Biometrika (1999), **86**, 1, *pp.* 59–70

Printed in Great Britain

Nonparametric estimation of the gap time distributions for serial events with censored data

By D. Y. LIN

Department of Biostatistics, Box 357232, University of Washington, Seattle, Washington 98195, U.S.A.

danyu@biostat.washington.edu

W. SUN

Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, New Jersey 07033, U.S.A.

wei.sun@spcorp.com

AND ZHILIANG YING

Department of Statistics, Hill Center, Busch Campus, Rutgers University, Piscataway, New Jersey 08855, U.S.A.

zying@stat.rutgers.edu

Summary

In many follow-up studies, each subject can potentially experience a series of events, which may be repetitions of essentially the same event or may be events of entirely different natures. This paper provides a simple nonparametric estimator for the multivariate distribution function of the gap times between successive events when the follow-up time is subject to right censoring. The estimator is consistent and, upon proper normalisation, converges weakly to a zero-mean Gaussian process with an easily estimated covariance function. Numerical studies demonstrate that both the distribution function estimator and its covariance function estimator perform well for practical sample sizes. An application to a colon cancer study is presented.

Some key words: Bivariate distribution; Correlated failure times; Dependent censoring; Kaplan–Meier estimator; Multiple events; Multivariate failure time; Recurrent events.

1. Introduction

In many scientific investigations, each study subject can potentially experience more than one event. These multiple events data normally fall into one of two categories, ‘parallel’ and ‘serial’. In the parallel system, several possibly dependent failure processes act concurrently, while in the serial system there is a natural ordering of times of occurrence of events. Data of the latter type arise more frequently than the former and are the focus of this paper. Medical examples of serial events include the recurrences of a given illness, such as infection episodes, and the progression of a disease through successive stages, such as HIV infection → AIDS → death.

There are two possible time scales for serial events: the gap time is the duration between two successive states or events while the total time is measured from the start of follow-up to the occurrence of the event. In many applications, the investigators are more interested in the gap time than the total time, e.g. Gail, Santner & Brown (1980). For example, when evaluating the efficacy of a treatment on an episodic illness, it is often important to assess whether or not the treatment delays the time from the initiation of the treatment to the first episode as well as the time from the first episode to the second episode, and so on. The total time from the initiation of the treatment to the second episode is of less interest because a treatment which delays the first episode will inevitably lengthen the total time to the second episode even if it becomes ineffective after the first episode.

A fundamental problem in analysing multiple events data is the estimation of the multivariate distribution function in the presence of right censoring. A number of nonparametric estimators are currently available for estimating the joint survival distribution of parallel events based on censored observations; see Andersen et al. (1993, pp. 688–704) for a survey. In dealing with the gap time distributions of several events, the main new difficulty is that, when the overall follow-up time is subject to independent right censoring, the gap times except the first one are subject to dependent censoring. Therefore, even the estimation of the marginal distributions of the gap times is nontrivial.

In this paper, we present a simple nonparametric approach to estimating the joint and marginal distributions of the gap times. We describe the proposed estimators and their asymptotic properties in the next section. We then report the results of some simulation studies in § 3 and present a real example in § 4. In the final section, we discuss some extensions and potential uses of the proposed estimators.

2. Nonparametric estimation

Suppose that an individual subject may experience K consecutive events at times $Y_1 < Y_2 < \ldots < Y_K$, which are measured from the start of the follow-up. We are primarily interested in the gap times $T_1 := Y_1$, $T_2 := Y_2 - Y_1, \ldots$ and $T_K := Y_K - Y_{K-1}$. As usual, assume that the follow-up time is subject to independent right censoring by C, which implies that $(Y_1, \ldots, Y_K)$ are independent of C. On the other hand, for any $k = 2, \ldots, K$, the gap time T_k is subject to right censoring by $C - Y_{k-1}$, which is naturally correlated with T_k unless T_k is independent of Y_{k-1}. The marginal distributions of $(T_2, \ldots, T_K)$ cannot therefore be estimated by the Kaplan–Meier method, and neither can the joint distribution of $(T_1, \ldots, T_K)$ be estimated by any existing estimator for parallel events.

We will develop a simple nonparametric estimator for the joint distribution of $(T_1, \ldots, T_K)$, which does not impose any assumption on the dependence structures of the gap times. To ease our presentation, we take $K = 2$ in this section and discuss the extension to the setting of $K > 2$ in § 5.

Suppose that there are n independent subjects in the study so that (Y_{1i}, Y_{2i}, C_i) $(i = 1, \ldots, n)$ are n independent replicates of (Y_1, Y_2, C). The observable data consist of $(\tilde{Y}_{1i}, \tilde{Y}_{2i}, \delta_{1i}, \delta_{2i})$ $(i = 1, \ldots, n)$, where $\tilde{Y}_{ki} = Y_{ki} \wedge C_i$ and $\delta_{ki} = I(Y_{ki} \leqslant C_i)$ $(k = 1, 2;$ $i = 1, \ldots, n)$. Here and in the sequel, $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$, $a^+ = \max(a, 0)$ and $I(.)$ is the indicator function.

Let F_1 and F_2 be the marginal distribution functions of T_1 and T_2 and let F be their joint distribution function. That is,

$$F_1(t) = \mathrm{pr}(T_1 \leqslant t), \quad F_2(t) = \mathrm{pr}(T_2 \leqslant t), \quad F(t_1, t_2) = \mathrm{pr}(T_1 \leqslant t_1, T_2 \leqslant t_2).$$

For simplicity of description, we assume that F is continuous. Clearly, $F(t_1, t_2) = H(t_1, 0) - H(t_1, t_2)$, where $H(t_1, t_2) = \mathrm{pr}(T_1 \leqslant t_1, T_2 > t_2)$. Thus, we will have an estimator for F if we know how to estimate H.

If there were no censoring, then $H(t_1, t_2)$ could be estimated by

$$n^{-1} \sum_{i=1}^{n} I(T_{1i} \leqslant t_1, T_{2i} > t_2). \tag{1}$$

In the presence of censoring, one observes $\tilde{Y}_{1i} := T_{1i} \wedge C_i$ and $\tilde{Y}_{2i} := (T_{1i} + T_{2i}) \wedge C_i$ instead of T_{1i} and T_{2i}. Motivated by expression (1), we consider the indicator function $I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2)$. Since $\tilde{Y}_{2i} - \tilde{Y}_{1i} > 0$ implies that $\tilde{Y}_{1i} = T_{1i}$, we have

$$I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2) = I(T_{1i} \leqslant t_1, T_{2i} > t_2, C_i > T_{1i} + t_2). \tag{2}$$

Thus,

$$E\{I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2) \mid T_{1i}, T_{2i}\} = I(T_{1i} \leqslant t_1, T_{2i} > t_2) G(T_{1i} + t_2), \tag{3}$$

where G is the common survival function of the censoring time variable, that is $G(t) = \mathrm{pr}(C > t)$. It follows from (3) that

$$E\left\{ \frac{I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2)}{G(T_{1i} + t_2)} \,\middle|\, T_{1i}, T_{2i} \right\} = I(T_{1i} \leqslant t_1, T_{2i} > t_2), \tag{4}$$

which indicates that

$$\tilde{H}(t_1, t_2) := n^{-1} \sum_{i=1}^{n} \frac{I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2)}{G(\tilde{Y}_{1i} + t_2)}$$

would be an unbiased estimator of $H(t_1, t_2)$ if G were known. Therefore, we estimate $H(t_1, t_2)$ by

$$\hat{H}(t_1, t_2) := n^{-1} \sum_{i=1}^{n} \frac{I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2)}{\hat{G}(\tilde{Y}_{1i} + t_2)},$$

where $\hat{G}$ is the Kaplan–Meier estimator of G based on the data $(\tilde{Y}_{1i}, 1 - \delta_{1i})$ $(i = 1, \ldots, n)$ or $(\tilde{Y}_{2i}, 1 - \delta_{2i})$ $(i = 1, \ldots, n)$. The corresponding estimator for $F(t_1, t_2)$ is

$$\hat{F}(t_1, t_2) := \hat{H}(t_1, 0) - \hat{H}(t_1, t_2).$$

In deriving equation (4) from (3), we implicitly assumed that the denominator $G(T_{1i} + t_2)$ is strictly positive for every i such that $T_{1i} \leqslant t_1$. This condition holds if $t_1 + t_2 < \tau_c$, where $\tau_c = \sup\{t : G(t) > 0\}$. In fact, $F(t_1, t_2)$ and $H(t_1, t_2)$ are not estimable if $t_1 + t_2 > \tau_c$. An intuitive explanation for the non-estimability is as follows: if $T_1 = \tau_c - t_2$, then T_2 is observable only when $T_2 \leqslant t_2$, which implies that there is no information in the data to estimate the probability of $T_2 > t_2$. As a result of this estimability constraint, the estimator $\hat{H}(t_1, t_2)$, or any other potential estimator, is naturally confined to $\{(t_1, t_2) : t_1 + t_2 < \tau_c\}$. Obviously, this constraint will be redundant if $\tau_c = \infty$ or $\tau_c \geqslant \tau_1 + \tau_2$, where $\tau_1 = \sup\{t : F_1(t) = 1\}$ and $\tau_2 = \sup\{t : F_2(t) = 1\}$. It is well known that the Kaplan–Meier estimator for $F_1(t)$ can be defined only up to τ_c.

Clearly, $\hat{F}$ reduces to the usual empirical distribution function in the absence of censoring. Note also that

$$\hat{F}(t, \infty) = \hat{H}(t, 0) = n^{-1} \sum_{i=1}^{n} \frac{\delta_{1i} I(\tilde{Y}_{1i} \leqslant t)}{\hat{G}(\tilde{Y}_{1i})}. \tag{5}$$

If $\hat{G}$ is the Kaplan–Meier estimator of G calculated from $(\tilde{Y}_{1i}, 1 - \delta_{1i})$ $(i = 1, \ldots, n)$, then (5) is the estimator of $F_1(t)$ given by Susarla, Tsai & Van Ryzin (1984). It can be shown that this estimator is identical to the Kaplan–Meier estimator of F_1.

As discussed previously, in general $F_2(t)$ is estimable only when $\tau_c > \tau_1 + t$. Assume for the moment that this condition holds. Then

$$\hat{H}(\infty, t) = n^{-1} \sum_{i=1}^{n} \frac{I(\tilde{Y}_{2i} - \tilde{Y}_{1i} > t)}{\hat{G}(\tilde{Y}_{1i} + t)} \tag{6}$$

is our estimator for $S_2(t) := 1 - F_2(t)$. As a result of dependent censoring, the Kaplan–Meier estimator for S_2 is invalid unless T_1 and T_2 are independent. If T_1 were a fixed constant and thus independent of T_2, then the right-hand side of (6) with $\hat{G}$ calculated from $(\tilde{Y}_{2i}, 1 - \delta_{2i})$ $(i = 1, \ldots, n)$ would reduce to the Susarla–Tsai–Van Ryzin estimator of $S_2(t)$. It can be easily shown that the Susarla–Tsai–Van Ryzin and Kaplan–Meier estimators have the same probability masses and that the latter accumulates masses forward from 0 to ∞ whereas the former does it backwards. Therefore, the two estimators are identical when the last observation time is uncensored. This connection, as well as the connections described in the previous paragraph, suggests that $\hat{F}(t_1, t_2)$ is a natural extension of the Kaplan–Meier approach to the estimation of the bivariate distribution function of two gap times and is likely to have high efficiency.

When $\tau_c < \tau_1 + t$, it is generally not possible to estimate the marginal distribution function $F_2(t)$, but it is possible to estimate the conditional distribution function

$$F_{2|1}(t_2 \mid t_1) := \mathrm{pr}(T_2 \leqslant t_2 \mid T_1 \leqslant t_1) = F(t_1, t_2)/F_1(t_1)$$

as long as $t_1 + t_2 < \tau_c$. In fact, this conditional distribution may be of interest even when F_2 is estimable. Naturally, $F_{2|1}(t_2 \mid t_1)$ is estimated by

$$\hat{F}_{2|1}(t_2 \mid t_1) := \hat{F}(t_1, t_2)/\hat{H}(t_1, 0) = 1 - \hat{H}(t_1, t_2)/\hat{H}(t_1, 0).$$

For definiteness, suppose that $\hat{G}$ is constructed from $(\tilde{Y}_{2i}, 1 - \delta_{2i})$ $(i = 1, \ldots, n)$. We show in the Appendix that the estimator $\hat{F}$ is strongly consistent, and the process $n^{\frac{1}{2}}\{\hat{F}(.,.) - F(.,.)\}$ converges weakly to a bivariate zero-mean Gaussian process with covariance function

$$\begin{aligned} \sigma(t_1, t_2; t_1', t_2') := E\Bigg[& \left\{ \frac{\delta_1 I(\tilde{Y}_1 \leqslant t_1)}{G(\tilde{Y}_1)} - \frac{I(\tilde{Y}_1 \leqslant t_1, \tilde{Y}_2 - \tilde{Y}_1 > t_2)}{G(\tilde{Y}_1 + t_2)} - F(t_1, t_2) \right\} \\ & \times \left\{ \frac{\delta_1 I(\tilde{Y}_1 \leqslant t_1')}{G(\tilde{Y}_1)} - \frac{I(\tilde{Y}_1 \leqslant t_1', \tilde{Y}_2 - \tilde{Y}_1 > t_2')}{G(\tilde{Y}_1 + t_2')} - F(t_1', t_2') \right\} \Bigg] \\ & - \int_0^{\tau_c} \frac{D(t_1, t_2; u) D(t_1', t_2'; u)\, d\Lambda_c(u)}{\mathrm{pr}(\tilde{Y}_2 > u)}, \end{aligned} \tag{7}$$

where $D(t_1, t_2; u) = \{F_1(t_1) - F_1(u)\}^+ - \{H(t_1, t_2) - H(u - t_2, t_2)\}^+$, and Λ_c is the cumulative hazard function of C. A consistent estimator of $\sigma(t_1, t_2; t_1', t_2')$ is

$$\begin{aligned} \hat{\sigma}(t_1, t_2; t_1', t_2') := n^{-1} \sum_{i=1}^{n} \Bigg[& \left\{ \frac{\delta_{1i} I(\tilde{Y}_{1i} \leqslant t_1)}{\hat{G}(\tilde{Y}_{1i})} - \frac{I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2)}{\hat{G}(\tilde{Y}_{1i} + t_2)} - \hat{F}(t_1, t_2) \right\} \\ & \times \left\{ \frac{\delta_{1i} I(\tilde{Y}_{1i} \leqslant t_1')}{\hat{G}(\tilde{Y}_{1i})} - \frac{I(\tilde{Y}_{1i} \leqslant t_1', \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2')}{\hat{G}(\tilde{Y}_{1i} + t_2')} - \hat{F}(t_1', t_2') \right\} \\ & - \frac{(1 - \delta_{2i}) \hat{D}(t_1, t_2; \tilde{Y}_{2i}) \hat{D}(t_1', t_2'; \tilde{Y}_{2i})}{n^{-2}\{1 + \sum_{j=1}^{n} I(\tilde{Y}_{2j} > \tilde{Y}_{2i})\} \sum_{l=1}^{n} I(\tilde{Y}_{2l} \geqslant \tilde{Y}_{2i})} \Bigg], \end{aligned} \tag{8}$$

where $\hat{D}(t_1, t_2; u) = \{\hat{H}(t_1, 0) - \hat{H}(u, 0)\}^+ - \{\hat{H}(t_1, t_2) - \hat{H}(u - t_2, t_2)\}^+$. Furthermore, the estimator $\hat{F}_{2|1}$ is also strongly consistent, and $n^{\frac{1}{2}}\{\hat{F}_{2|1}(.,.) - F_{2|1}(.,.)\}$ converges weakly to a zero-mean Gaussian process with a covariance function which can be consistently estimated by

$$\hat{\sigma}(t_2; t_2' | t_1) := \frac{n^{-1}}{\hat{H}^2(t_1, 0)} \sum_{i=1}^{n} \left[\delta_{1i} I(\tilde{Y}_{1i} \leqslant t_1) \left\{ \frac{\hat{H}(t_2 | t_1)}{\hat{G}(\tilde{Y}_{1i})} - \frac{I(\tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2)}{\hat{G}(\tilde{Y}_{1i} + t_2)} \right\} \right.$$
$$\times \left\{ \frac{\hat{H}(t_2' | t_1)}{\hat{G}(\tilde{Y}_{1i})} - \frac{I(\tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2')}{\hat{G}(\tilde{Y}_{1i} + t_2')} \right\}$$
$$\left. - \frac{(1 - \delta_{2i}) \hat{B}(t_1, t_2; \tilde{Y}_{2i}) \hat{B}(t_1, t_2'; \tilde{Y}_{2i})}{n^{-2}\{1 + \sum_{j=1}^{n} I(\tilde{Y}_{2j} > \tilde{Y}_{2i})\} \sum_{l=1}^{n} I(\tilde{Y}_{2l} \geqslant \tilde{Y}_{2i})} \right], \quad (9)$$

where $\hat{H}(t_2 | t_1) = \{\hat{H}(t_1, t_2)\}/\{\hat{H}(t_1, 0)\}$,

$$\hat{B}(t_1, t_2; u) = \hat{H}(t_2 | t_1)\{\hat{H}(t_1, 0) - \hat{H}(u, 0)\}^+ - \{\hat{H}(t_1, t_2) - \hat{H}(u - t_2, t_2)\}^+.$$

3. Simulation studies

Two sets of simulations were carried out to assess the finite-sample performance of the joint distribution function estimator $\hat{F}$ and conditional distribution function estimator $\hat{F}_{2|1}$ as well as their variance estimators. The gap times (T_1, T_2) were generated from Gumbel's (1960) bivariate distribution function

$$F(t_1, t_2) = F_1(t_1) F_2(t_2)[1 + \theta\{1 - F_1(t_1)\}\{1 - F_2(t_2)\}]$$

with unit exponential margins. The parameter θ was set to 0 and 1 in the first and second sets of simulations, respectively, corresponding to 0 and 0·25 correlation between T_1 and T_2. The follow-up time was subject to right censoring by an independent Un[0, 4] variable so that about 25% of T_1 and 50% of T_2 were censored. In each simulation study, 10 000 samples were generated, each with 100 subjects.

Table 1 summarises the main findings of the simulations. For the joint distribution function, the results are given at pairs of time points (t_1, t_2), where t_1 and t_2 take values 0·2231, 0·5108, 0·9163 and 1·6094, corresponding to marginal survival probabilities of 0·8, 0·6, 0·4 and 0·2. At each (t_1, t_2), both the estimator $\hat{F}(t_1, t_2)$ and its standard error estimator appear to be unbiased. For the conditional distribution, the results are displayed at $t_2 =$ 0·2231, 0·5108, 0·9163 or 1·6094 and $t_1 = 0{\cdot}5108$ or 1·6094. Again, both the estimator $\hat{F}_{2|1}(t_2 | t_1)$ and its standard error estimator are virtually unbiased, although there seems to be slight underestimation of the true variability at $t_2 = 1{\cdot}6094$.

4. A real example

Each year, cancer of the colon afflicts over 100 000 persons in the United States. In approximately 80% of the patients, the diagnosis is made at a sufficiently early stage when all apparent diseased tissue can be surgically removed. Those who have regional nodal involvement that is clinically completely resected are referred to as having Duke's Stage C disease. Unfortunately, about one-half of these patients have residual cancer existing in an occult and probably microscopic stage, which leads to recurrence of disease and death within 5 years. A recent clinical trial on Duke's Stage C patients demonstrated that therapy

Table 1. *Simulation summary statistics for $\hat{F}$ and $\hat{F}_{2|1}$ under bivariate exponential models*: (a) *true probabilities*, (b) *empirical means of estimated probabilities*, (c) *empirical standard errors of estimated probabilities, and* (d) *empirical means of standard error estimates*

			Independent gap times				Dependent gap times			
			$t_2=$				$t_2=$			
	t_1		0·2231	0·5018	0·9163	1·6094	0·2231	0·5108	0·9163	1·6094
$\hat{F}(t_1, t_2)$	0·2231	(a)	0·04	0·08	0·12	0·16	0·066	0·118	0·158	0·186
		(b)	0·040	0·080	0·120	0·160	0·066	0·118	0·158	0·185
		(c)	0·022	0·031	0·036	0·041	0·027	0·034	0·038	0·040
		(d)	0·021	0·030	0·036	0·040	0·026	0·034	0·038	0·040
	0·5108	(a)	0·08	0·16	0·24	0·32	0·118	0·218	0·298	0·358
		(b)	0·080	0·160	0·240	0·320	0·118	0·218	0·297	0·358
		(c)	0·031	0·043	0·050	0·055	0·035	0·045	0·050	0·053
		(d)	0·031	0·042	0·049	0·054	0·035	0·045	0·050	0·053
	0·9163	(a)	0·12	0·24	0·36	0·48	0·158	0·298	0·418	0·518
		(b)	0·120	0·240	0·360	0·480	0·158	0·297	0·417	0·518
		(c)	0·039	0·052	0·059	0·063	0·042	0·053	0·058	0·061
		(d)	0·039	0·051	0·058	0·062	0·042	0·053	0·058	0·060
	1·6094	(a)	0·16	0·32	0·48	0·64	0·186	0·358	0·518	0·666
		(b)	0·160	0·320	0·481	0·640	0·185	0·357	0·517	0·664
		(c)	0·047	0·062	0·069	0·071	0·048	0·061	0·067	0·069
		(d)	0·064	0·061	0·068	0·069	0·048	0·061	0·067	0·068
$\hat{F}_{2\|1}(t_2\|t_1)$	0·5108	(a)	0·2	0·4	0·6	0·8	0·296	0·544	0·744	0·896
		(b)	0·200	0·399	0·600	0·800	0·297	0·545	0·744	0·896
		(c)	0·074	0·094	0·098	0·089	0·082	0·092	0·085	0·068
		(d)	0·073	0·091	0·096	0·086	0·080	0·090	0·083	0·061
	1·6094	(a)	0·2	0·4	0·6	0·8	0·232	0·448	0·648	0·832
		(b)	0·200	0·400	0·600	0·800	0·231	0·447	0·646	0·830
		(c)	0·057	0·071	0·075	0·072	0·058	0·070	0·072	0·070
		(d)	0·056	0·070	0·073	0·068	0·057	0·070	0·072	0·066

with levamisole plus fluorouracil delayed the time to cancer recurrence as well as time to death as measured from the time of randomisation (Moertel et al., 1990). An important issue that was not addressed by the investigators is whether or not therapy had any benefit on survival after cancer recurrence. The methods developed in this paper were motivated by the need to answer such questions.

There were 315 and 304 patients in the observation and therapy groups, respectively. The database available for this analysis contains considerably richer long-term information than that used in the original report, with maximum follow-up time of more than 8 years. By the end of the study, 177 patients in the observation group had cancer recurrence, among whom 155 died, while in the therapy group 119 patients had cancer recurrence, among whom 108 died. Figures 1(a) and (b) display the Kaplan–Meier estimates for the cumulative probabilities of cancer recurrence and death. The observed chi-squared values of the log-rank statistics are 19·1 and 11·2 for cancer recurrence and death, respectively, providing strong evidence for the benefit of therapy.

Under the proposed framework for analysing gap times, T_1 is the time from randomisation to cancer recurrence and T_2 is the time from cancer recurrence to death. Table 2 presents the estimates for the joint distribution function $F(t_1, t_2)$ and the conditional distribution function $F_{2|1}(t_2|t_1)$ for $t_1 = 1, 2, 3, 4, 5$ years and $t_2 = 1, 2, 3$ years. The joint

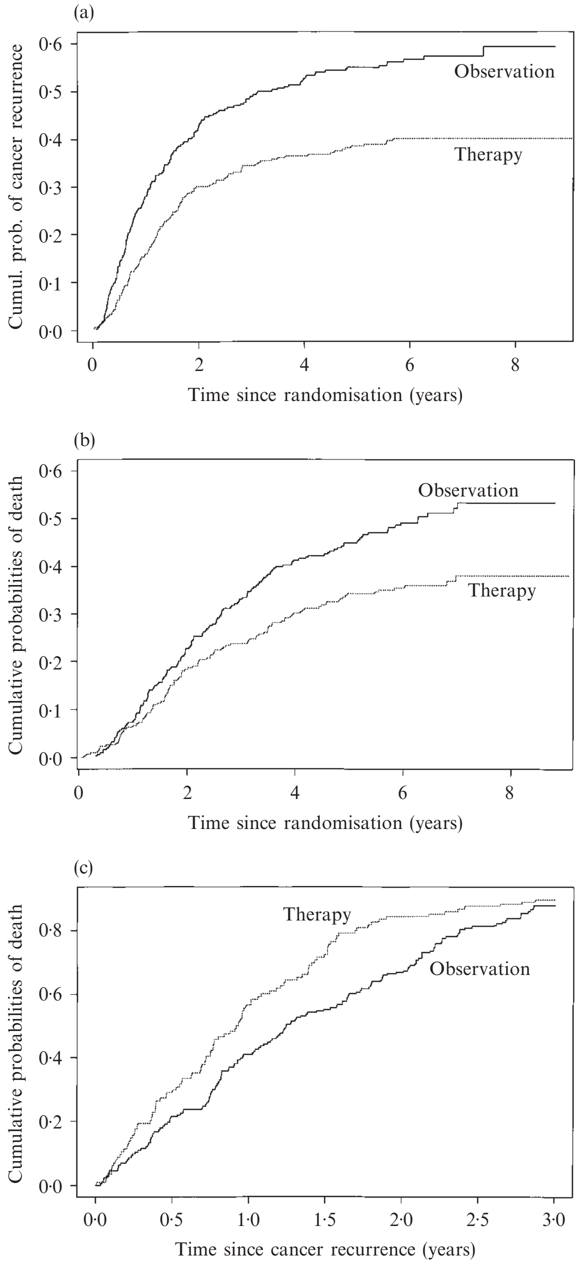

Fig. 1. Nonparametric estimation of the cumulative probabilities of cancer recurrence and death in the colon cancer study: (a) Kaplan–Meier estimates for the cumulative probabilities of cancer recurrence since randomisation, (b) Kaplan–Meier estimates for the cumulative probabilities of death since randomisation, (c) proposed estimates for the cumulative probabilities of death since cancer recurrence. The estimates for the observation group are shown by the solid curves and those of the therapy group by the dotted curves.

distribution function estimates of the observation group are considerably higher than those of the therapy group. However, the opposite is true for the conditional distribution function. For each t_2, a smaller value of t_1 tends to be associated with a higher value of $\hat{F}_{2|1}(t_2|t_1)$, which suggests a positive correlation between T_1 and T_2. As evident from

Fig. 1(a), cancer recurrence normally takes place within 5 years it at all. Thus, we use $\hat{F}_{2|1}(t_2|t_1)$ evaluated at $t_1 = 5$ to estimate the marginal distribution of T_2. Figure 1(c) shows these estimates when t_2 varies continuously between 1 and 3 years. Evidently, nearly 90% of the patients who suffered from cancer recurrence died within 3 years of recurrence. More importantly, the patients who were on therapy died faster after cancer recurrence compared to those in the observation group. This suggests that the benefit of therapy may not outweigh its risk or toxicities after cancer recurrence.

Table 2. *Estimates of the joint and conditional distribution functions for the colon cancer study. Standard error estimates are given in parentheses*

	$\hat{F}(t_1, t_2)$			$\hat{F}_{2\|1}(t_2, t_1)$		
Therapy	$t_2 = 1$	$t_2 = 2$	$t_3 = 3$	$t_2 = 1$	$t_2 = 2$	$t_3 = 3$
		$t_1 = 1$			$t_1 = 1$	
No	0·151 (0·020)	0·231 (0·024)	0·263 (0·025)	0·547 (0·054)	0·835 (0·040)	0·952 (0·023)
Yes	0·116 (0·018)	0·146 (0·020)	0·149 (0·021)	0·742 (0·064)	0·935 (0·036)	0·957 (0·030)
		$t_1 = 2$			$t_1 = 2$	
No	0·198 (0·023)	0·316 (0·026)	0·384 (0·028)	0·470 (0·044)	0·751 (0·038)	0·913 (0·025)
Yes	0·185 (0·022)	0·252 (0·025)	0·265 (0·026)	0·633 (0·052)	0·862 (0·037)	0·906 (0·032)
		$t_1 = 3$			$t_1 = 3$	
No	0·213 (0·023)	0·351 (0·027)	0·425 (0·028)	0·439 (0·041)	0·723 (0·037)	0·876 (0·028)
Yes	0·199 (0·023)	0·286 (0·026)	0·305 (0·027)	0·590 (0·049)	0·849 (0·036)	0·905 (0·030)
		$t_1 = 4$			$t_1 = 4$	
No	0·216 (0·023)	0·364 (0·028)	0·450 (0·029)	0·417 (0·039)	0·703 (0·037)	0·868 (0·031)
Yes	0·202 (0·023)	0·298 (0·027)	0·317 (0·028)	0·565 (0·048)	0·836 (0·037)	0·889 (0·035)
		$t_1 = 5$			$t_1 = 5$	
No	0·221 (0·024)	0·364 (0·028)	0·473 (0·029)	0·409 (0·038)	0·667 (0·041)	0·873 (0·030)
Yes	0·212 (0·024)	0·316 (0·027)	0·335 (0·028)	0·566 (0·047)	0·843 (0·035)	0·894 (0·034)

5. Remarks

This paper provides a surprisingly simple solution to a long-standing problem. Recently, Wang & Wells (1998) proposed an estimator for the bivariate survival function of (T_1, T_2) by estimating the cumulative conditional hazard of T_2 given $T_1 > t_1$. The estimator was shown to be consistent and asymptotically normal, but no analytical variance expression was given. In an unpublished paper, M. C. Wang and S. H. Chang studied an estimator for a recurrent survival function under the restrictive condition that the gap times of the recurrent events have the same marginal distribution.

Our method, as well as those of Wang & Wells and Wang & Chang, requires the censoring time C to be completely independent of the gap times $(T_1, T_2, \ldots, T_K)$. This could be a practical limitation in some contexts. The method by Visser (1996), on the other hand, can deal with situations where censoring may depend upon previous gap times, but his method relies on estimating the cumulative conditional hazard of T_2 given $T_1 = t_1$ and requires discrete censoring time and gap times.

The estimator $\hat{F}(t_1, t_2)$ is not always a proper distribution function in that it may have negative mass points, though it converges to a proper distribution function as $n \to \infty$. It is difficult to deal with the issue of negative mass for the bivariate estimation with censored data, even for the simpler case of parallel events; see Pruitt (1991). However, it is easy to obtain proper estimators of the marginal distribution functions. The estimator $\hat{F}(t, \infty)$ for $F_1(t)$ shown in (5) is clearly a proper distribution function. We can modify the estimator $\hat{H}(\infty, t)$ given in (6) to produce a proper estimator for $1 - F_2(t)$ while preserving the desired asymptotic properties. Specifically, let $\hat{H}(\infty, 0) = 1$ and $\hat{H}^*(\infty, t) = \inf_{s \leqslant t} \hat{H}(\infty, s)$. Then, $\hat{H}^*(\infty, t)$ is a proper survival function. Furthermore, by the arguments of Lin & Ying (1994, pp. 64–5), the difference between $\hat{H}(\infty, t)$ and $\hat{H}^*(\infty, t)$ is of order $o(n^{-\frac{1}{2}})$ so that $\hat{H}^*(\infty, t)$ has the same limiting distribution as $\hat{H}(\infty, t)$. A similar modification can be made to $\hat{F}_{2|1}$.

The nonparametric estimators such as $\hat{F}$ and $\hat{F}_{2|1}$ are not only important in their own right, but also have many statistical applications. As evident in § 4, it is often of interest to compare the gap time distributions between two or more groups. The estimators $\hat{F}$ and $\hat{F}_{2|1}$ and their variance estimators enable one to make such comparisons at a fixed time point or at a given set of time points. To compare the entire distributions, it is desirable to perform Kolmogorov–Smirnov-type or log-rank-type tests. We are currently exploring the use of the proposed nonparametric estimators in constructing such tests, in assessing the degree of dependence between gap times and in semiparametric regression analysis.

It is straightforward to extend the results of § 2 to the general case of K events. Let $\bar{t} = (t_1, \ldots, t_K)$, $\bar{t}_0 = (t_1, \ldots, t_{K-1}, 0)$,

$$H(\bar{t}) = \mathrm{pr}(T_1 \leqslant t_1, \ldots, T_{K-1} \leqslant t_{K-1}, T_K > t_K), \quad F(\bar{t}) = \mathrm{pr}(T_1 \leqslant t_1, \ldots, T_K \leqslant t_K).$$

Clearly, $F(\bar{t}) = H(\bar{t}_0) - H(\bar{t})$. We estimate $H(\bar{t})$ by $\hat{H}(\bar{t}) := n^{-1} \sum_{i=1}^n \hat{H}_i(\bar{t})$, where

$$\hat{H}_i(\bar{t}) = \frac{I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} \leqslant t_2, \ldots, \tilde{Y}_{K-1,i} - \tilde{Y}_{K-2,i} \leqslant t_{K-1}, \tilde{Y}_{Ki} - \tilde{Y}_{K-1,i} > t_K)}{\hat{G}(\tilde{Y}_{K-1,i} + t_K)},$$

and $\hat{G}$ is the Kaplan–Meier estimator of G based on $(\tilde{Y}_{Ki}, 1 - \delta_{Ki})$ $(i = 1, \ldots, n)$. We then estimate $F(\bar{t})$ by $\hat{F}(\bar{t}) := \hat{H}(\bar{t}_0) - \hat{H}(\bar{t})$. Using the techniques given in the Appendix, we can show that $n^{\frac{1}{2}}\{\hat{H}(\bar{t}) - H(\bar{t})\}$ converges weakly to a K-variate zero-mean Gaussian process with a simple covariance function. From this result, we can also show that $n^{\frac{1}{2}}\{\hat{F}(\bar{t}) - F(\bar{t})\}$ converges weakly to a zero-mean Gaussian process with a covariance function which can be consistently estimated by

$$n^{-1} \sum_{i=1}^n \left[\{\hat{H}_i(\bar{t}_0) - \hat{H}_i(\bar{t}) - \hat{F}(\bar{t})\}\{\hat{H}_i(\bar{t}'_0) - \hat{H}_i(\bar{t}') - \hat{F}(\bar{t}')\} - \frac{(1 - \delta_{Ki})\hat{D}(\bar{t}; \tilde{Y}_{Ki})\hat{D}(\bar{t}'; \tilde{Y}_{Ki})}{n^{-2}\{1 + \sum_{j=1}^n I(\tilde{Y}_{Kj} > \tilde{Y}_{Ki})\} \sum_{l=1}^n I(\tilde{Y}_{Kl} \geqslant \tilde{Y}_{Ki})} \right],$$

where $\hat{D}(\bar{t}; u) = n^{-1} \sum_{i=1}^n \{\hat{H}_i(\bar{t}_0) I(\tilde{Y}_{K-1,i} \geqslant u) - \hat{H}_i(\bar{t}) I(\tilde{Y}_{K-1,i} \geqslant u - t_K)\}$.

Acknowledgement

This work was supported by the U.S. National Institutes of Health, National Science Foundation and National Security Agency. The authors are grateful to the reviewers for their helpful comments.

APPENDIX

Asymptotic properties of the estimators

As argued in § 2, $\hat{H}$ is an unbiased estimator of H. It follows from the strong law of large numbers and the strong consistency of the Kaplan–Meier estimator $\hat{G}$ (Shorack & Wellner, 1986, pp. 304–6) that $\hat{H}$ is strongly consistent, which implies that $\hat{F}$ and $\hat{F}_{2|1}$ are strongly consistent.

It is crucial to study the weak convergence of $W(t_1, t_2) := n^{\frac{1}{2}}\{\hat{H}(t_1, t_2) - H(t_1, t_2)\}$ since both $\hat{F}$ and $\hat{F}_{2|1}$ are functionals of $\hat{H}$. Clearly,

$$W(t_1, t_2) = n^{\frac{1}{2}}\{\tilde{H}(t_1, t_2) - H(t_1, t_2)\} + n^{\frac{1}{2}}\{\hat{H}(t_1, t_2) - \tilde{H}(t_1, t_2)\}.$$

In view of (2),

$$n^{\frac{1}{2}}\{\tilde{H}(t_1, t_2) - H(t_1, t_2)\} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left\{ \frac{I(T_{1i} \leqslant t_1, T_{2i} > t_2, C_i > T_{1i} + t_2)}{G(T_{1i} + t_2)} - H(t_1, t_2) \right\}, \tag{A1}$$

$$n^{\frac{1}{2}}\{\hat{H}(t_1, t_2) - \tilde{H}(t_1, t_2)\} = n^{-\frac{1}{2}} \sum_{i=1}^{n} I(T_{1i} \leqslant t_1, T_{2i} > t_2, C_i > T_{1i} + t_2) \times \left\{ \frac{G(T_{1i} + t_2) - \hat{G}(T_{1i} + t_2)}{G(T_{1i} + t_2)\hat{G}(T_{1i} + t_2)} \right\}. \tag{A2}$$

By a martingale representation for the Kaplan–Meier estimator (Fleming & Harrington, 1991, p. 97),

$$\frac{G(t) - \hat{G}(t)}{G(t)} = \int_0^t \frac{\hat{G}(u-)}{G(u)} \frac{\sum_{i=1}^n dM_i^c(u)}{\sum_{i=1}^n I(\tilde{Y}_{2i} \geqslant u)}, \quad t \leqslant \max_{1 \leqslant i \leqslant n} \tilde{Y}_{2i},$$

where

$$M_i^c(t) = I(C_i \leqslant t \wedge Y_{2i}) - \int_0^t I(\tilde{Y}_{2i} \geqslant u)\, d\Lambda_c(u).$$

Thus, (A2) can be written as

$$\begin{aligned} n^{\frac{1}{2}}\{\hat{H}(t_1, t_2) - \tilde{H}(t_1, t_2)\} &= n^{-\frac{1}{2}} \int_0^{\tau_c} \left\{ n^{-1} \sum_{i=1}^{n} \frac{I(T_{1i} \leqslant t_1, T_{2i} > t_2, C_i > T_{1i} + t_2) I(T_{1i} \geqslant u - t_2)}{\hat{G}(T_{1i} + t_2)} \right\} \\ &\quad \times \frac{\hat{G}(u-)}{G(u)} \frac{\sum_{j=1}^n dM_j^c(u)}{n^{-1} \sum_{j=1}^n I(\tilde{Y}_{2j} \geqslant u)} + o_p(1) \\ &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau_c} \frac{\{H(t_1, t_2) - H(u - t_2, t_2)\}^+}{\mathrm{pr}(\tilde{Y}_2 > u)} dM_i^c(u) + o_p(1), \end{aligned}$$

where the second equality follows from the consistency of $\hat{H}$ and $\hat{G}$, the continuity of H and the fact that $G(u-)/\{G(u)\,\mathrm{pr}(\tilde{Y}_2 \geqslant u)\} = 1/\mathrm{pr}(\tilde{Y}_2 > u)$.

Combining the preceding result with (A1), we have

$$\begin{aligned} W(t_1, t_2) &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \left\{ \frac{I(T_{1i} \leqslant t_1, T_{2i} > t_2, C_i > T_{1i} + t_2)}{G(T_{1i} + t_2)} - H(t_1, t_2) \right\} \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau_c} \frac{\{H(t_1, t_2) - H(u - t_2, t_2)\}^+}{\mathrm{pr}(\tilde{Y}_2 > u)} dM_i^c(u) + o_p(1), \end{aligned} \tag{A3}$$

which is a sum of n independent and identically distributed zero-mean random variables. By the multivariate central limit theorem, $W(t_1, t_2)$ converges in finite-dimensional distributions to a zero-mean Gaussian process with covariance function

$$
\begin{aligned}
&\sigma_H(t_1, t_2; t_1', t_2') \\
&= E\left[\frac{I(T_1 \leqslant t_1, T_2 > t_2, C > T_1 + t_2)}{G(T_1 + t_2)} - H(t_1, t_2) + \int_0^{\tau_c} \frac{\{H(t_1, t_2) - H(u - t_2, t_2)\}^+}{\mathrm{pr}(\tilde{Y}_2 > u)} \, dM^c(u)\right] \\
&\quad \times \left[\frac{I(T_1 \leqslant t_1', T_2 > t_2', C > T_1 + t_2')}{G(T_1 + t_2')} - H(t_1', t_2') + \int_0^{\tau_c} \frac{\{H(t_1', t_2') - H(u - t_2', t_2')\}^+}{\mathrm{pr}(\tilde{Y}_2 > u)} \, dM^c(u)\right] \\
&= E\left[\left\{\frac{I(T_1 \leqslant t_1, T_2 > t_2, C > T_1 + t_2)}{G(T_1 + t_2)} - H(t_1, t_2)\right\} \left\{\frac{I(T_1 \leqslant t_1', T_2 > t_2', C > T_1 + t_2')}{G(T_1 + t_2')} - H(t_1', t_2')\right\}\right] \\
&\quad + \int_0^{\tau_c} \frac{\{H(t_1, t_2) - H(u - t_2, t_2)\}^+ \{H(t_1', t_2') - H(u - t_2', t_2')\}^+}{\{\mathrm{pr}(\tilde{Y}_2 > u)\}^2} \mathrm{pr}(\tilde{Y}_2 \geqslant u)\{1 - \Delta\Lambda_c(u)\} \, d\Lambda_c(u) \\
&\quad - E\left[\frac{I(T_1 \leqslant t_1, T_2 > t_2, C > T_1 + t_2)}{G(T_1 + t_2)} \int_0^{T_1 + t_2} \frac{\{H(t_1', t_2') - H(u - t_2', t_2')\}^+}{\mathrm{pr}(\tilde{Y}_2 > u)} I(\tilde{Y}_2 \geqslant u) \, d\Lambda_c(u)\right] \\
&\quad - E\left[\frac{I(T_1 \leqslant t_1', T_2 > t_2', C > T_1 + t_2')}{G(T_1 + t_2')} \int_0^{T_1 + t_2'} \frac{\{H(t_1, t_2) - H(u - t_2, t_2)\}^+}{\mathrm{pr}(\tilde{Y}_2 > u)} I(\tilde{Y}_2 \geqslant u) \, d\Lambda_c(u)\right] \\
&= E\left[\left\{\frac{I(T_1 \leqslant t_1, T_2 > t_2, C > T_1 + t_2)}{G(T_1 + t_2)} - H(t_1, t_2)\right\} \left\{\frac{I(T_1 \leqslant t_1', T_2 > t_2', C > T_1 + t_2')}{G(T_1 + t_2')} - H(t_1', t_2')\right\}\right] \\
&\quad - \int_0^{\tau_c} \frac{\{H(t_1, t_2) - H(u - t_2, t_2)\}^+ \{H(t_1', t_2') - H(u - t_2', t_2')\}^+}{\mathrm{pr}(\tilde{Y}_2 > u)} \, d\Lambda_c(u).
\end{aligned}
$$

Note that familiar results for martingale integrals (Fleming & Harrington, 1991, Theorem 2.6.2) were used to obtain the second equality.

To complete the weak convergence proof, we need to establish the tightness of W. It suffices to show the tightness of the right-hand side of (A3). By the functional central limit theorem (Pollard, 1990, p. 53), the tightness of the first term reduces to the manageability of

$$\{n^{-\frac{1}{2}} I(T_{1i} \leqslant t_1, T_{2i} > t_2, C_i > T_{1i} + t_2)/G(T_{1i} + t_2); \; i = 1, \ldots, n\}.$$

However, the manageability is a direct consequence of the fact that each of these n terms can be written as a product of four monotone functions of t_1 or t_2. For the tightness of the second term, we note that, via the integration-by-parts formula, it suffices to show the weak convergence of

$$\int_0^{\tau_c} n^{-\frac{1}{2}} \sum_{i=1}^n M_i^c(u) \, d[\{H(t_1, t_2) - H(u - t_2, t_2)\}^+ / \mathrm{pr}(\tilde{Y}_2 > u)].$$

This follows from the facts that the integrand, which is a martingale process, converges weakly (Fleming & Harrington, 1991, Theorem 5.3.5) and that the integral with the integrand replaced by its limiting process is continuous in t_1 and t_2.

Since $\hat{F}(t_1, t_2) = \hat{H}(t_1, 0) - \hat{H}(t_1, t_2)$, the weak convergence for $\hat{F}$ follows immediately from that of W. By the arguments used in evaluating σ_H, it can be shown that (7) is indeed the covariance function for $\hat{F}$. It is natural to estimate $\sigma(t_1, t_2; t_1', t_2')$ by

$$
\begin{aligned}
\hat{\sigma}(t_1, t_2 \colon t_1', t_2') &:= n^{-1} \sum_{i=1}^n \left\{\frac{\delta_{1i} I(\tilde{Y}_{1i} \leqslant t_1)}{\hat{G}(\tilde{Y}_{1i})} - \frac{I(\tilde{Y}_{1i} \leqslant t_1, \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2)}{\hat{G}(\tilde{Y}_{1i} + t_2)} - \hat{F}(t_1, t_2)\right\} \\
&\quad \times \left\{\frac{\delta_{1i} I(\tilde{Y}_{1i} \leqslant t_1')}{\hat{G}(\tilde{Y}_{1i})} - \frac{I(\tilde{Y}_{1i} \leqslant t_1', \tilde{Y}_{2i} - \tilde{Y}_{1i} > t_2')}{\hat{G}(\tilde{Y}_{1i} + t_2')} - \hat{F}(t_1', t_2')\right\} \\
&\quad - \int_0^{\tau_c} \frac{\hat{D}(t_1, t_2; u) \hat{D}(t_1', t_2'; u) \, d\hat{\Lambda}_c(u)}{n^{-1}\{1 + \sum_{j=1}^n I(\tilde{Y}_{2j} > u)\}},
\end{aligned} \tag{A4}
$$

which is obtained by replacing the unknown parameters on the right-hand side of (7) by their respective sample estimators. Here,

$$\hat{\Lambda}_c(t) := \int_0^t \frac{\sum_{i=1}^n (1-\delta_{2i})\, dI(\tilde{Y}_{2i} \leqslant u)}{\sum_{j=1}^n I(\tilde{Y}_{2j} \geqslant u)}$$

is the Nelson–Aalen estimator for $\Lambda_c(t)$ based on $(\tilde{Y}_{2i}, 1-\delta_{2i})$ $(i=1,\ldots,n)$. Note that we estimate $\mathrm{pr}(\tilde{Y}_2 > u)$ by $n^{-1}\{1+\sum I(\tilde{Y}_{2j} > u)\}$ rather than $n^{-1}\sum I(\tilde{Y}_{2j} > u)$ to avoid the possibility of having a zero denominator in (A4). It is easy to see that (A4) equals (8). The strong law of large numbers, together with the strong consistency of $\hat{H}$, $\hat{G}$ and $\hat{\Lambda}_c$, implies that $\hat{\sigma}$ is a consistent estimator of σ.

It is straightforward to show that

$$\begin{aligned} n^{\frac{1}{2}}\{\hat{F}_{2|1}(t_2|t_1) - F_{2|1}(t_2|t_1)\} = \frac{n^{\frac{1}{2}}}{F_1(t_1)}[\{1-F_{2|1}(t_2|t_1)\}\{\hat{H}(t_1,0)-H(t_1,0)\} \\ - \{\hat{H}(t_1,t_2)-H(t_1,t_2)\}] + o_p(1), \end{aligned}$$

which indicates that the weak convergence for $\hat{F}_{2|1}$ also follows directly from that of W. In addition, calculations similar to those used in deriving σ_H produce the limiting covariance function

$$\begin{aligned} \sigma(t_2; t_2'|t_1) := \frac{1}{F_1^2(t_1)}\Bigg(E\Bigg[\left\{H(t_2|t_1)\frac{\delta_1 I(\tilde{Y}_1 \leqslant t_1)}{G(\tilde{Y}_1)} - \frac{I(\tilde{Y}_1 \leqslant t_1, \tilde{Y}_2 - \tilde{Y}_1 > t_2)}{G(\tilde{Y}_1 + t_2)}\right\} \\ \times \left\{H(t_2'|t_1)\frac{\delta_1 I(\tilde{Y}_1 \leqslant t_1)}{G(\tilde{Y}_1)} - \frac{I(\tilde{Y}_1 \leqslant t_1, \tilde{Y}_2 - \tilde{Y}_1 > t_2')}{G(\tilde{Y}_1 + t_2')}\right\}\Bigg] \\ - \int_0^{\tau_c} \frac{B(t_1, t_2; u)B(t_1, t_2'; u)\, d\Lambda_c(u)}{\mathrm{pr}(\tilde{Y}_2 > u)}\Bigg), \end{aligned} \tag{A5}$$

where $H(t_2|t_1) = H(t_1, t_2)/F_1(t_1)$ and

$$B(t_1, t_2; u) = H(t_2|t_1)\{F_1(t_1) - F_1(u)\}^+ - \{H(t_1, t_2) - H(u - t_2, t_2)\}^+.$$

Replacement of the unknown parameters in (A5) by their respective sample estimators yields formula (9).

REFERENCES

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. & KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

GAIL, M. H., SANTNER, T. J. & BROWN, C. C. (1980). An analysis of comparative carcinogenisis experiments based on multiple times to tumor. *Biometrics* **36**, 255–66.

GUMBEL, E. J. (1960). Bivariate exponential distributions. *J. Am. Statist. Assoc.* **55**, 698–707.

LIN, D. Y. & YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

MOERTEL, C. G., FLEMING, T. R., MCDONALD, J. S. et al. (1990). Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New Engl. J. Med.* **322**, 352–8.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. Hayward, CA: Inst. Math. Statist.

PRUITT, R. C. (1991). On negative mass assigned by the bivariate Kaplan–Meier estimator. *Ann. Statist.* **19**, 433–53.

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

SUSARLA, V., TSAI, W.-Y. & VAN RYZIN, J. (1984). A Buckley–James estimator for the mean with censored data. *Biometrika* **71**, 624–5.

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

WANG, W. & WELLS, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85**, 561–72.

[*Received May* 1997. *Revised December* 1997]