Biometrika (1998), **85**, 3, *pp.* 561–572
Printed in Great Britain

Nonparametric estimation of successive duration times under dependent censoring

BY WEIJING WANG
Institute of Statistics, Academia Sinica, Taipei, Taiwan
wjwang@stat.sinica.edu.tw

AND MARTIN T. WELLS
Department of Social Statistics, Cornell University, Ithaca, New York 14853, U.S.A.
mtw1@cornell.edu

SUMMARY

This paper considers nonparametric inference for duration times of two successive events. Since the second duration process becomes observable only if the first event has occurred, the length of the first duration affects the probability of the second duration being censored. Dependent censoring arises if the two duration times are correlated, which is often the case. Standard approaches to this problem fail because of dependent censoring mechanism. A new product-limit estimator for the second duration variable and a path-dependent joint survival function estimator are proposed, both modified for the dependent censoring. Properties of the estimators are discussed. An example from Lawless (1982) is studied for illustrative purposes as well as a simulation study.

Some key words: Auxiliary endpoints; Bivariate survival data; Dependent censoring; Informative censoring; Product-limit estimator; Serial failure events.

1. INTRODUCTION

In the analysis of censored lifetime data it is often assumed that the failure time variables are independent of the nuisance censoring variables to ensure identifiability of the marginal survival function. This assumption is usually referred to as 'non-informative censoring' and has been implicitly used to construct most survival function estimators, such as the well-known Kaplan & Meier (1958) estimator and estimators of a bivariate survival function proposed by Campbell & Földes (1982), Tsai, Leurgans & Crowley (1986), Dabrowska (1988), Prentice & Cai (1992), Lin & Ying (1993) and Tsai & Crowley (1988), for instance.

In this paper we investigate a situation in which the independent censoring assumption is not plausible. Consider the duration times of two successive events, which occur in a particular order. Such a sampling scheme is very useful for describing the evolution of a multistage disease process or a process of recurrent events. For example, it is known that the development of AIDS evolves from the HIV incubation period to the period of clinical AIDS. The joint behaviour, especially their association, of the two duration variables is often of interest. However, censoring can occur to both duration processes as well as to their sum. The second duration process becomes observable only if the first event has occurred. The longer the first duration, the greater is the chance that the second duration time will be censored. If the two duration variables are correlated, the second duration time is censored by a dependent variable related to the first duration time. We consider nonparametric estimation of the joint survival function of the two duration times under this dependent censoring mechanism.

In the development of the proposed method, the information on the first duration is used to weight individual observations to 'unbias' the effect of dependent censoring. There has been recent interest in improved estimation of the survival function using auxiliary information from disease markers. Papers by Dabrowska (1987), Jewell & Kabfleisch (1992), Longini et al. (1989), Malani, Redfearn & Nielsen (1992), Robins (1992), Robins & Rotnizky (1992), Fleming et al. (1994), Gary (1994) and Malani (1995) give results in this direction. Disease progression can also be described by a multistage compartment model. Frydman (1995) considers modelling and estimation of these transition probabilities. Strong assumptions such as Markov or semi-Markov structures on transition probabilities and the duration distributions are imposed.

In § 2 we discuss the characteristics of the dependent censoring mechanism for successive events and investigate the effect of censoring on inference. In § 3 we propose a path-dependent estimator of the joint survival function. Alternative estimators are also discussed. Asymptotic properties of the proposed estimator are discussed in § 4. In § 5 an example adapted from Lawless (1982, p. 477) is analysed to illustrate the proposed ideas. Simulation results are presented in § 6. Visser (1996) studied the same problem of dependent censoring but assumed that the duration and censoring variables are discrete; the proposed method can handle a more general situation.

2. NOTATION AND THE EFFECT OF CENSORING

Let (X, Y) be the duration times of two consecutive and adjacent events with the joint survival function, let $F(x, y) = \text{pr}(X > x, Y > y)$ and let $F_i(.)$ $(i = 1, 2)$ be the marginal survival functions of X and Y, respectively. Both X and Y are subject to right censoring. Figure 1 indicates that C_1 actually governs the censoring for two univariate processes, namely X and $X + Y$. Throughout the paper it is assumed that C_1 is independent of both X and Y. Therefore the marginal and joint survival functions of X, $Z = X + Y$, and (X, Z) can be estimated by standard methods by assuming independent censoring. Note that Y is censored by $C_2 = (C_1 - X)I(X \leqslant C_1)$, which is possibly correlated with Y if (X, Y) are correlated. The observed variables are $(\tilde{X}, \tilde{Y}, \delta_x, \delta_y)$, where $\tilde{X} = X \wedge C_1$, $\tilde{Y} = Y \wedge C_2$, $\delta_x = I(X \leqslant C_1)$ and $\delta_y = I(Y \leqslant C_2)$. It can be seen that

Case 1, $(\delta_x, \delta_y) = (0, 0)$ if $C_1 < X$,
Case 2, $(\delta_x, \delta_y) = (1, 0)$ if $X \leqslant C_1 < X + Y$,
Case 3, $(\delta_x, \delta_y) = (1, 1)$ if $C_1 \geqslant X + Y$.

Denote the observed sample by $S := \{(\tilde{X}_i, \tilde{Y}_i, \delta_{x_i}, \delta_{y_i}, i = 1, \ldots, n\}$ which consists of n identically and independently distributed realisations from $(\tilde{X}, \tilde{Y}, \delta_x, \delta_y)$. Note that, if

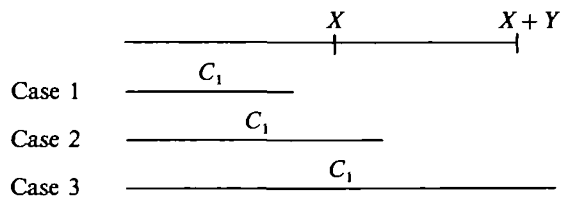

Fig. 1. Censoring conditions for successive duration times.

$\delta_{x_i}=0$, then $\delta_{y_i}=0$ and $\tilde{Y}_i=0$, which implies that no information about Y_i is available. Consider a subset of S, denoted by S^*, consisting of observations in Case 2 and Case 3. In S^*, $\delta_{x_i}=1$ and Y_i is subject to censoring by $C_{2i}=C_{1i}-x_i>0$, where x_i is the observed value of X_i and can be treated as a covariate. We will see that x_i affects the probability of the corresponding Y_i being censored. Since $F_2(y)=F(0, y)$, we shall discuss the effect of informative censoring on the estimation of $F(x, y)$.

There exist several methods for estimating $F(x, y)$ nonparametrically. The most common approach is to estimate $F(x, y)$ via estimable components. The independent censoring assumption plays a crucial role in making these components identifiable and can simplify the estimation. For example, consider the following path-dependent decomposition of $F(x, y)$:

$$\begin{aligned} F(x, y) &= \mathrm{pr}(X>x, Y>y)=\mathrm{pr}(Y>y \mid X>x)\,\mathrm{pr}(X>x) \\ &= \prod_{v \leqslant y}\{1-\Lambda_{Y|X>x}(dv)\}F_1(x), \end{aligned} \tag{1}$$

where $\Lambda_{Y|X>x}(y)$ is the cumulative conditional hazard of Y given $X>x$. The marginal $F_1(x)$ can be estimated by the product limit method based on $(\tilde{X}_i, \delta_{x_i})$ $(i=1, \ldots, n)$. Assuming that $(X, Y) \perp (C_1, C_2)$, Campbell & Földes (1982) proposed the following estimator of $\Lambda_{Y|X>x}(dv)$:

$$\hat{\Lambda}_{Y|X>x}(\Delta v)=\frac{\sum_{i=1}^{n} I(\tilde{X}_i>x, \tilde{Y}_i=v, \delta_{y_i}=1)}{\sum_{i=1}^{n} I(\tilde{X}_i>x, \tilde{Y}_i \geqslant v)}. \tag{2}$$

Note that, as $x=0$, (2) reduces to the Nelson–Aalen estimator of $\Lambda_Y(dv)$, where $\Lambda_Y(.)$ is the cumulative hazard of Y, and the Campbell–Földes estimator reduces to the Kaplan–Meier estimator of $F_2(y)$. Let $R(v \mid x)$ be the risk set of Y at time v given $X>x$. If $(X, Y) \perp (C_1, C_2)$, each observation in $R(v \mid x)$ is subject to the same censoring effect which can be cancelled in calculating $\Lambda^{\mathrm{CF}}_{Y|X>x}(\Delta v)$. In other words, if censoring is non-informative, observations are censored regardless of the survival status and hence $R(v \mid x)$, at each x and v, is still homogeneous so $\Lambda^{\mathrm{CF}}_{Y|X>x}(\Delta v)$ is a reasonable estimate of $\Lambda_{Y|X>x}(dv)$. Under the dependent censoring structure described earlier, it is easy to see that

$$\Lambda^{\mathrm{CF}}_{Y|X>x}(\Delta v) \rightarrow \frac{\mathrm{pr}(X>x, Y \in dv, C_1>X+v)}{\mathrm{pr}(X>x, Y \geqslant v, C_1>X+v)}=\frac{\int_x^{\infty} G_1(u+v)F(du, dv)}{\int_x^{\infty} -G_1(u+v)F(du, v-)},$$

in probability. Hence the Campbell–Földes estimator of $F(x, y)$ and the corresponding Kaplan–Meier estimator of $F_2(y)$ are not consistent. Similar arguments can be applied to other bivariate estimators of $F(x, y)$ assuming independent censoring.

3. Nonparametric estimation of $F(x, y)$

3·1. *The recommended estimator*

When X and Y occur consecutively, the path which describes the trajectory of (X, Y) on the plane can be determined by their relative ordering. Therefore the path-dependent decomposition discussed earlier is a natural choice for the problem. We now propose a modified estimate of $\Lambda_{Y|X>x}(dv)$. Previous analysis indicates that, under the dependent censoring structure, the risk set $R(v \mid x)$ for estimating $\Lambda_{Y|X>x}(dv)$ may not be homogeneous. Note that, for $v>0$, $R(v \mid x) \subset S^*$. An observation i with the first duration observed to be

x_i is included in $R(v|x)$ $(v>0)$ with probability

$$\begin{aligned}\mathrm{pr}\{i\in R(v|x)\} &= \mathrm{pr}(\tilde{X}_i\in x_i, x_i>x, \delta_{x_i}=1, \tilde{Y}_1\geqslant v)\\ &= \mathrm{pr}(X\in x_i, x_i>x, Y\geqslant v)\,\mathrm{pr}(C_1>x_i+v),\end{aligned} \tag{3}$$

Where $X\in x$ is the abbreviation of $X\in(x, x+\Delta)$ as $\Delta\to 0$. Equation (3) implies that x_i affects the probability of the corresponding Y_i being included in $R(v|x)$. To adjust the heterogeneity, one can weight each observation in $R(v|x)$ by an estimate of the reciprocal of its including probability, namely $1/G_1(x_i+v)$. Note that, since C_1 is censored independently by $X+Y$, $G_1(.)$ can be estimated by the Kaplan–Meier estimator based on $(\tilde{X}_i+\tilde{Y}_i, 1-\delta_{x_i}\delta_{x_i})$ $(i=1,\ldots,n)$. Hence the proposed estimator of $\Lambda_{Y|X>x}(dy)$ is given by

$$\begin{aligned}\hat{\Lambda}_{Y|X>x}(\Delta v) &= \frac{\sum_{i\in R_Y(v|x)} I(\tilde{Y}_i=v, \delta_{y_i}=1)/\hat{G}_1(x_i+v)}{\sum_{i\in R_Y(v|x)} I(\tilde{Y}_i\geqslant v)/\hat{G}_1(x_i+v)}\\ &= \frac{\sum_{i=1}^n I(\tilde{X}_i>x, \delta_{x_i}=1, \tilde{Y}_i=v, \delta_{y_i}=1)/\hat{G}_1(\tilde{X}_i+v)}{\sum_{i=1}^n I(\tilde{X}_i>x, \delta_{x_i}=1, \tilde{Y}_i\geqslant v)/\hat{G}_1(\tilde{X}_i+v)},\end{aligned} \tag{4}$$

where $\hat{G}_1(.)$ is the Kaplan–Meier estimator of $G_1(.)$. If we plug (4) into (1), the proposed path-dependent estimator is given by

$$\hat{F}(x, y) = \prod_{v\leqslant y}\{1-\hat{\Lambda}_{Y|X>x}(\Delta v)\}\hat{F}_1(x), \tag{5}$$

where $\hat{F}_1(x)$ is the Kaplan–Meier estimator of $F_1(x)$. The marginal survival function of Y can be estimated by $\hat{F}(0, y)$.

A potential problem with $\hat{F}(x, y)$ exists if $\hat{G}_1(.)=0$. Let $\tilde{X}_{(n)}+\tilde{Y}_{(n)}$ be the observed largest value of $(\tilde{X}_i+\tilde{Y}_i)$ $(i=1,\ldots,n)$ and let $(\delta_{x_{(n)}}, \delta_{y_{(n)}})$ be the corresponding indicators. When $y\geqslant\tilde{Y}_{(n)}$, one has to compute $\hat{\Lambda}_{Y|X>x}(\Delta\tilde{Y}_{(n)})$, which involves calculating $1/\hat{G}_1(\tilde{X}_{(n)}+\tilde{Y}_{(n)})$ if $\tilde{X}_{(n)}>x$. When C_1 is observed at $\tilde{X}_{(n)}+\tilde{Y}_{(n)}$, that is $\delta_{x_{(n)}}\delta_{y_{(n)}}=0$, $\hat{G}_1(\tilde{X}_{(n)}+\tilde{Y}_{(n)})=0$. Note that, in this case, the numerator in (4) becomes

$$I(\tilde{X}_{(n)}>x, \delta_{x_{(n)}}=1, \tilde{Y}_i=\tilde{Y}_{(n)}, \delta_{y_{(n)}}=1)=0.$$

By the convention that $0/0=0$, we can set $\hat{\Lambda}_{Y|X>x}(\Delta\tilde{Y}_{(n)})=0$ when $\delta_{x_{(n)}}\delta_{y_{(n)}}=0$.

3·2. *Alternative estimators*

Recall that X and $Z=X+Y$ are both subject to censoring by C_1, independent of both X and Z. Therefore the joint survival function of (X, Z), denoted by $F_{X,Z}(x, z)=\mathrm{pr}(X>x, Z>z)$, can be estimated by standard procedures such as those of Dabrowska (1988) and Prentice & Cai (1992) or Lin & Ying (1993) and Tsai & Crowley (1998) in the case of univariate censoring. Let $\hat{F}_{X,Z}(.,.)$ be an estimator of $F_{X,Z}(.,.)$. It is possible to estimate $F(x, y)$ by transforming $\hat{F}_{X,Z}(.,.)$. Specifically,

$$\hat{F}^{\#}(x, y)=\sum_{u>x}\sum_{z>u+y}\hat{F}_{X,Z}(\Delta u, \Delta z), \tag{6}$$

where

$$\hat{F}_{X,Z}(\Delta x, \Delta z)=\hat{F}_{X,Z}(x, z)-\hat{F}_{X,Z}(x, z-)-\hat{F}_{X,Z}(x-, z)+\hat{F}_{X,Z}(x-, z-)$$

is the estimated mass of (X, Z) at (x, z). In general, $\hat{F}^{\#}(x, y)$ does not have an explicit formula. A possibly more serious problem is the negative mass problem that occurs with most bivariate survival estimators, including those mentioned earlier. It is not clear how $\hat{F}^{\#}(x, y)$ is affected if $\hat{F}_{X,Z}(\Delta u, \Delta z)<0$ for some (u, v).

In principle, the weighting idea can be applied to any decomposition expression of $F(x, y)$ or other functions of interest in which the effect of censoring can be measured. For example, define $D(x, y) = \text{pr}(X \leqslant x, Y \leqslant y)$ to be the distribution function of (X, Y), which can be expressed as

$$D(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{\tilde{H}_{11}(du, dv)}{G_1(u+v)}, \tag{7}$$

where $\tilde{H}_{11}(u, v) = \text{pr}(\tilde{X} \leqslant u, \tilde{Y} \leqslant v, \delta_x = \delta_y = 1)$ is a sub-distribution function. Burke (1988) proposes an estimate of $\tilde{F}(x, y)$ based on (7) given by

$$\hat{D}(x, y) = \sum_{i=1}^{n} \frac{I(\tilde{X}_i \leqslant x, \tilde{Y}_i \leqslant y, \delta_{x_i} = \delta_{y_i} = 1)}{n\hat{G}_1(\tilde{X}_i + \tilde{Y}_i)}, \tag{8}$$

and an estimate of $F(x, y)$ is given by

$$\hat{F}(x, y) = 1 - \hat{D}(x, \infty) - \hat{D}(\infty, y) + \hat{D}(x, y).$$

Note that the Burke estimator only uses data with $\delta_x = \delta_y = 1$, Case 3.

Visser (1996) suggests estimating $F(x, y)$ by estimating $\Lambda_{Y|X=x}(dv)$, which is related to $F(x, y)$ through the identity

$$F(x, y) = -\int_{u>x} \text{pr}(Y > y \mid X = u) F_1(du) = -\int_{u>x} \prod_{v \leqslant v} \{1 - \lambda_{Y|X=u}(dv)\} F_1(du), \tag{10}$$

where $F_1(du) = F_1(u) - F_1(u-)$. Note that, in some applications, $\text{pr}(Y > y \mid X = x)$, which assesses the effect of X on the survival probability of Y, is the quantity of interest. The main advantage of estimating $\Lambda_{Y|X=x}(dv)$ is that

$$\lambda_{Y|X=u}(dv) = \text{pr}(Y \in dv \mid Y \geqslant v, X = u) = \text{pr}(\tilde{Y} \in dv, \delta_y = 1 \mid \tilde{X} = u, \delta_x = 1, \tilde{Y} \geqslant v). \tag{11}$$

If X is discrete, $\Lambda_{Y|X=x}(dv)$ can be estimated by a ratio of empirical functions:

$$\hat{\Lambda}^0_{Y|X=x}(dy) = \frac{\sum_{i=1}^{n} I(\tilde{X}_i = x, \delta_{x_i} = 1, \tilde{Y}_i = y, \delta_{y_i} = 1)}{\sum_{i=1}^{n} I(\tilde{X}_i = x, \delta_{x_i} = 1, \tilde{Y}_i \geqslant y)}. \tag{12}$$

Hence $F(x, y)$ can be estimated by plugging $\hat{\Lambda}^0_{Y|X=x}(dy)$ into (10), giving $\hat{F}^0(x, y)$, say. With additional assumptions on the discreteness of Y and C_1, several analytical properties of $\hat{\Lambda}^0_{Y|X=x}(dy)$ and $\hat{F}^0(x, y)$, such as their asymptotic variances, have been deduced. However, when X is continuous, estimating $\Lambda_{Y|X=x}(dv)$ requires special smoothing techniques and can be very complicated when the dependent censoring condition is taken into account; see Dabrowska (1987) for references in this direction.

If X is discrete or grouped, so that $\text{pr}(X = u) > 0$, one can alternatively apply the weighting technique to estimate $\text{pr}(X = u, Y > y)$ directly. Since

$$F(x, y) = \sum_{u>x} \text{pr}(X = u, Y > y),$$

$F(x, y)$ can be estimated by

$$\hat{F}^{**}(x, y) = \sum_{u>x} \sum_{i=1}^{n} \frac{I(\tilde{X}_i = u, \delta_{x_i} = 1, \tilde{Y}_i > y)}{n\hat{G}_1(u+y)} = \sum_{i=1}^{n} \frac{I(\tilde{X}_i > x, \delta_{x_i} = 1, \tilde{Y}_i > y)}{n\hat{G}_1(\tilde{X}_i + y)}. \tag{13}$$

Previous analysis shows that there are several ways of estimating $F(x, y)$ under the dependent censoring structure. Apparently $\hat{F}(x, y)$ can handle more general situations and, compared to the Burke estimator, utilises more data. It is clear that the path decomposition accounts for the serial feature of the data and can take the advantage of usable information.

4. ASYMPTOTIC PROPERTIES OF $\hat{F}(x, y)$

In this section we discuss the large-sample behaviour of the proposed estimate. We first claim that $\hat{F}(x, y) \to F(x, y)$ in probability. Note that, since $\hat{G}_1(.) \to G_1(.)$, in probability, it follows that

$$\sum_{i=1}^{n} \frac{I(\tilde{X}_i > x, \delta_{x_i} = 1, \tilde{Y}_i = v, \delta_{y_i} = 1)}{n\hat{G}_1(\tilde{X}_i + v)} = \sum_{u>x} \sum_{i=1}^{n} \frac{I(\tilde{X}_i = u, \delta_{x_i} = 1, \tilde{Y}_i = v, \delta_{y_i} = 1)}{n\hat{G}_1(u + v)}$$

$$\to \int_{u>x} \frac{\mathrm{pr}(X \in du, C_1 > u, Y \in dv, C_1 - u > v)}{\mathrm{pr}(C_1 > u + v)}$$

$$= \int_{u>x} F(du, dv)$$

in probability. Similarly, also in probability,

$$\sum_{i=1}^{n} \frac{I(\tilde{X}_i > x, \delta_{x_i} = 1, \tilde{Y}_i \geqslant v)}{n\hat{G}_1(\tilde{X}_i + v)} \to \int_{u>x} \frac{\mathrm{pr}(X \in du, Y > v, C_1 - u > v)}{\mathrm{pr}(C_1 > u + v)} = -\int_{u>x} F(du, v).$$

It follows that

$$\hat{\Lambda}_{Y|X>x}(\Delta v) \to -\frac{F(x, dv)}{F(x, v-)} = \Lambda_{Y|X>x}(dv)$$

in probability. Since each component in (5) is consistent, $\hat{F}(x, y) \to F(x, y)$ in probability.

It can be shown that the limiting distribution of $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$ converges weakly to a zero-mean Gaussian process on $D([0, \tau_1] \times [0, \tau_2])$, where D denotes the cadlag space containing right continuous functions with left-hand limit, and (τ_1, τ_2) satisfies $\mathrm{pr}(\tilde{X} > \tau_1, \tilde{Y} > \tau_2) > 0$. A sketch of the proof is given in the Appendix. The variance of the limiting processes of $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$, however, is quite complex. One can use Efron's (1981) bootstrap to construct reliable standard error estimates.

5. AN EXAMPLE

We study a dataset adapted from Lawless (1982, p. 477) to illustrate the proposed idea. Experiments were conducted to investigate the failure of epoxy electrical cable insulation specimens under a constant voltage stress of 55 kilovolts. The data are presented in Table 1, where X denotes the time to initiate a defect and Y the subsequent additional elapsed time to failure. Three specimens (6, 13, 16) had not developed the first failure at the end of the study and induced complete censorship for the subsequent failures.

To compute $\hat{\Lambda}_{Y|X>x}(\Delta\tilde{Y}_j)$ based on (4) one has to assign the weight, $1/\hat{G}_1(x_i + \tilde{Y}_j)$, to observations in $R_{Y|X>x}(\tilde{Y}_j)$. The estimate $\hat{G}_1(.)$ based on the original dataset is listed in Table 2. It can be seen that $1/\hat{G}_1(t) = 1$ for $t < 1740$. It turns out that $1/\hat{G}_1(x_i + \tilde{Y}_j) = 1$ for all observations in S^*, which implies that all the observations contributing to $\hat{\Lambda}_{Y|X>x}(\Delta\tilde{Y}_j)$ for x, $\tilde{Y}_j > 0$ receive equal weights. Thus, based on the original dataset, the proposed estimator $\hat{F}(x, y)$ is the same as the Campbell–Földes estimator, and $\hat{F}_2(y)$ is the same as

Table 1. *Cable insulation failure data; X, the initiation time of the first failure; Y, the time to subsequent failure*

Specimen	X	Y	Specimen	X	Y
1	228	30	11	1227	39
2	106	8	12	254	46
3	246	66	13	>2440	—
4	700	72	14	435	85
5	473	25	15	1155	85*
6	>1740	—	16	>2600	—
7	155	7	17	195	27
8	414	30	18	117	27
9	1374	90	19	724	21
10	128	4	20	300	96

* is replaced by '>85' for the modified data.

Table 2. *Cable insulation failure data: Estimates of $G_1(.)$ based on original and modified data; $\hat{G}_1(.)$, the estimates based on original data; $\hat{G}_1^*(.)$, the estimates based on the modified data*

t	114	132	144	162	222	258	300	312	396	444
$\hat{G}_1(t)$	1	1	1	1	1	1	1	1	1	1
$\hat{G}_1^*(t)$	1	1	1	1	1	1	1	1	1	1

t	498	520	745	772	1240	1266	1464	1740	2440	2600
$\hat{G}_1(t)$	1	1	1	1	1	1	1	0·67	0·33	0
$\hat{G}_1^*(t)$	1	1	1	1	0·83	0·83	0·83	0·56	0·28	0

the Kaplan–Meier estimator. For biomedical data, patients usually enter the study at different times and hence it seldom happens that all censored cases correspond to the patients whose first survival times exceed the whole study time.

To see the effect of weighting, we slightly modify the dataset by letting Y_{15} be censored, that is $Y_{15} > 85$. The estimates of $G_1(.)$ based on the modified data, denoted by $\hat{G}_1^*(.)$, are also given in Table 2. To compute $\hat{F}(117, 4)$, one has to calculate $\hat{F}_1(117)$ and $\hat{\Lambda}_{Y|X>117}(\Delta 4)$. Note that $R_{Y|X>117}(4)$ contains observations 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 14, 15, 17, 18, 19, 20. In calculating $\hat{\Lambda}_{Y|X>117}(\Delta 4)$, an observation i in $R_{Y|X>117}(4)$ receives weight $1/G_1^*(x_i + 4)$. It turns out that observation #9 receives weight $1/\hat{G}_1^*(1374 + 4) = 0{\cdot}83$, and the rest all receive weight 1. It follows that $\hat{\Lambda}_{Y|X>117}(\Delta 4) = 0{\cdot}0617$. Since $F_1(117) = 0{\cdot}9$, $\hat{F}(117, 4) = 0{\cdot}8444$. The estimates of $\hat{F}(x, y)$ based on the modified data are summarised in Table 3. Note that $\hat{F}(x, y)$ is not monotone in the direction of x. It is well known that most nonparametric bivariate estimators assign negative mass (Pruitt, 1991). Also, the fairly irregular weight assigned to $\hat{F}(x, y)$ may have some contribution on the negative mass. When the sample size increases, the effect of negative mass becomes less serious. In the following simulation studies, $\hat{F}(x, y)$ performs fairly well under moderate sample size.

Table 3. *Cable insulation failure data: Estimates of $\hat{F}(x, y)$ based on the modified data*

	y													
x	0	4	7	8	21	25	27	30	39	46	66	72	85	90
0	1·0	0·94	0·88	0·83	0·77	0·71	0·60	0·48	0·41	0·36	0·30	0·24	0·19	0·08
106	0·95	0·89	0·83	0·83	0·77	0·72	0·60	0·49	0·42	0·36	0·30	0·24	0·19	0·09
117	0·90	0·84	0·78	0·78	0·72	0·66	0·61	0·49	0·42	0·37	0·30	0·24	0·19	0·09
128	0·85	0·85	0·79	0·79	0·73	0·67	0·61	0·50	0·42	0·37	0·31	0·25	0·19	0·09
155	0·80	0·80	0·80	0·80	0·74	0·68	0·62	0·50	0·43	0·37	0·31	0·25	0·19	0·09
195	0·75	0·75	0·75	0·75	0·69	0·63	0·63	0·51	0·44	0·37	0·31	0·25	0·20	0·09
228	0·70	0·70	0·70	0·70	0·64	0·58	0·58	0·52	0·44	0·38	0·32	0·26	0·20	0·09
246	0·65	0·65	0·65	0·65	0·59	0·53	0·53	0·46	0·39	0·33	0·33	0·26	0·20	0·09
254	0·60	0·60	0·60	0·60	0·54	0·47	0·47	0·41	0·33	0·33	0·33	0·27	0·21	0·09
300	0·55	0·55	0·55	0·55	0·48	0·42	0·42	0·35	0·28	0·28	0·28	0·21	0·15	0
414	0·50	0·50	0·50	0·50	0·43	0·36	0·36	0·36	0·28	0·28	0·28	0·22	0·15	0
435	0·45	0·45	0·45	0·45	0·38	0·31	0·31	0·31	0·23	0·23	0·23	0·15	0·15	0
473	0·40	0·40	0·40	0·40	0·33	0·33	0·33	0·33	0·24	0·24	0·24	0·16	0·16	0
700	0·35	0·35	0·35	0·35	0·27	0·27	0·27	0·27	0·18	0·18	0·18	0·18	0·18	0
724	0·30	0·30	0·30	0·30	0·30	0·30	0·30	0·30	0·19	0·19	0·19	0·19	0·19	0
1155	0·25	0·25	0·25	0·25	0·25	0·25	0·25	0·25	0·13	0·13	0·13	0·13	0·13	0
1227	0·20	0·20	0·20	0·20	0·20	0·20	0·20	0·20	0·20	0·20	0·20	0·20	0·20	0
1374	0·15	0·15	0·15	0·15	0·15	0·15	0·15	0·15	0·15	0·15	0·15	0·15	0·15	0

6. SIMULATION RESULTS

In this section we study the finite sample performance of estimators of $F_2(y)$ and $F(x, y)$ discussed earlier. In the simulations (X, Y) are generated from a continuous bivariate model, proposed by Clayton (1978), whose survival functions are of the form

$$F(x, y) = \left[\left\{\frac{1}{F_1(x)}\right\}^{\alpha-1} + \left\{\frac{1}{F_2(y)}\right\}^{\alpha-1} - 1\right]^{-1/(\alpha-1)},$$

where α is an association parameter related to Kendall's tau, denoted by τ, by

$$\tau = (\alpha - 1)/(\alpha + 1). \tag{14}$$

Note that, when X and Y are independent, $\tau = 0$, but the converse is not true. We used the algorithm by Prentice & Cai (1992) to generate (X, Y) by letting $X \sim \mathrm{Ex}(1)$ and $Y \sim \mathrm{Ex}(1)$. Here C_1 is generated from an exponential distribution and

$$C_2 = (C_1 - X)I(X \leqslant C_1).$$

The censoring rate of X is around 30% and the censoring rates of Y and $X + Y$ are around 50%.

A series of Monte Carlo simulations with $n = 250$ were performed. The degree of association between X and Y varies from $\tau = 0$ to $\tau = 0{\cdot}5$. The average bias and standard deviation of the estimates on some selected points are presented. Table 4 summarises the results for the marginal estimators of $F_2(y)$. Note that, even when X and Y are slightly correlated, the Kaplan–Meier estimator can be quite biased. With additional estimation of the weights, the proposed estimator, $\hat{F}(0, y)$, shows larger variation, especially when y approaches the tail region, since $1/\hat{G}_1(.)$ becomes more variable.

Tables 4 and 5 summarise the results for three estimators of $F(x, y)$, namely the proposed

Table 4. *Simulation summary statistics for the marginal estimators of* $F_2(y)$: (a), *the average bias* $\times 10^3$ (*standard deviation* $\times 10^3$) *of the proposed estimator*; (b), *the average bias* $\times 10^3$ (*standard deviation* $\times 10^3$) *of the Kaplan–Meier estimator. The replication number is* 1000

		$F_2(y)=0{\cdot}973$	$F_2(y)=0{\cdot}719$	$F_2(y)=0{\cdot}562$	$F_2(y)=0{\cdot}313$
		$n=250$			
$\tau=0$	(a)	0·146 (14·52)	0·454 (41·88)	0·174 (46·30)	−1·783 (48·80)
	(b)	0·364 (12·59)	−0·404 (36·69)	−0·100 (40·73)	−0·824 (43·00)
$\tau=0{\cdot}2$	(a)	0·191 (12·46)	−2·544 (39·35)	−3·770 (45·30)	−6·489 (51·29)
	(b)	−2·926 (13·17)	−31·20 (38·06)	−42·37 (41·41)	−48·54 (41·33)
$\tau=0{\cdot}5$	(a)	0·025 (11·32)	−2·536 (36·66)	−4·021 (43·31)	−6·750 (51·01)
	(b)	−6·901 (14·06)	−70·23 (39·55)	−90·23 (41·79)	−106·68 (37·47)
		$n=100$			
$\tau=0$	(a)	−0·508 (23·33)	−0·405 (63·90)	−1·893 (74·73)	−4·880 (78·71)
	(b)	−0·181 (20·26)	−0·638 (56·35)	−1·145 (66·79)	−4·172 (69·95)
$\tau=0{\cdot}2$	(a)	−0·327 (20·07)	−3·272 (60·74)	−6·260 (73·64)	−12·66 (81·66)
	(b)	−3·244 (21·04)	−31·23 (58·15)	−43·33 (67·25)	−52·80 (65·72)
$\tau=0{\cdot}5$	(a)	0·553 (17·51)	−3·136 (56·73)	−6·447 (69·40)	−15·82 (83·39)
	(b)	−6·880 (22·08)	−70·18 (60·96)	−97·15 (66·54)	−111·81 (61·14)

estimator in (5), the Burke estimator (1988) and the Campbell–Földes estimator. The Dabrowska estimator was also studied giving results almost identical to those of the Campbell–Földes estimator. Note that previous studies indicted that the Dabrowska estimator usually outperforms the Campbell–Földes estimator when X and Y do not follow a particular order; see Pruitt (1990) and van der Laan (1996) for more on comparisons between various estimates of the bivariate survival function. We also carried out the same study at $n=100$ and found similar results.

When $\tau=0$, see Table 5(a), the Campbell-Földes estimator performs best but, as τ increases, it may produce large bias. Our estimator, $\hat{F}(x, y)$, in general performs well in terms of bias and variation under all choices of τ and sample size. We also carried out the same simulation at $n=100$ and found that, although the Burke estimator also accounts for dependent censoring, it shows poor performance and in general has larger variance. This may be because the effective sample size for the Burke estimator is smaller. It is interesting to note that the Burke estimator performs better as y approaches the tail, whereas $\hat{F}(x, y)$ behaves in an opposite way. A possible explanation is that the Burke estimator estimates $\mathrm{pr}(X \leqslant x, Y \leqslant y)$ first so that more data are included when y gets larger.

Acknowledgement

The authors would like to thank Bruce Turnbull for his insightful comments on the paper. This work is partially funded by the Faculty Research and Professional Development Fund at North Carolina State University, the National Institutes of Health and the National Science Foundation.

Table 5. *Simulation summary statistics for the bivariate estimators when* $(X, Y) \sim$ Clayton ($\tau = 0$), $(X, Y) \sim$ Clayton ($\tau = 0{\cdot}2$) *and* $(X, Y) \sim$ Clayton ($\tau = 0{\cdot}5$), *with* X 30% *censored and* Y 50% *censored with* $n = 250$.

(*a*) $(X, Y) \sim$ Clayton ($\tau = 0$)

(x, y)	$F(x, y)$	Proposed	Burke	C&F
(0·027, 0·027)	0·949	−0·810 (17·42)	2·830 (19·10)	−0·625 (15·69)
(0·330, 0·027)	0·701	−1·126 (30·64)	2·089 (44·47)	−0·963 (29·84)
(0·577, 0·027)	0·549	−2·086 (34·96)	0·954 (50·22)	−2·001 (33·48)
(1·160, 0·027)	0·306	−1·279 (35·03)	4·450 (52·29)	−1·233 (34·61)
(0·330, 0·330)	0·517	−0·006 (43·78)	4·091 (52·43)	−0·873 (39·72)
(0·577, 0·330)	0·404	−0·301 (44·03)	3·249 (55·31)	−1·095 (40·77)
(1·160, 0·330)	0·225	1·398 (39·21)	7·392 (51·83)	0·972 (36·82)
(0·577, 0·577)	0·315	−0·413 (45·18)	3·830 (55·49)	−1·042 (41·34)
(1·160, 0·577)	0·176	0·128 (40·09)	6·953 (50·85)	−0·549 (37·69)
(1·160, 1·160)	0·098	−1·053 (37·81)	7·320 (46·60)	−0·868 (34·58)

(*b*) $(X, Y) \sim$ Clayton ($\tau = 0{\cdot}2$)

(x, y)	$F(x, y)$	Proposed	Burke	C&F
(0·027, 0·027)	0·948	0·381 (15·52)	3·817 (15·99)	−2·607 (15·86)
(0·330, 0·027)	0·702	0·545 (29·94)	3·685 (37·21)	−1·411 (29·84)
(0·577, 0·027)	0·550	0·841 (32·71)	2·935 (43·34)	−0·523 (32·66)
(1·160, 0·027)	0·307	−0·634 (34·91)	2·677 (49·44)	−1·121 (34·96)
(0·330, 0·330)	0·540	−1·179 (40·98)	2·870 (45·43)	−19·20 (40·22)
(0·577, 0·330)	0·435	−0·206 (40·21)	2·416 (48·55)	−12·91 (39·57)
(1·160, 0·330)	0·256	−0·675 (38·53)	2·788 (49·82)	−5·833 (28·27)
(0·577, 0·577)	0·357	−1·645 (42·69)	1·477 (49·58)	−19·94 (40·63)
(1·160, 0·577)	0·220	−2·398 (39·65)	1·643 (50·33)	−10·66 (38·40)
(1·160, 1·160)	0·150	−5·286 (42·62)	0·518 (51·30)	−16·55 (38·55)

(*c*) $(X, Y) \sim$ Clayton ($\tau = 0{\cdot}5$)

(x, y)	$F(x, y)$	Proposed	Burke	C&F
(0·027, 0·027)	0·949	0·329 (14·65)	3·045 (15·16)	−6·253 (15·16)
(0·330, 0·027)	0·709	−0·345 (28·15)	2·268 (32·11)	−3·091 (32·11)
(0·577, 0·027)	0·556	−0·099 (33·63)	2·199 (39·91)	−1·420 (39·91)
(1·160, 0·027)	0·312	−0·293 (34·63)	1·472 (43·54)	−0·540 (34·67)
(0·330, 0·330)	0·589	−1·062 (35·97)	1·368 (38·60)	−33·380 (38·29)
(0·577, 0·330)	0·493	−0·842 (36·70)	0·856 (44·10)	−18·310 (38·46)
(1·160, 0·330)	0·298	−0·938 (35·40)	0·874 (43·69)	−4·658 (36·19)
(0·577, 0·577)	0·432	−2·476 (35·35)	−0·676 (46·86)	−33·490 (40·61)
(1·160, 0·577)	0·282	−1·326 (35·95)	0·715 (43·73)	−8·648 (37·16)
(1·160, 1·160)	0·252	−4·153 (40·45)	−0·140 (44·16)	−23·080 (41·18)

In the left-hand column, the first item is the selected grid point and the next is the true survival probability. The first number in each cell is the average bias ($\times 10^3$), and the second number, in parentheses, is the standard deviation ($\times 10^3$) of the estimate based on 1000 replications.

Proposed, proposed estimator (5); Burke, Burke (1988) estimator; C&F, Campbell–Földer estimator.

Appendix

Weak convergence results

We briefly discuss the weak convergence result for $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$. We have

$$n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\} = n^{\frac{1}{2}}\hat{F}(y\,|\,x)\{\hat{F}_1(x) - F_1(x)\} + n^{\frac{1}{2}}F_1(x)\{\hat{F}(y\,|\,x) - F(y\,|\,x)\}, \tag{A1}$$

where

$$\hat{F}(y\,|\,x) = \prod_{v \leqslant y}\{1 - \hat{\Lambda}_{Y|X>x}(dv)\}$$

is the product limit estimator of $F(y\,|\,x) = \mathrm{pr}(Y > y\,|\,X > x)$. It can be shown, by tedious Taylor expansions and integration by parts, that $n^{\frac{1}{2}}\{\hat{\Lambda}_{Y|X>x}(y) - \Lambda_{Y|X>x}(y)\}$ is a smooth Hadamard differentiable functional of $n^{\frac{1}{2}}\{\hat{G}_1(.) - G_1(.)\}$ and the following processes:

$$n^{\frac{1}{2}}\{\hat{H}_{11}(x, y) - H_{11}(x, y)\} = n^{\frac{1}{2}}\left\{\frac{1}{2}\sum_{i=1}^{n} I(\tilde{X}_i > x, \tilde{Y}_i > y, \delta_{x_i} = 1, \delta_{y_i} = 1) - \mathrm{pr}(\tilde{X} > x, \tilde{Y} > y, \delta_x > 1, \delta_y = 1)\right\},$$

$$n^{\frac{1}{2}}\{\hat{H}_{10}(x, y) - H_{10}(x, y)\} = n^{\frac{1}{2}}\left\{\frac{1}{n}\sum_{i=1}^{n} I(\tilde{X}_i > x, \tilde{Y}_i > y, \delta_{x_i} = 1) - \mathrm{pr}(\tilde{X} > x, \tilde{Y} > y, \delta_x > 1)\right\}.$$

Weak convergence of $n^{\frac{1}{2}}\{\hat{\Lambda}_{Y|X>x}(y) - \Lambda_{Y|X>x}(y)\}$ can be derived by known results on empirical processes and the univariate Kaplan–Meier estimator. Since, for each $x \in [0, \tau_1]$, $F(y\,|\,x)$ is a functional of $\Lambda_{Y|X>x}(y)$ and is compactly differentiable on $D[0, \tau_2]$ (Andersen et al., 1991, Proposition II 8.7), the weak convergence of $n^{\frac{1}{2}}\{\hat{\Lambda}_{Y|X>x}(y) - \Lambda_{Y|X>x}(y)\}$ implies the weak convergence of $n^{\frac{1}{2}}\{\hat{F}(y\,|\,x) - F(y\,|\,x)\}$. Since $\hat{F}(y\,|\,x)$ and $\hat{F}_1(x)$ are consistent estimators and $n^{\frac{1}{2}}\{\hat{F}_1(x) - F_1(x)\}$ converges weakly to a zero mean Gaussian process on $D[0, \tau_1]$ weak convergence of $n^{\frac{1}{2}}\{\hat{F}(x, y) - F(x, y)\}$ on $D([0, \tau_1] \times [0, \tau_2])$ can be established. Since $\hat{F}_2(y)$ is simply $\hat{F}(0, y)$, weak convergence of $n^{\frac{1}{2}}\{\hat{F}_2(y) - F_2(y)\}$ on $D[0, \tau_2]$ follows.

References

Andersen P. K., Borgan, O., Gill, R. D. & Keiding, N. (1992). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

Burke, M. D. (1988). Estimation of a bivariate distribution under random censoring. *Biometrika* **75**, 379–82.

Campbell, G. & Földes, A. (1982). Large sample properties of nonparametric statistical inference. In *Colloquia Methemetica-Societatis, János Bolyai*, Ed. B. V. Gnedenko, M. L. Puri and I. Vincze, pp. 103–22. Amsterdam: North Holland.

Clayton, 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. (1987). Nonparametric regression with censored data survival times. *Scand. J. Statist.* **14**, 181–97.

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

Efron, B. (1981). Censored data and the bootstrap. *J. Am. Statist. Assoc.* **76**, 312–21.

Fleming, T. R., Prentice, R. L., Pepe, M. S. & Glidden, D. (1994). Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. *Statist. Med.* **13**, 956–68.

Frydman, H. (1995). Semiparametric estimation in three-state duration-dependent Markov model from interval-censored observations with applications to AIDS data. *Biometrics* **51**, 502–11.

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

Gray, R. J. (1994). A kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika* **81**, 527–40.

Jewell, N. P. & Kalbfleisch, J. D (1992). Marker models in survival analysis and applications to issues associated with AIDS. In *AIDS Epidemiology: Methodological Issues*, Ed. N. Jewell, K. Keitz and V. Farewell, pp. 211–30. Boston: Birkhäuser-Boston.

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

LAWLESS, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: Wiley.

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

LONGINI, I. M., CLARK, W. S., BEYERS, R. H., WARD, J. W., DARROW, W. W., LEMP, G. F. & HETHCOTE, H. W. (1989). Statistical analysis of the stages of HIV infection using a Markov model. *Statist. Med.* **8**, 831–43.

MALANI, H. M. (1995). A modification of the redistribution to the right algorithm using disease markers. *Biometrika* **82**, 515–26.

MALANI, H. M., REDFEARN, W. J. & NIELSEN, J. P. (1992). A note on the asymptotic variance of survival function in the semi-Markov model. *Statist. Prob. Lett.* **18**, 19–25.

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

PRUITT, R. C. (1990). Small sample comparison of six bivariate survival curve estimators. *J. Statist. Comp. Simul.* **45**, 147–67.

PRUITT, R. C. (1991). On negative mass assigned by the bivariate Kaplan–Meier estimation. *Ann. Statist.* **19**, 443–53.

ROBINS J. M. (1992). Analytical methods for estimating HIV-treatment and cofactor effects. In *Methodological Issues of AIDS Behavioral Research*, Ed. D. G. Ostrow and R. Kessler, pp. 213–90. New York: Plenum.

ROBINS, J. M. & ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology: Methodological Issues*, Ed. N. Jewell, K. Dietz and V. Farewell, pp. 297–331. Boston: Birkhäuser-Boston.

TSAI, W.-Y. & CROWLEY, J. (1998). A note on nonparametric estimators of the bivariate survival function under univariate censoring. *Biometrika* **85**, 573–80.

TSAI, W.-Y., LEURGANS, S. & CROWLEY, 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.

VISSER M. (1996). Nonparametric estimation of the bivariate survival function with application to vertically transmitted AIDS. *Biometrika* **83**, 507–18.

[*Received December* 1995. *Revised January* 1998]