it may well be that practitioners possess useful information which may simplify the analysis. For instance, when the relationship between C_1 and C_2 is known or can be modelled, the estimation of $G(x, y)$ can be simplified. Hence in general we will use

$$\hat{F}(x, y) = \frac{\hat{H}(x, y)}{\hat{G}(x, y)} \tag{1·2}$$

to estimate $F(.,.)$. The denominator will be estimated differently for each proposed censoring model. In § 2, we discuss estimation of $G(x, y)$ under various simplifications to the censoring scheme. In § 3 we discuss the large sample properties of the proposed estimators. Simulation results which demonstrate the finite sample properties are presented in § 4, and a real data example is given in § 5. The proofs of the results are sketched in the Appendix.

2. The proposed estimators under simplified censoring conditions

2·1. *Univariate censoring*

In some studies of failure times it may be reasonable to assume that censoring is univariate, that is, $C_1 \equiv C_2 \equiv C$. In the case of univariate censoring one can show that

$$\text{pr}(\tilde{X} < \tilde{Y}, \delta^x = 0, \delta^y = 1) = \text{pr}(C < Y, C < X, C \geq Y) = 0,$$

and similarly that

$$\text{pr}(\tilde{X} > \tilde{Y}, \delta^x = 1, \delta^y = 0) = \text{pr}(X > C, X \leq C, C > Y) = 0.$$

Furthermore $(\delta^x, \delta^y) = (0, 0)$ implies $\tilde{X} = \tilde{Y} = C$. The above relationships provide necessary conditions for a univariate censoring mechanism. Under univariate censoring, the monotonicity of a survival function implies

$$G(x, y) = \text{pr}(C_1 > x, C_2 > y) = G_1(x) \wedge G_2(y), \tag{2·1a}$$

where $G_1(x) = \text{pr}(C_1 > x)$ and $G_2(y) = \text{pr}(C_2 > y)$. The quantity $G_1(x) \wedge G_2(y)$ is called the upper Fréchet bound of $G(x, y)$ (Genest & MacKay, 1986). Hence, under univariate censoring, the bivariate survival function $G(x, y)$ can be estimated by

$$\hat{G}(x, y) = \hat{G}_1(x) \wedge \hat{G}_2(y), \tag{2·1b}$$

where $\hat{G}_i(.)$ is the Kaplan–Meier estimator of $G_i(.)$ $(i = 1, 2)$. When ties are absent, $\hat{G}_i(t) = \hat{H}_i(t)/\hat{F}_i(t)$ (Shorack & Wellner, 1986, p. 295), where $\hat{H}_i(.)$ $(i = 1, 2)$ are the empirical survival functions of $\tilde{X}$ and $\tilde{Y}$, respectively, and $\hat{F}_i(.)$ are the Kaplan–Meier estimators of $F_i(.)$ $(i = 1, 2)$ respectively. Lin & Ying (1993) proposed a survival function estimate under univariate censoring through the Kaplan–Meier estimator based on $\tilde{C}_i = C_i \wedge (X_i \vee Y_i) = \tilde{X}_i \vee \tilde{Y}_i$ and $\delta_i^c = 1 - \delta_i^x \delta_i^y$ $(i = 1, \ldots, n)$. The proposed estimator provides an alternative to the Lin–Ying estimator, and it will be shown in § 3·2 that our estimator has smaller asymptotic variance.

2·2. *Independent censoring*

In some case-control studies, it may be reasonable to assume that the patients in the case and control groups are censored via independent mechanisms. Hence,

$$G(x, y) = G_1(x)G_2(y), \tag{2·2a}$$

and a reasonable estimate of the bivariate survival function $G(.,.)$ is given by

$$\hat{G}(x, y) = \hat{G}_1(x)\hat{G}_2(y). \tag{2·2b}$$

The independence assumption between C_1 and C_2 can be checked by a test such as those proposed by Oakes (1982), Pons (1986) and Dabrowska (1986).

2·3. *The copula censoring model*

Suppose that there exists a known function $C_\alpha(.,.)$ such that

$$G(x, y) = C_\alpha\{G_1(x), G_2(y)\}, \tag{2·3a}$$

where $C_\alpha(.,.)$: $[0, 1] \times [0, 1] \rightarrow [0, 1]$ and α is a dependence parameter. If $G(x, y)$ can be modelled as a function of its marginals, it can be estimated by plugging estimates of the marginals, such as the Kaplan–Meier estimates, into that function. If one can find an estimator of α, denoted by $\hat{\alpha}$, then the estimate of the bivariate survival function $G(.,.)$ is given by

$$\hat{G}(x, y) = C_{\hat{\alpha}}\{\hat{G}_1(x), \hat{G}_2(y)\}. \tag{2·3b}$$

There has been substantial work in modelling the dependent relationship of two variables by separating their marginal effects (Clayton, 1978; Oakes, 1989; Genest & Rivest, 1993). The so-called 'copula' function $C_\alpha(.,.)$, which by itself is a survival function on $[0, 1] \times [0, 1]$, describes the local dependence structure. The parameter α measures global association and is related to Kendall's tau, denoted by τ, through the equation

$$\tau = 4 \int_0^1 \int_0^1 C_\alpha(u, v)C_\alpha(du, dv) - 1.$$

In the two extreme cases of $C_1 = C_2$ $(\tau = 1)$ and, where C_1 and C_2 $(\tau = 0)$ are independent, (2·3a) reduces to (2·1a) and (2·2a) respectively.

There have been several papers on estimation of α when the dependence structure $C_\alpha(.,.)$ is specified up to α but the marginal distributions remain unspecified, including the pseudo-likelihood approach of Genest, Ghoudi & Rivest (1995), and a cumulative hazard variate correlation approach of Hsu & Prentice (1996).

3. PROPERTIES OF THE PROPOSED ESTIMATORS

3·1. *General*

In this section, we discuss the consistency and weak convergence of the proposed estimators. All the limit results are as $n\to\infty$. Specifically, in the development of weak convergence results of $n^{\frac{1}{2}}\{\hat{F}(x, y)-F(x, y)\}$, by (1·1), (1·2) and consistency of $\hat{F}(.,.)$, $\hat{G}(.,.)$ and $\hat{H}(.,.)$, it follows that

$$\begin{aligned} n^{\frac{1}{2}}\{\hat{F}(x, y)-F(x, y)\} &= n^{\frac{1}{2}}\left\{\frac{\hat{H}(x, y)}{\hat{G}(x, y)}-\frac{H(x, y)}{G(x, y)}\right\} \\ &= \frac{\hat{F}(x, y)}{\hat{H}(x, y)} n^{\frac{1}{2}}\{\hat{H}(x, y)-H(x, y)\}-\frac{F(x, y)}{\hat{G}(x, y)} n^{\frac{1}{2}}\{\hat{G}(x, y)-G(x, y)\} \\ &= F(x, y)\left[\frac{1}{H(x, y)} n^{\frac{1}{2}}\{\hat{H}(x, y)-H(x, y)\}\right. \\ &\qquad \left.-\frac{1}{G(x, y)} n^{\frac{1}{2}}\{\hat{G}(x, y)-G(x, y)\}\right]+o_p(1). \end{aligned} \tag{3·1}$$

Define the following processes:

$$\begin{aligned} A(x, y) &= \frac{1}{H(x, y)} n^{\frac{1}{2}}\{\hat{H}(x, y)-H(x, y)\}, \\ B_i(z) &= -\frac{1}{H_i(z)} n^{\frac{1}{2}}\{\hat{H}_i(z)-H_i(z)\} \quad (i=1, 2), \\ C_i(z) &= \frac{1}{F_i(z)} n^{\frac{1}{2}}\{\hat{F}_i(z)-F_i(z)\} \quad (i=1, 2), \\ D_i(z) &= \frac{1}{G_i(z)} n^{\frac{1}{2}}\{\hat{G}_i(z)-G_i(z)\} \quad (i=1, 2). \end{aligned} \tag{3·2}$$

Using similar arguments as in (3·1) and by consistency of $H_i(.)$ and $F_i(.)$, we can show that

$$D_i(z)=-\{B_i(z)+C_i(z)\}+o_p(1) \quad (i=1, 2).$$

It is clear that each process in (3·2) converges weakly to a zero-mean Gaussian process. We will show that, under the simplified censoring mechanisms and some mild smoothness conditions, $n^{\frac{1}{2}}\{\hat{G}(x, y)-G(x, y)\}$ and $n^{\frac{1}{2}}\{\hat{F}(x, y)-F(x, y)\}$ can be expressed as the sum of the above processes.

3·2. *Univariate censoring*

The proposed estimator under univariate censoring is constructed by plugging (2·1b) into (1·2), that is

$$\hat{F}(x, y)=\frac{\hat{H}(x, y)}{\hat{G}_1(x)\wedge\hat{G}_2(y)}. \tag{3·3}$$

Note that, under univariate censoring, we set $\tilde{G}(.)=G_1(.)=G_2(.)$ and hence $\tilde{G}(x\vee y)=G(x, y)$. The strong consistency and weak convergence of $\hat{F}(x, y)$ under univariate censoring are established in the next two results.

THEOREM 1. *Assume $C \equiv C_1 \equiv C_2$ and that C and (X, Y) are independent. Then, for x and y such that $G(x, y) > 0$,*

$$\hat{G}(x, y) = \hat{G}_1(x) \wedge \hat{G}_2(y) \to G_1(x)$$

with probability 1 *if $x \geqslant y$, and*

$$\hat{G}(x, y) = \hat{G}_1(x) \wedge \hat{G}_2(y) \to G_2(y)$$

with probability 1 *if $x \leqslant y$. Furthermore, for x and y such that $H(x, y) = F(x, y)G(x, y) > 0$, $\hat{F}(x, y) \to F(x, y)$ with probability* 1.

One can show that, for the estimator in (3·3), for $x \geqslant y$, $G(x, y) = \tilde{G}(x) = G_1(x)$ and

$$\begin{aligned} n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\} &= F(x, y)\{A(x, y) + B_1(x) + C_1(x)\} + o_p(1) \\ &= F(x, y)\{A(x, y) - D_1(x)\} + o_p(1). \end{aligned} \tag{3·4a}$$

Similarly, for $x \leqslant y$, $G(x, y) = \tilde{G}(y) = G_2(y)$ and

$$\begin{aligned} n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\} &= F(x, y)\{A(x, y) + B_2(y) + C_2(y)\} + o_p(1) \\ &= F(x, y)\{A(x, y) - D_2(y)\} + o_p(1). \end{aligned} \tag{3·4b}$$

THEOREM 2. *Assume $C \equiv C_1 \equiv C_2$ and that C and (X, Y) are independent. Then, for x and y such that $H(x, y) = F(x, y)G(x, y) > 0$, $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$ converges weakly to a mean zero Gaussian process with asymptotic variance given by:*

(i) *for $x \geqslant y$,*

$$\sigma^2(x, y) = F^2(x, y)\left\{\frac{1}{H(x, y)} - \frac{1}{H_1(x)} + \int_0^x \frac{\Lambda_1(du)}{H_1(u)}\right\} \tag{3·5a}$$

$$= \frac{F(x, y)}{\tilde{G}(x \vee y)} - F^2(x, y)\left\{1 - \int_0^x \frac{d\tilde{G}(u)}{F_1(u)\tilde{G}^2(u)}\right\}; \tag{3·5b}$$

(ii) *for $x \leqslant y$,*

$$\sigma^2(x, y) = F^2(x, y)\left\{\frac{1}{H(x, y)} - \frac{1}{H_2(y)} + \int_0^y \frac{\Lambda_2(du)}{H_2(u)}\right\} \tag{3·6a}$$

$$= \frac{F(x, y)}{\tilde{G}(x \vee y)} - F^2(x, y)\left\{1 - \int_0^y \frac{d\tilde{G}(u)}{F_2(u)\tilde{G}^2(u)}\right\}; \tag{3·6b}$$

where $\Lambda_1(du) = -dF_1(u)/F_1(u)$ and $\Lambda_2(du) = -dF_2(u)/F_2(u)$ are marginal hazards of X and Y, respectively.

The limiting variance $\sigma^2(x, y)$ can be easily estimated by plugging in empirical estimates of each unknown component. For example, (3·5a) can be estimated by

$$\hat{\sigma}^2(x, y) = \hat{F}^2(x, y)\left[\frac{1}{\hat{H}(x, y)} - \frac{1}{\hat{H}_1(x)} + \sum_{i: \tilde{x}_i \leqslant x} \frac{n \sum_j I(\tilde{x}_j = \tilde{x}_i, \delta_j^x = 1)}{\{\sum_j I(\tilde{x}_j \geqslant \tilde{x}_i)\}^2}\right]. \tag{3·7}$$

Since each estimator in (3·7) is consistent, it is easy to see that $\hat{\sigma}^2(x, y) \to \sigma^2(x, y)$ in probability. The asymptotic covariance of $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$ involves lengthy expressions and is not very useful in practice; hence it is not presented. It will be interesting to compare $\sigma^2(x, y)$ with the asymptotic variance of the Lin & Ying (1993) estimator which is given by

$$\tilde{\sigma}^2(x, y) = \frac{F(x, y)}{\tilde{G}(x \vee y)} - F^2(x, y)\left\{1 - \int_0^{x \vee y} \frac{d\tilde{G}(u)}{\mathrm{pr}(X \vee Y \geqslant u)\tilde{G}^2(u)}\right\}. \tag{3·8}$$

It is clear to see from (3·5b) and (3·8) that variances of the two estimators only differ by one term. Since $F_i(u) \leqslant \mathrm{pr}(X \vee Y \geqslant u)$ $(i = 1, 2)$ for all u and $d\tilde{G}(.) \leqslant 0$, it is clear that $\tilde{\sigma}^2(x, y) \geqslant \sigma^2(x, y)$.

3·3. *Independent censoring*

Under the independence of the censoring time model, the proposed estimator is constructed by plugging (2·2b) into (1·2), that is

$$\hat{F}(x, y) = \frac{\hat{H}(x, y)}{\hat{G}_1(x)\hat{G}_2(y)}. \tag{3·9}$$

Each component in (3·9) is strongly consistent. It will be shown that, under the hypothesis $C_1 \perp\!\!\!\perp C_2$, $\hat{F}(x, y)$ in (3·9) is also strongly consistent.

THEOREM 3. *Assume that both C_1 and C_2 and (C_1, C_2) and (X, Y) are independent. Then, for x and y such that $G(x, y) > 0$, $\hat{G}(x, y) = \hat{G}_1(x)\hat{G}_2(y) \to G(x, y)$ with probability 1. Furthermore, for x and y such that $H(x, y) = F(x, y)G(x, y) > 0$, $\hat{F}(x, y) \to F(x, y)$ with probability 1.*

To deduce weak convergence if C_1 and C_2 are independent, using the notation in (3·2) one can write

$$\begin{aligned} n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\} &= F(x, y)\{A(x, y) + B_1(x) + C_1(x) + B_2(y) + C_2(y)\} + o_p(1) \\ &= F(x, y)\{A(x, y) - D_1(x) - D_2(y)\} + o_p(1). \end{aligned} \tag{3·10}$$

THEOREM 4. *Assume that both C_1 and C_2 and (C_1, C_2) and (X, Y) are independent. Then, for x and y such that $H(x, y) = F(x, y)G(x, y) > 0$, $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$ converges weakly to a zero-mean Gaussian process with the asymptotic variance, $\sigma^2(x, y)$, equal to*

$$\begin{aligned} F^2(x, y) \times \Bigg(&\frac{1}{H(x, y)} + \int_0^x \frac{dG_1(u)}{F_1(u)G_1^2(u)} + \int_0^y \frac{dG_2(u)}{F_2(u)G_2^2(u)} + 1 \\ &+ 2\Bigg[\frac{H(x, y)}{H_1(x)H_2(y)} + \frac{1}{H_1(x)}\int_0^y \frac{\mathrm{pr}(\tilde{X} \geqslant x, \tilde{Y} \in [v, v + dv), \delta^y = 1) - H(x, v)\Lambda_2(dv)}{H_2(v)} \\ &+ \frac{1}{H_2(y)}\int_0^x \frac{\mathrm{pr}(\tilde{X} \in [u, u + du), \delta^x = 1, \tilde{Y} \geqslant y) - H(u, y)\Lambda_1(du)}{H_1(u)} \\ &+ \int_0^y \int_0^x \frac{H(u, v)}{H_1(u)H_2(v)} \times \Bigg\{\Lambda_{11}(du, dv) - \frac{H(du, v)}{H(u, v)}\Lambda_2(dv) \\ &- \frac{H(u, dv)}{H(u, v)}\Lambda_1(du) + \Lambda_1(du)\Lambda_2(dv)\Bigg\}\Bigg]\Bigg). \end{aligned} \tag{3·11}$$

Note that the expression in (3·11) is quite complex, but each component in (3·11) is estimable, so it is easy to construct a consistent estimator by plugging in the empirical versions of the unknown functions.

3·4. *General censoring condition*

Under the marginal model in (2·3a), one can estimate $F(x, y)$ by plugging (2·3b) into (1·2), that is

$$\hat{F}(x, y) = \frac{\hat{H}(x, y)}{C_{\hat{\alpha}}\{\hat{G}_1(x), \hat{G}_2(y)\}}, \tag{3·12}$$

where $\hat{\alpha}$ is an estimator of α. As long as $C_\alpha(.,.)$ satisfies some smoothness conditions and $\hat{\alpha}$ is a reasonable estimator of α, then the estimator will be consistent and converge weakly to $G(x, y)$ at rate $n^{-\frac{1}{2}}$ and so will $\hat{F}(x, y)$. Specifically, one can write

$$\hat{G}(x, y) - G(x, y) = C_\alpha\{\hat{G}_1(x), \hat{G}_2(y)\} - C_\alpha\{G_1(x), G_2(y)\} + C_{\hat{\alpha}}\{\hat{G}_1(x), \hat{G}_2(y)\} - C_\alpha\{\hat{G}_1(x), \hat{G}_2(y)\}. \tag{3·13}$$

If we assume that $C_\alpha(.,.)$ is twice differentiable in both arguments, a Taylor expansion yields

$$\begin{aligned} &C_\alpha\{\hat{G}_1(x), \hat{G}_2(y)\} - C_\alpha\{G_1(x), G_2(y)\} \\ &= \left.\frac{\partial C_\alpha(u, v)}{\partial u}\right|_{\{G_1(x), G_2(y)\}} \{\hat{G}_1(x) - G_1(x)\} + \left.\frac{\partial C_\alpha(u, v)}{\partial v}\right|_{\{G_1(x), G_2(y)\}} \{\hat{G}_2(y) - G_2(y)\} \\ &+ \frac{1}{2}\left.\frac{\partial^2 C_\alpha(u, v)}{\partial u^2}\right|_{(u^*, v^*)} \{\hat{G}_1(x) - G_1(x)\}^2 + \frac{1}{2}\left.\frac{\partial^2 C_\alpha(u, v)}{\partial v^2}\right|_{(u^*, v^*)} \{\hat{G}_2(y) - G_2(y)\}^2 \\ &+ \left.\frac{\partial^2 C_\alpha(u, v)}{\partial u\, \partial v}\right|_{(u^*, v^*)} \{\hat{G}_1(x) - G_1(x)\}\{\hat{G}_2(y) - G_2(y)\}, \end{aligned} \tag{3·14}$$

where $(u^*, v^*) \in (0, 1) \times (0, 1)$ are intermediate values. If $C_\alpha(.,.)$ is twice differentiable at α,

$$C_{\hat{\alpha}}\{\hat{G}_1(x), \hat{G}_2(y)\} - C_\alpha\{\hat{G}_1(x), \hat{G}_2(y)\} = \frac{\partial C_\alpha(\hat{G}_1, \hat{G}_2)}{\partial \alpha}\{\hat{\alpha} - \alpha\} + \frac{1}{2}\left.\frac{\partial^2 C_\alpha(\hat{G}_1, \hat{G}_2)}{\partial \alpha^2}\right|_{\alpha^*} \{\hat{\alpha} - \alpha\}^2, \tag{3·15}$$

where α^* is an intermediate value. The following theorem states that, by strong consistency of $\hat{G}_i(.)$ $(i = 1, 2)$ and $\hat{\alpha}$ and boundedness of the derivatives, strong consistency of $\hat{G}(x, y)$ and $\hat{F}(x, y)$ can be established.

THEOREM 5. *Suppose the marginal model in* (2·3a) *holds*, $C_\alpha(.,.)$ *is twice differentiable and has bounded derivatives in the two arguments*, $\hat{G}_i(.) \to G_i(.)$ $(i = 1, 2)$ *and* $\hat{\alpha} \to \alpha$ *with probability* 1. *Then, for* x *and* y *such that* $G(x, y) > 0$,

$$\hat{G}(x, y) = C_{\hat{\alpha}}\{\hat{G}_1(x), \hat{G}_2(y)\} \to G(x, y)$$

with probability 1, *and hence, for* x *and* y *such that* $H(x, y) = F(x, y)G(x, y) > 0$, $\hat{F}(x, y) \to F(x, y)$ *with probability* 1.

To simplify the notation, define

$$K_1(u, v, \alpha) = \frac{u}{C_\alpha(u, v)} \frac{\partial C_\alpha(u, v)}{\partial u}, \quad K_2(u, v, \alpha) = \frac{v}{C_\alpha(u, v)} \frac{\partial C_\alpha(u, v)}{\partial v},$$
$$K_3(u, v, \alpha) = \frac{1}{C_\alpha(u, v)} \frac{\partial C_\alpha(u, v)}{\partial \alpha}. \tag{3·16}$$

By (3·14) and (3·16), one can write

$$\frac{1}{G(x, y)} n^{\frac{1}{2}}[C_\alpha\{\hat{G}_1(x), \hat{G}_2(y)\} - C_\alpha\{G_1(x), G_2(y)\}] = K_1\{G_1(x), G_2(y), \alpha\}D_1(x) + K_2\{G_1(x), G_2(y), \alpha\}D_2(y) + R_{1n}. \tag{3·17a}$$

Similarly, one can write

$$\frac{1}{G(x,y)} n^{\frac{1}{2}}[C_{\hat{\alpha}}\{\hat{G}_1(x),\hat{G}_2(y)\} - C_\alpha\{\hat{G}_1(x),\hat{G}_2(y)\}] = K_3\{\hat{G}_1(x),\hat{G}_2(y),\alpha\}n^{\frac{1}{2}}(\hat{\alpha}-\alpha) + R_{2n}. \tag{3·17b}$$

It will be shown that the remainder terms, R_{1n} and R_{2n}, are both $o_p(1)$. It follows that

$$\frac{1}{G(x, y)} n^{\frac{1}{2}}\{\hat{G}(x, y) - G(x, y)\} = K_1\{G_1(x), G_2(y), \alpha\}D_1(x) + K_2\{G_1(x), G_2(y), \alpha\}D_2(y) + K_3\{G_1(x), G_2(y), \alpha\}n^{\frac{1}{2}}(\hat{\alpha} - \alpha) + o_p(1). \tag{3·18}$$

Hence, by (3·1),

$$n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\} = F(x, y)[A(x, y) - K_1\{G_1(x), G_2(y), \alpha\}D_1(x) - K_2\{G_1(x), G_2(y), \alpha\}D_2(y) - K_3\{G_1(x), G_2(y), \alpha\}n^{\frac{1}{2}}(\hat{\alpha} - \alpha)] + o_p(1). \tag{3·19}$$

Note that, under univariate censoring,

$$C(u, v) = u \wedge v = \{(u + v) - |u - v|\}/2,$$

which is continuous but not differentiable at $u = v$. However, under the independence of C_1 and C_2, all the models reduce to $C(u, v) = uv$, which satisfies the differentiability condition. In this case, $K_1(u, v) = K_2(u, v) = 1$. Thus Theorem 4 is a special case of the following more general result.

THEOREM 6. *Suppose the marginal model* (2·3a) *holds and*
(i) $C_\alpha(.,.) \in \mathscr{C}^2$,
(ii) C_α *is twice differentiable at* α *and both derivatives are bounded.*
(iii) $n^{\frac{1}{2}}(\hat{\alpha} - \alpha) = O_p(1)$.
Then, for x *and* y *such that* $H(x, y) = F(x, y)G(x, y) > 0$, $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$ *converges weakly to a zero-mean Gaussian process. When* α *is known, the asymptotic variance of*

$n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$ *is*

$$\sigma^2(x, y) = F^2(x, y)\left(\frac{1}{H(x, y)} - 1 + K_1^2\{G_1(x), G_2(y), \alpha\}\left\{1 - \int_0^x \frac{dG_1(u)}{F_1(u)G_1^2(u)}\right\}\right.$$

$$+ K_2^2\{G_1(x), G_2(y), \alpha\}\left\{1 - \int_0^y \frac{dG_2(u)}{F_2(u)G_2^2(u)}\right\}$$

$$+ 2K_1\{G_1(x), G_2(y), \alpha\}\int_0^x \frac{dG_1(u)}{F_1(u)G_1^2(u)}$$

$$+ 2K_2\{G_1(x), G_2(y), \alpha\}\int_0^y \frac{dG_2(u)}{F_2(u)G_2^2(u)}$$

$$+ 2K_1\{G_1(x), G_2(y), \alpha\}K_2\{G_1(x), G_2(y), \alpha\}$$

$$\times\left[\frac{H(x, y)}{H_1(x)H_2(y)} - 1 + \frac{1}{H_1(x)}\int_0^y \frac{\text{pr}(\tilde{X} \geqslant x, \tilde{Y} \in [v, v+dv), \delta^y = 1) - H(x, v)\Lambda_2(dv)}{H_2(v)}\right.$$

$$+ \frac{1}{H_2(y)}\int_0^x \frac{\text{pr}(\tilde{X} \in [u, u+du), \delta^x = 1, \tilde{Y} \geqslant y) - H(u, y)\Lambda_1(du)}{H_1(u)}$$

$$+ \int_0^y\int_0^x \frac{H(u, v)}{H_1(u)H_2(v)}\left\{\Lambda_{11}(du, dv) - \frac{H(du, v)}{H(u, v)}\Lambda_2(dv)\right.$$

$$\left.\left.\left. - \frac{H(u, dv)}{H(u, v)}\Lambda_1(du) + \Lambda_1(du)\Lambda_2(dv)\right\}\right]\right). \tag{3·20}$$

When α is unknown, from (3·19) it can be seen that $\sigma^2(x, y)$ contains extra terms involving the variance of $n^{\frac{1}{2}}(\hat{\alpha} - \alpha)$ and its covariances with $A(x, y)$ and $D_i(.)$ $(i = 1, 2)$ which, however, may be difficult to derive even when the form of $\hat{\alpha}$ is given. In this case, $\sigma^2(x, y)$ can be estimated using the bootstrap.

Generally speaking, the proposed estimator in (3·12) imposes a semiparametric structure on (C_1, C_2) and hence is subject to model misspecification and the error produced by $\hat{\alpha}$. The robustness of the proposed estimator will be studied through simulations.

4. Simulation results

A series of 1000 simulations were carried out for comparing the finite sample $(n = 60, 250)$ performance of different estimators of $F(x, y)$, namely the Dabrowska (1988) estimator, the Prentice & Cai (1992) estimator, the Lin & Ying (1993) estimator and the proposed simplified estimators under different censoring mechanisms. In the simulations, (X, Y) and (C_1, C_2) are both generated by the Clayton (1978) family using the algorithm discussed in Prentice & Cai (1992). The dependence structure of Clayton's family is of the form

$$C_\alpha(u, v) = (u^{1-\alpha} + v^{1-\alpha} - 1)^{-1/(\alpha-1)} \quad (u, v \in (0, 1)), \tag{4·1a}$$

where

$$\tau = (\alpha - 1)/(\alpha + 1). \tag{4·1b}$$

As $\alpha \to \infty$, $\tau \to 1$ and as $\alpha \to 1$, $\tau \to 0$. We will denote by $\tau_{x,y}$ and τ_{c_1,c_2} the Kendall's tau measures of (X, Y) and (C_1, C_2), respectively. In each simulation run, marginal censoring rates are around 30–40%.

Table 1 presents the simulation results under univariate censorship for $\tau_{x,y} = 0{\cdot}67$. Four estimators are compared based on the average bias and standard deviation on 9 points. Relative performance of the four estimators is quite consistent regardless of $\tau_{x,y}$ and the sample size. The Dabrowska and Prentice–Cai estimators perform very similarly and have slightly smaller variation than the two simple estimators specified for univariate censored data. The proposed estimator in (3·3) in general outperforms the Lin–Ying estimator in terms of both accuracy and precision. However, the differences in bias and standard

Table 1. *Simulation summary statistics for estimators of bivariate survival probabilities under univariate censoring and* $(X, Y) \sim$ Clayton $(\tau_{x,y} = 0{\cdot}67)$ *based on* 1000 *samples*

		$n=250$			$n=60$		
y		$x=0{\cdot}092$	$x=0{\cdot}336$	$x=0{\cdot}829$	$x=0{\cdot}092$	$x=0{\cdot}336$	$x=0{\cdot}829$
0·092	True	0·854	0·697	0·456	0·854	0·697	0·456
	Dab.	−1·35	−3·30	−2·26	−2·80	−5·46	−2·26
		(2·27)	(2·99)	(3·56)	(4·56)	(6·03)	(7·09)
	P.-C.	−1·35	−3·31	−2·28	−2·81	−5·52	−2·36
		(2·27)	(2·99)	(3·56)	(4·56)	(6·03)	(7·09)
	L.-Y.	−1·61	−3·97	−3·41	−3·26	−7·30	−3·80
		(2·31)	(3·11)	(3·66)	(4·58)	(6·18)	(7·30)
	(3·3)	−0·34	−3·31	−2·28	−1·30	−5·83	−2·63
		(2·29)	(3·02)	(3·56)	(4·54)	(6·06)	(7·08)
0·336	True	0·697	0·625	0·427	0·697	0·625	0·427
	Dab.	−1·55	−2·56	−2·31	−3·50	−5·38	−3·14
		(2·86)	(3·09)	(3·50)	(6·26)	(6·47)	(7·07)
	P.-C.	−1·57	−2·59	−2·39	−3·55	−5·50	−3·43
		(2·86)	(3·09)	(3·50)	(6·26)	(6·47)	(7·07)
	L.-Y.	−2·13	−2·93	−3·41	−4·77	−5·87	−4·31
		(2·96)	(3·20)	(3·64)	(6·39)	(6·64)	(7·34)
	(3·3)	−1·78	0·46	−2·40	−3·54	0·73	−3·15
		(2·92)	(3·14)	(3·55)	(6·32)	(6·55)	(7·12)
0·861	True	0·456	0·427	0·369	0·456	0·427	0·369
	Dab.	−2·53	−2·77	−1·78	−3·63	−4·52	−2·93
		(3·46)	(3·42)	(3·40)	(7·29)	(7·12)	(7·01)
	P.-C.	−2·55	−2·85	−1·91	−3·72	−4·80	−2·50
		(3·46)	(3·42)	(3·40)	(7·29)	(7·12)	(7·02)
	L.-Y.	−3·27	−3·65	−2·41	−6·51	−7·22	−2·38
		(3·56)	(3·57)	(3·56)	(7·48)	(7·40)	(7·34)
	(3·3)	−2·78	−3·17	2·12	−3·60	−4·39	6·98
		(3·48)	(3·48)	(3·48)	(7·32)	(7·24)	(7·22)

True, true survival probability; Dab., Dabrowska estimator; P.-C., Prentice–Cai estimator; L.-Y., Lin–Ying estimator; (3·3), estimator proposed in (3·3).
In each cell, top figure is average bias ($\times 10^3$) of estimate; bottom figure, in parentheses, is standard deviation ($\times 10^2$) of estimate.

deviation among all the estimators are no greater than the order of 10^{-3}. We found these conclusions also held for $\tau_{x,y}=2$. Table 2 compares the proposed estimator in (3·9) and the two fully nonparametric estimators under independent censorship ($\tau_{c_1,c_2}=0$) and $\tau_{x,y}=0{\cdot}67$. The proposed estimator behaves roughly as the Dabrowska and Prentice–Cai estimators at most points but seems to be more variable in the tail region.

Table 3 examines the performance of the proposed estimator in (3·12) for $n=60$, when (C_1, C_2) is generated from Clayton's family with moderate correlation ($\tau_{c_1,c_2}=0{\cdot}5$) but fitted by three different semiparametric models, namely Clayton's family, Frank's family (Genest, 1987) and the log-copula family (Genest & Rivest, 1993). The dependence structure for Frank's family is of the form

$$C_\alpha(u, v)=\log_\alpha\left\{1+\frac{(\alpha^u-1)(\alpha^v-1)}{\alpha-1}\right\}, \tag{4·2a}$$

where $\log_\alpha$ denotes the logarithm with base α ($a>0$), which is related to Kendall's tau by

$$\tau=4\gamma^{-1}\{D_1(\gamma)-1\}, \tag{4·2b}$$

Table 2. *Simulation summary statistics for estimators of bivariate survival probabilities under independent censoring and* $(X, Y)\sim$ Clayton ($\tau_{x,y}=0{\cdot}67$) *based on* 1000 *samples*

			$n=250$			$n=60$	
y		$x=0{\cdot}092$	$x=0{\cdot}336$	$x=0{\cdot}829$	$x=0{\cdot}092$	$x=0{\cdot}336$	$x=0{\cdot}829$
0·092	True	0·854	0·697	0·456	0·854	0·697	0·456
	Dab.	0·25	−2·18	0·68	−0·07	−1·06	1·17
		(2·35)	(3·14)	(3·72)	(4·69)	(6·21)	(7·04)
	P.-C.	0·25	−2·20	0·24	−0·08	−1·13	1·01
		(2·35)	(3·14)	(3·72)	(4·69)	(6·21)	(7·04)
	(3·9)	0·26	−2·16	0·29	−0·07	0·32	0·69
		(2·42)	(3·30)	(3·90)	(4·85)	(6·60)	(7·36)
0·336	True	0·697	0·625	0·427	0·697	0·625	0·427
	Dab.	−0·57	−2·39	−0·09	0·59	−0·10	1·58
		(3·06)	(3·29)	(3·68)	(6·07)	(6·45)	(6·99)
	P.-C.	−0·58	−2·48	−0·26	0·51	−0·52	0·62
		(3·06)	(3·30)	(3·69)	(6·07)	(6·47)	(6·98)
	(3·9)	−0·64	−2·19	−0·08	0·84	0·27	0·22
		(3·24)	(3·76)	(4·18)	(6·33)	(7·40)	(7·99)
0·829	True	0·456	0·427	0·369	0·456	0·427	0·369
	Dab.	−1·12	−1·64	0·04	−3·70	−4·21	−1·04
		(3·62)	(3·53)	(3·42)	(7·11)	(6·80)	(6·98)
	P.-C.	−1·16	−1·86	−0·71	−3·83	−4·97	−3·91
		(3·62)	(3·54)	(3·43)	(7·11)	(6·81)	(6·89)
	(3·9)	−1·58	−1·89	−0·67	−3·98	−4·42	−2·45
		(3·75)	(4·06)	(4·45)	(7·40)	(7·84)	(9·05)

True, true survival probability; Dab., Dabrowska estimator; P.-C., Prentice–Cai estimator proposed in (3·9).
In each cell, top figure is average bias ($\times 10^3$) of estimate; bottom figure, in parentheses, is standard deviation ($\times 10^2$) of estimate.

Table 3. *Robustness study for the estimator proposed in* (3·12) *when* $(X, Y) \sim$ Clayton $(\tau_{x,y} = 0{\cdot}67)$ *and* $(C_1, C_2) \sim$ Clayton $(\tau_{c_1,c_2} = 0{\cdot}5)$ *but* (C_1, C_2) *is fitted by different models without estimated parameters*; $n = 60$

y		$x = 0{\cdot}092$	$x = 0{\cdot}336$	$x = 0{\cdot}829$
0·092	True	0·854	0·698	0·346
	$\hat{F}_{\alpha}^{\text{Cln}}$	−2·72 (4·90)	−3·02 (6·32)	−2·90 (7·32)
	$\hat{F}_{\alpha}^{\text{lc}}$	−0·50 (4·79)	−1·18 (6·44)	−3·32 (7·46)
	$\hat{F}_{\alpha}^{\text{Frk}}$	−2·24 (4·78)	−4·49 (6·41)	−5·42 (7·44)
	$\hat{F}_{\hat{\alpha}}^{\text{Cln}}$	1·79 (4·90)	−0·61 (6·32)	−0·22 (7·34)
	$\hat{F}_{\hat{\alpha}}^{\text{lc}}$	−0·48 (4·79)	−0·97 (6·43)	−2·84 (7·41)
	$\hat{F}_{\hat{\alpha}}^{\text{Frk}}$	−3·85 (6·10)	−4·86 (6·76)	−4·79 (7·39)
0·336	True	0·697	0·625	0·427
	$\hat{F}_{\alpha}^{\text{Cln}}$	−4·83 (6·51)	−4·91 (7·11)	−3·46 (7·73)
	$\hat{F}_{\alpha}^{\text{lc}}$	−0·04 (6·19)	−0·33 (6·96)	−0·08 (7·79)
	$\hat{F}_{\alpha}^{\text{Frk}}$	−3·30 (6·16)	−8·27 (6·87)	−6·72 (7·67)
	$\hat{F}_{\hat{\alpha}}^{\text{Cln}}$	−2·45 (6·49)	1·99 (7·09)	5·50 (7·80)
	$\hat{F}_{\hat{\alpha}}^{\text{lc}}$	0·24 (6·19)	0·50 (6·88)	−4·59 (7·74)
	$\hat{F}_{\hat{\alpha}}^{\text{Frk}}$	−3·38 (6·54)	−7·37 (7·08)	1·56 (7·62)
0·829	True	0·346	0·427	0·369
	$\hat{F}_{\alpha}^{\text{Cln}}$	−1·52 (7·37)	−0·80 (7·82)	−1·43 (8·19)
	$\hat{F}_{\alpha}^{\text{lc}}$	1·51 (7·21)	3·95 (7·54)	6·26 (8·33)
	$\hat{F}_{\alpha}^{\text{Frk}}$	−0·58 (7·18)	−2·73 (7·43)	−3·11 (8·12)
	$\hat{F}_{\hat{\alpha}}^{\text{Cln}}$	1·14 (7·38)	8·20 (7·87)	14·68 (8·29)
	$\hat{F}_{\hat{\alpha}}^{\text{lc}}$	2·02 (7·20)	5·54 (7·47)	9·53 (8·31)
	$\hat{F}_{\hat{\alpha}}^{\text{Frk}}$	1·11 (7·16)	−0·58 (7·35)	0·49 (8·02)

True, true survival probability; Cln, Clayton; lc, log-copula; Frk, Frank.
In each cell, first figure is average bias ($\times 10^3$) of estimate; second figure, in parentheses, is standard deviation ($\times 10^2$) of estimate.

where $\gamma = -\log \alpha$ and $D_1(.)$ is the Debye function of order 1 defined by

$$D_1(\gamma) = \frac{1}{\gamma} \int_0^{\gamma} \frac{t}{e^t - 1}\, dt.$$

The log-copula family in general has two parameters with the dependence structure

$$C_\alpha(u, v) = \exp(\alpha\gamma[1 - \{(1 - \log u)^{\alpha+1} + (1 - \log v)^{\alpha+1} - 1\}^{1/(\alpha+1)}]). \qquad (4{\cdot}3\text{a})$$

Letting $\alpha\gamma = 1$, we have

$$\tau = \int_0^{\infty} \frac{\exp(-2t)}{(1+t)^{\alpha}}\, dt. \qquad (4{\cdot}3\text{b})$$

Recall that the estimator in (3·12) has the following general expression:

$$\hat{F}(x, y) = \frac{\hat{H}(x, y)}{C_{\hat{\alpha}}\{\hat{G}_1(x), \hat{G}_2(y)\}}.$$

Denote by $\hat{F}_\alpha^{\mathrm{Cln}}$ the estimator when $C_\alpha(.,.)$ is fitted by Clayton's family and α is the value corresponding to the true value of τ_{c_1,c_2}. If $\tau = 0{\cdot}5$ in equation (4·1b), $\alpha = 3$. Similarly, denote by $\hat{F}_\alpha^{\mathrm{Frk}}$ the estimator in (3·12) by fitting Frank's family with $\alpha = 5{\cdot}8$, which is obtained by inverting equation (4·2b). Now $\hat{F}_\alpha^{\mathrm{lc}}$ is the proposed estimator by fitting the log-copula family with $\alpha = 3{\cdot}3$. When the value of τ_{c_1,c_2} is unknown, it is estimated using the method by Brown, Hollander & Korwar (1974), which adjusts the relative concordance/discordance probabilities of (C_1, C_2) by the Kaplan–Meier estimators of $G_i(.)$ $(i = 1, 2)$. Once an estimate of τ_{c_1,c_2} is obtained, the formulae in (4·1b), (4·2b) and (4·3b) can be inverted to estimate the corresponding values of α, denoted by $\hat{\alpha}$. In principle, $\hat{F}_\alpha^{\mathrm{Cln}}$ can be thought of as the possibly best candidate for (3·12) since $C_\alpha(.,.)$ is fitted by the true dependence structure and the true parameter value. Also an estimate of α obtained through estimating Kendall's tau, even under consistency, often produces large variation (Maguluri, 1993). We used this method to estimate α in the simulations since it is easy to compute and also it may be interesting to see what is the effect of such a poor estimate of a nuisance parameter on the performance of $\hat{F}(x, y)$. Table 3 shows that all the estimators perform reasonably well even when the model is misspecified. The effect of estimating α seems to vary for different models. Estimator $\hat{F}_{\hat{\alpha}}^{\mathrm{Cln}}$ has much poorer performance than that of $\hat{F}_\alpha^{\mathrm{Cln}}$ in the tail region but is fine at other points. The differences between $\hat{F}_{\hat{\alpha}}$ and $\hat{F}_\alpha$ for Frank's model and the log-copula model are less obvious. In the case with $n = 250$ the biases were roughly the same but the variances were about half the size of those with $n = 60$.

5. A real data example

In this section, we study the well-known dataset discussed in Holt & Prentice (1974) and Lin & Ying (1993). The dataset consists of 11 paired survival times, in days, of closely

Table 4. *Survival days of skin grafts on burn patients*

(*a*) *Raw data*

	Patient i										
	1	2	3	4	5	6	7	8	9	10	11
$\tilde{X}_i$	37	19	57+	93	16	22	20	18	63	29	60+
$\tilde{Y}_i$	29	13	15	26	11	17	26	21	43	15	40

(*b*) *Estimation of* $\tilde{G}(x \vee y)$ *using the method proposed in* (2·1b)

	$x \vee y$										
	37	19	57	93	16	22	26	21	63	29	60
$\hat{G}_1(x)$	1	1	1	0·5	1	1	1	1	0·5	1	0·75
$\hat{G}_2(y)$	1	1	1	1	1	1	1	1	1	1	1
$\hat{G}_1 \wedge \hat{G}_2$	1	1	1	0·5	1	1	1	1	0·5	1	0·75

(*c*) *Estimation of* $\hat{G}(x \vee y)$ *using the Lin–Ying approach*

	$x \vee y$										
	37	19	57	93	16	22	26	21	63	29	60
δ^c	0	0	1	0	0	0	0	0	0	0	1
$\hat{G}_{\mathrm{LY}}(.)$	1	1	1	0·5	1	1	1	1	0·5	1	0·75

$\tilde{X}_i$, close match survival time; $\tilde{Y}_i$, poor match survival time.
$\hat{G}_1(x) \wedge \hat{G}_2(y)$, proposed estimator of $\tilde{G}(x \vee y)$.
$\hat{G}_{\mathrm{LY}}(x \vee y)$, Lin–Ying estimator of $\tilde{G}(x \vee y)$.

and poorly matched skin grafts on the same burn patient. This is a perfect example of univariate censorship. From the raw data listed in Table 4(a), it can be seen that, for the two singly censored data points, (57+, 15) and (60+, 40), the censored component must be greater than the observed component. Tables 4(b) and (c) show how $\tilde{G}(x \vee y)$ is estimated using the method proposed in (2·1b) and the Lin–Ying estimator, respectively. For this dataset, $\delta_i^y = 1$ $(i = 1, \ldots, n)$ and hence $\hat{G}_2(y) = 1$ for all y, which implies that $\hat{G}_1(x) \wedge \hat{G}_2(y) = \hat{G}_1(x)$ for all x. Also note that $\delta_i^c = 1 - \delta_i^x \delta_i^y = 1 - \delta_i^x$ $(i = 1, \ldots, n)$, that δ^c determines the censoring status of $\tilde{C} = \tilde{X} \vee \tilde{Y}$ using the Lin–Ying estimator of $\tilde{G}(x \vee y)$ and that $1 - \delta^x$ determines the censoring status for computing $\hat{G}_1(x)$. Thus the two estimators coincide with each other in this example.

Acknowledgement

This paper benefited from discussions with Bruce Turnbull and Louis-Paul Rivest. The support of the U.S. National Institute of Health and National Science Foundation is gratefully acknowledged.

Appendix

Proofs

In the following proofs, $(\Omega, \mathscr{F}, \mathscr{P})$ denotes the underlying probability space. Strong consistency results for the marginal quantities $\hat{H}_i(.)$, $\hat{F}_i(.)$ and $\hat{G}_i(.)$ $(i = 1, 2)$ will be used. Since the univariate censoring structure does not satisfy the differentiability condition on $C(.,.)$, Theorem 1 is proved separately. Theorem 3 is a special case of Theorem 5 and thus only the proof of the latter is provided. The weak convergence derivations in Theorems 2, 4 and 6 are also presented here.

Proof of Theorem 1. Consider the case that $x \geqslant y$; by monotonicity of the survival function, $G_1(x) \leqslant G_2(y)$. If $G_1(x) = G_2(y)$, then, by strong consistency of $\hat{G}_i(.)$ $(i = 1, 2)$, $\hat{G}_1(x) = \hat{G}_2(y)$ with probability 1. Consider another case when $x > y$ and $G_1(x) < G_2(y)$. We would like to show that $\hat{G}(x, y) = \hat{G}_1(x) \wedge \hat{G}_2(y) \to G_1(x)$ with probability 1. Note that, for any $x > y$ with $G_1(x) < G_2(y)$ and $\omega \in \Omega$,

$$|\hat{G}(x, y) - G(x, y)| = |\hat{G}_1(x, \omega) - G_1(x)| I\{\hat{G}_1(x, \omega) \leqslant \hat{G}_2(y, \omega)\} + |\hat{G}_2(y, \omega) - G_1(x)| I\{\hat{G}_1(x, \omega) > \hat{G}_2(y, \omega)\}, \quad (A·1)$$

where $\hat{G}_i(t, \omega)$ denotes the value of the random function $\hat{G}_i(.)$ evaluated at t for the outcome $\omega \in \Omega$. We show that, for any $x > y$ with $G_1(x) < G_2(y)$,

$$\mathrm{pr}[\limsup \{\hat{G}_1(x, \omega) > \hat{G}_2(y, \omega)\}] = 0.$$

Letting $G_2(y) - G_1(x) = \delta > 0$, we get

$$\{\omega: \hat{G}_1(x, \omega) > \hat{G}_2(y, \omega)\} \subseteq \{\omega: |\hat{G}_1(x, \omega) - G_1(x)| + |\hat{G}_2(y, \omega) - G_2(y)| > \delta\}. \quad (A·2)$$

It can be shown that, for two random variables $X(\omega)$ and $Y(\omega)$ and any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\{\omega: |X(\omega)| + |Y(\omega)| \geqslant \varepsilon\} \subseteq \{\omega: |X(\omega)| \geqslant \delta\} \cup \{\omega: |Y(\omega)| \geqslant \delta\}.$$

Hence we can conclude that there exists $\delta^* > 0$ such that

$$\{\omega: \hat{G}_1(x, \omega) > \hat{G}_2(y, \omega)\} \subseteq \{\omega: |\hat{G}_1(x, \omega) - G_1(x)| \geqslant \delta^*\} \cup \{\omega: |\hat{G}_2(y, \omega) - G_2(y)| \geqslant \delta^*\}. \quad (A·3)$$

Taking lim sup in (A·3) and applying a property of the lim sup function (Billingsley, 1985, p. 60), we get that there exists $\delta^* > 0$ such that

$$\limsup\{\omega: \hat{G}_1(x,\omega) > \hat{G}_2(y,\omega)\}$$
$$\subseteq \limsup\{\omega: |\hat{G}_1(x,\omega) - G_1(x)| \geqslant \delta^*\} \cup \limsup\{\omega: |\hat{G}_2(y,\omega) - G_2(y)| \geqslant \delta^*\}.$$

Strong consistency of $\hat{G}_i(.)$ $(i = 1, 2)$ leads to $\mathrm{pr}[\limsup\{\omega: \hat{G}_1(x,\omega) > \hat{G}_2(y,\omega)\}] = 0$. Hence by (A·1) and monotonicity it follows that, for $x > y$,

$$|\{\hat{G}_1(x) \wedge \hat{G}_2(y)\} - G_1(x)| = |\hat{G}_1(x) - G_1(x)| \quad (A{\cdot}4)$$

with probability 1. By strong consistency of $\hat{G}_1(.)$, we can show that, for any $\varepsilon > 0$,

$$\mathrm{pr}\left[\sup_{x \geqslant y} |\{\hat{G}_1(x) \wedge \hat{G}_2(y)\} - G_1(x)| > \varepsilon\right] \leqslant \mathrm{pr}\left\{\sup_x |\hat{G}_1(x,\omega) - G_1(x)| > \varepsilon\right\} = 0.$$

Similar arguments can be established for the case when $x \leqslant y$. □

Proof of Theorems 3 *and* 5. Since Theorem 3 is a special case of Theorem 5, we shall prove the more general result. Consider the expression in (3·13). We first show that

$$C_\alpha\{\hat{G}_1(x), \hat{G}_2(y)\} \to C_\alpha\{G_1(x), G_2(y)\}$$

with probability 1.

Since $C_\alpha(.,.)$ has continuous first and second derivatives in both arguments and is defined on a compact set, the derivatives are bounded. Specifically, there exist $M_i < \infty$ $(i = 1, 2, 3, 4, 5)$ such that

$$\left|\frac{\partial C_\alpha(u,v)}{\partial u}\right| \leqslant M_1, \quad \left|\frac{\partial C_\alpha(u,v)}{\partial v}\right| \leqslant M_2, \quad \frac{1}{2}\left|\frac{\partial^2 C_\alpha(u,v)}{\partial u^2}\right| \leqslant M_3,$$
$$\frac{1}{2}\left|\frac{\partial^2 C_\alpha(u,v)}{\partial v^2}\right| \leqslant M_4, \quad \left|\frac{\partial^2 C_\alpha(u,v)}{\partial u\,\partial v}\right| \leqslant M_5.$$

By the Taylor expansion in (3·14), for any x, y and $\omega \in \Omega$, it follows that

$$\sup_{x,y} |\hat{G}(x,y;\omega) - G(x,y)| \leqslant M_1 \sup_x |\hat{G}_1(x,\omega) - G_1(x)| + M_2 \sup_y |\hat{G}_2(y,\omega) - G_2(y)|$$
$$+ M_3 \sup_x |\hat{G}_1(x,\omega) - G_1(x)|^2 + M_4 \sup_y |\hat{G}_2(y,\omega) - G_2(y)|^2$$
$$+ M_5 \sup_{x,y} |\hat{G}_1(x,\omega) - G_1(x)|\,|\hat{G}_2(y,\omega) - G_2(y)|. \quad (A{\cdot}5)$$

By similar arguments as in the previous proof and taking the lim sup, one can show, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\mathrm{pr}\left\{\omega: \sup_{x,y} |\hat{G}(x,y;\omega) - G(x,y)| \geqslant \varepsilon\right\}$$
$$\leqslant \mathrm{pr}\left\{\omega: \sup_x |\hat{G}_1(x,\omega) - G_1(x)| \geqslant \delta/M_1\right\} + \mathrm{pr}\left\{\omega: \sup_y |\hat{G}_2(y,\omega) - G_2(y)| \geqslant \delta/M_2\right\}$$
$$+ \mathrm{pr}\left\{\omega: \sup_x |\hat{G}_1(x,\omega) - G_1(x)|^2 \geqslant \delta/M_3\right\} + \mathrm{pr}\left\{\omega: \sup_y |\hat{G}_2(y,\omega) - G_2(y)|^2 \geqslant \delta/M_4\right\}$$
$$+ \mathrm{pr}\left\{\omega: \sup_{x,y} |\hat{G}_1(x,\omega) - G_1(x)|\,|\hat{G}_2(y,\omega) - G_2(y)| \geqslant \delta/M_5\right\}. \quad (A{\cdot}6)$$

Since each M_i is bounded and thus $\delta/M_i > 0$ for $\delta > 0$, each term on the right-hand side of (A·6) has probability zero. Hence

$$\mathrm{pr}\left\{\omega: \sup_{x,y} |\hat{G}(x,y;\omega) - G(x,y)| \geqslant \varepsilon\right\} = 0$$

and $C_\alpha\{\hat{G}_1(x), \hat{G}_2(y)\} \to C_\alpha\{G_1(x), G_2(y)\}$ with probability 1. By (3·15), we can show that, if $\partial C_\alpha/\partial\alpha$ and $\partial^2 C_\alpha/\partial\alpha^2$ are bounded and $\hat{\alpha} \to \alpha$ with probability one, then

$$C_{\hat{\alpha}}\{\hat{G}_1(x), \hat{G}_2(y)\} - C_\alpha\{\hat{G}_1(x), \hat{G}_2(y)\} = 0$$

with probability 1. Therefore $\hat{G}(x, y) \to G(x, y)$ with probability 1. By (1·2) it is easy to show that $\hat{F}(x, y) \to F(x, y)$ with probability 1. □

Proof of Theorems 2, 4 *and* 6 *on weak convergence of* $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$. Each process in (3·2) converges to a mean zero Gaussian process. By (3·4), (3·10) and (3·19), $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$ can be expressed in terms of the sum of these processes and thus will converge to a mean zero Gaussian process on the bivariate Càdlàg space on $[0, t_1] \times [0, t_2]$, where (t_1, t_2) is such that $H(t_1, t_2) > 0$. To derive its asymptotic variance in each case, the following expressions are useful:

$$\begin{aligned}
n^{\frac{1}{2}}\{\hat{H}(x, y) - H(x, y)\} &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \{I(\tilde{X}_i \geq x, \tilde{Y}_i \geq y) - H(x, y)\}, \\
n^{\frac{1}{2}}\{\hat{H}_1(x) - H_1(x)\} &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \{I(\tilde{X}_i \geq x) - H_1(x)\}, \\
n^{\frac{1}{2}}\{\hat{H}_2(y) - H_2(y)\} &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \{I(\tilde{Y}_i \geq y) - H_2(y)\}, \\
n^{\frac{1}{2}}\{\hat{F}_1(x) - F_1(x)\} &= -n^{-\frac{1}{2}} F_1(x) \left\{\sum_{i=1}^{n} \int_0^x \frac{dM_{1i}(u)}{H_1(u)}\right\} + o_p(1), \\
n^{\frac{1}{2}}\{\hat{F}_2(y) - F_2(y)\} &= -n^{-\frac{1}{2}} F_2(y) \left\{\sum_{i=1}^{n} \int_0^y \frac{dM_{2i}(u)}{H_2(u)}\right\} + o_p(1),
\end{aligned} \tag{A·7}$$

where

$$\begin{aligned}
dM_{2i}(u) &= dI(\tilde{X}_i \leq u, \delta_i^x = 1) - I(\tilde{X} \geq u)\Lambda_1(du), \\
dM_{2i}(u) &= dI(\tilde{Y}_i \leq u, \delta_i^y = 1) - I(\tilde{Y} \geq u)\Lambda_2(du).
\end{aligned} \tag{A·8}$$

Denote by avar(.) the asymptotic variance of (.). By elementary probability arguments and some known results on the univariate Kaplan–Meier estimator, it can be shown that

$$\operatorname{avar}\{A(x, y)\} = \frac{1}{H(x, y)} - 1, \quad \operatorname{avar}\{B_i(t)\} = \frac{1}{H_i(t)} \quad (i = 1, 2),$$

$$\operatorname{avar}\{C_i(t)\} = \int_0^t \frac{\Lambda_i(du)}{H_i(u)}, \quad \operatorname{acov}\{B_i(t), C_i(t)\} = -\int_0^t \frac{\Lambda_i(du)}{H_i(u)} \quad (i = 1, 2),$$

$$\operatorname{acov}\{A(x, y), B_1(x)\} = -\frac{1}{H_1(x)} + 1, \quad \operatorname{acov}\{A(x, y), B_2(y)\} = -\frac{1}{H_2(y)} + 1,$$

$$\operatorname{acov}\{A(x, y), C_1(x)\} = \int_0^x \frac{\Lambda_1(du)}{H_1(u)}, \quad \operatorname{acov}\{A(x, y), C_2(y)\} = \int_0^y \frac{\Lambda_2(du)}{H_2(u)},$$

$$\operatorname{acov}\{B_1(x), B_2(y)\} = \frac{H(x, y)}{H_1(x)H_2(y)} - 1,$$

$$\operatorname{acov}\{B_1(x), C_2(y)\} = \frac{1}{H_1(x)} \int_0^y \frac{\operatorname{pr}\{\tilde{X} \geq x, \tilde{Y} \in [v, v + dv), \delta^y = 1\} - H(x, v)\Lambda_2(dv)}{H_2(v)},$$

$$\operatorname{acov}\{B_2(y), C_1(x)\} = \frac{1}{H_2(y)} \int_0^x \frac{\operatorname{pr}\{\tilde{X} \in [u, u + du), \delta^x = 1, \tilde{Y} \geq y\} - H(u, y)\Lambda_1(du)}{H_1(u)},$$

$$\operatorname{acov}\{C_1(x), C_2(y)\} = \int_0^y \int_0^x \frac{H(u,v)}{H_1(u)H_2(v)} \left\{ \Lambda_{11}(du,dv) - \frac{H(du,v)}{H(u,v)} \Lambda_2(dv) - \frac{H(u,dv)}{H(u,v)} \Lambda_1(du) + \Lambda_1(du)\Lambda_2(dv) \right\}.$$

Also, by integration by parts, one can derive the following useful identities:

$$\int_0^z \frac{\Lambda_i(du)}{H_i(u)} = -\int_0^z \frac{dF_i(u)}{F_i^2(u)G_i(u)} = \frac{1}{H_i(z)} - 1 + \int_0^z \frac{dG_i(u)}{F_i(u)G_i^2(u)} \quad (i=1,2). \tag{A·9}$$

Hence

$$\operatorname{avar}\{D_i(z)\} = 1 - \int_0^z \frac{dG_i(u)}{F_i(u)G_i^2(u)}, \quad \operatorname{acov}\{A, D_i(z)\} = -\int_0^z \frac{dG_i(u)}{F_i(u)G_i^2(u)} \quad (i=1,2). \tag{A·10}$$

The identities in (A·9) and (A·10) are then used to deduce (3·5b) and (3·6b). □

References

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: John Wiley.

Billingsley, P. (1985). *Probability and Measures*, 2nd ed. New York: John Wiley.

Brown, B. W., Hollander, M. & Korwar, R. M. (1974). Nonparametric tests of independence for censored data, with applications to heart transplant studies. In *Reliability and Biometry: Statistical Analysis of Lifelength*, Ed. F. Proschan and R. J. Serfling, pp. 327–54. Philadelphia: SIAM.

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**, 141–51.

Dabrowska, D. M. (1986). Rank tests for independence for bivariate censored data. *Ann. Statist.* **14**, 250–64.

Dabrowska, D. M. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16**, 1475–89.

Genest, G., Ghoudi, K. & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–52.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74**, 549–55.

Genest, C. & MacKay, J. (1986). The joy of copulas: bivariate distributions with uniform marginal. *Am. Statistician* **40**, 280–3.

Genest, C. & Rivest, L. P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Am. Statist. Assoc.* **88**, 1034–43.

Gill, R. D. (1992). Multivariate survival analysis. *Theory Prob. Applic.* **37**, 18–31.

Hanley, J. A. & Parnes, M. N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrika* **39**, 129–39.

Hsu, L. & Prentice, R. L. (1996). On assessing the strength of dependency between failure time variates. *Biometrika* **83**, 491–507.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations, I, II. *J. Am. Statist. Assoc.* **53**, 457–81, 562–3.

Lin, D. Y. & Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80**, 573–81.

Magaluri, G. (1993). Semiparametric estimation of association in a bivariate survival model. *Ann. Statist.* **21**, 1648–62.

Oakes, D. (1982). A concordance test for independence in the presence of censoring. *Biometrics* **28**, 451–55.

Oakes, D. (1989). Bivariate survival models induced by frailty. *J. Am. Statist. Assoc.* **84**, 487–93.

Pons, O. (1986). A test of independence between two censored survival times. *Scand. J. Statist.* **13**, 173–85.

Prentice, R. L. & Cai, J. (1992). Covariance and survival function estimation using censored multivariate failure time data. *Biometrika* **79**, 495–512.

Pruitt, R. C. (1993). Identifiability of bivariate survival curves from censored data. *J. Am. Statist. Assoc.* **88**, 573–9.

Shorack, G. R. & Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* New York: Wiley.

TSAI, W. Y., LEURGANS, S. & CROWLEY, J. J. (1986). Nonparametric estimation of a bivariate survival function in presence of censoring. *Ann. Statist.* **14**, 1351–65.
VAN DER LAAN, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.* **24**, 596–627.

[*Received January* 1996. *Revised February* 1997]